Biomarker-Based Characterization of Chemical Exposures and Physiological Responses

by

Vy Kim Nguyen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in The University of Michigan
2020

Doctoral Committee:

Assistant Professor Justin Colacino, Co-Chair
Professor Olivier Jolliet, Co-Chair
Professor Kayvan Najarian
Associate Professor Laura Rozek
Associate Professor Maureen Sartor
Dr. John Wambaugh, U.S. Environmental Protection Agency

Vy Kim Nguyen

nguyenvy@umich.edu

ORCID iD: 0000-0002-6128-0523

DEDICATION

I dedicate this work to my mom to honor her for shaping me into the person who I am today. She sacrificed her dream job as a chemical engineer in Vietnam, so that my sister and I can be educated in the United States. She tutored me in science and math, particularly in 2nd grade math because I was failing at the time, thus instilling within me a passion for science and math that has carried me through to this milestone.

ACKNOWLEDGEMENTS

I would first like to acknowledge my thesis co-advisors, Olivier Jolliet and Justin Colacino. I am very grateful that Olivier responded to my email titled “Your research is awesome!”, because his response facilitated this amazing and rewarding journey of our collaboration. Olivier is a dynamic and entertaining orator, inspiring me to incorporate such attributes into my own presentations. Olivier also has a profound skill in data interpretation, and I am so grateful to have him to challenge me to make informative data visualization tools. I am very grateful that Justin agreed to co-mentor me since the beginning of my graduate studies. Justin offered structure and organization to complement Olivier’s big research ideas, which was integral in helping me navigate the non-linear workflow of research. Justin is also an amazing writer with his mentorship helping to substantially improve my own scientific writing. Since using the ARC-TS cluster involves being charged by the second, Justin said that he would prefer not to receive a bill charged with an amount equivalent to my stipend, therefore encouraging me to optimize my code and helping me to achieve a dream goal of learning how to parallelize. I learned not only how to be a good scientist from Olivier and Justin, but due to their compassion for my wellbeing, I have also learned how to live a balanced life.

I would also like to acknowledge Chirag Patel’s mentorship. Chirag inspired me to conduct research in medicine, and due to his support, my research took an interesting turn into biomedical informatics. Chirag made me learn the R tidyverse package, and it would be an understatement to say that learning tidyverse has been life changing. I code faster and analyze data more efficiently

than ever before. I learned not only how to be a rigorous computational scientist from Chirag, but due to his superb etiquette, I have also learned how to be polite and constructive.

I would also like to acknowledge the other members of my dissertation committee, Maureen Sartor, Kayvan Najarian, Laura Rozek, and John Wambaugh for their outstanding input and support throughout this process. I am grateful for Maureen's and Kayvan's inputs to ground my research in data science. I am grateful for Laura's and John's endless enthusiasm for my research projects and plots. I am honored to have been mentorship by such exceptional scientists.

Within the Jolliet (iMod) lab, I would like to thank Lei Huang, Jacob Kvasnicka, Katerina Stylianou, Karin Veltman, Cedric Wannaz, and Xingyue Zhang. Their guidance, mentoring, support, and feedback have contributed to help me grow as a rigorous scientist. In addition, their friendship has made the past four years so enjoyable as characterized with laughter and jokes in the office, lab outings, culinary adventures, and travels in Europe. I would particularly like to thank Jacob for his assistance and patience in poring over my writing and providing thoughtful feedback. I would like to thank Katerina as she played an integral part in helping me reach candidacy as we spend hours discussing how to address edits for my preliminary exam. I would also like to thank Cedric for his eagerness to share his computational knowledge along with his free and very honest advice.

Within the Colacino Lab, I would like to thank Anagha Tapaswi, Nick Cemalovic, Chanese Forte, Juliet Heidt, Evan Hill, Adam Kahana, Sarah Karram, Lauren Middleton, Rachel Morgan, Nick Polakowski, Katelyn Polemi, Tasha Thong, and Sabrina Rocco. Their support, exchange of idea, and friendship have made my graduate studies that much more enjoyable. I cherish our summer outings, party sessions, conference attendances in Baltimore and Andover, and culinary adventures. I particularly like to thank Lauren Middleton for her unwavering support in all that I

do and for her insightful and nuanced advice. I particularly like to thank Lauren, Rachel, Katelyn, and Sabrina for their thoughtfulness in putting together a care package to encourage me to finish this degree.

Within the Patel (Rag) group, I would like to thank Jake Chung, Kajal Claypool, Andrew Deonarine, Undina Gisladottir, Alan Le Goallec, Chirag Lakhani, Nicole Parker, Nam Pho, Danielle Rasooly, Sivateja Tangirala, Braden Tierney, Tom van der Meer, and Jiaqi Xie. Their hospitality, encouragement, feedback, and support make my externship at Harvard a pleasure and honor. My repertoire of computational skills substantially improved as they provide the support and advice to further ground my dissertation in data science. I particularly would like to thank Alan for sharing his coding knowledge, because without his him, the mortality project would have been substantial harder. I would like to thank Jake for always be game to exchange idea and discuss analysis. I always look forward to these meetings, and I walk away so inspired. I would also like to thank Nicole for handling all the logistics to help facilitate the collaboration with Michigan and Harvard.

I would like to thank the Margit Burmeister, Julia Eussen, and Maureen Sartor of for being pillars of support for me in the Department of Computational Medicine and Bioinformatics. I would like to thank Margit for fighting for me to come to University of Michigan as I was on the waitlist. I particularly would like to thank Julia for her encouragement and compassion as she was willing to comfort me when I was struggling with a class.

I would like to acknowledge my collaborators Jon Arnot, Kelly Bakulski, Li Li, Joe Okeme, Hyeong-Moo Shin, Ming Xu, and Bu Zhao. I thank them for working with my crazy ideas and providing me guidance to help me grow as a better scientist.

I would like to thank the executive committee members of Girls Who Code: Brooke Wolford, Zena Lapp, Marlena Duda, Negar Farzaneh, Shweta Ramdas, Rucheng Diao, Verity Sturm, Stephanie Thiede, Saige Rutherford, Gabby Dotson, Kelly Sovacool, Ruma Deb, Hayley Falk, Morgan Oneka, Katie Furman, and Sarah Haynes. I'm grateful to this group of inspirational female scientists for creating a supportive environment to teach high school women not only how to code but also how to conduct research in data science. They have enabled me to meet and mentor Sarah, and seeing her present at an international conference as a high school student was one of the most rewarding experiences of my life.

I would like to acknowledge the work of the Department of Environmental Health's tremendous staff: Sue Crawford, Charlotte Carlson, Kelsey Hargesheimer, Jennifer Lewis, and Carmen Rey, and Patrice Sommerville. I'm grateful of their work to handle the logistics tied to research, because I would have been lost without them.

I would not have accomplished any of this work without the friendship of Ashton Baker, Catherine Barnier, Tricia Charles-Jones, Rucheng Diao, Negar Farzenah, Gayatri Iyer, Brandon Govindarajoo, and Hera Shi. I particularly like to thank Ashton, Tricia, Rucheng, and Negar for their support in our bioinformatics and math classes as well as in my personal life. The first few years of grad school was very challenging, and I would have dropped out if not for their encouragement and compassion. I would like to thank Gayatri for all her time spent studying and helping me in our introductory machine learning class. I would like to thank Catherine for sending me a slew of pictures of cute animals to encourage me to finish this degree. I would like to thank Brandon for helping me to achieve candidacy as we discussed for hours on how to ground my proposal in data science for my preliminary examination. I would like to thank Hera, because

without her, I would not have pass, my hardest course: the machine learning course from the department of Electrical Engineering and Computer Science.

Finally, I would like to thank my sister, Nita and my parents, Paul and Holly for their unconditional love and support over the years. My sister was the first to support my decision to change career from being a physician to a scientist. Though younger than me, my sister amazes me with her creativity, values, and humor. My parents sacrificed their stable life in Vietnam and immigrated to the United States to ensure that my sister and I would have a better life. I grew up seeing my parents work so hard to move us from a one-bedroom apartment in the city to a two-story house in the suburbs. The manifestation of their hard work instilled a strong work ethic with in me that can help me achieve any goal. This degree is truly an attestment to them for achieving the American Dream and thus helping me to achieve mine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

The chemisome is the chemical components of the exposome, defined as the totality of all exposures and their impact on health. Most current approaches, however, are limited in addressing this "totality" by only studying one chemical or one chemical family at a time in one exposed population. In addition, studying the links between chemical exposures and health is challenging due to an incomplete understanding of how physiological responses are associated with adverse health outcomes. This challenge is further complicated due to how chemical exposures change with demographics such as age, sex, race, and occupation. Thus, this dissertation aims to address these challenges by applying an unbiased approach to datasets of chemical biomarker levels and physiological measurements to systematically identify susceptible populations using the National Health and Nutrition Examination Survey.

In the first project, I use quadratic regression models to characterize non-linear, age-based trends of chemical exposure in a sample comprised of 74,942 participants. I screen across 141 chemicals to identify those of higher concentrations in children relative to the older population. Children exhibit higher exposures to chemicals in consumer products such as phthalates, brominated flame retardants, lead, and tungsten. In contrast, restricted and highly persistent chemicals such as polychlorinated biphenyls and dioxins are higher in the older population.

In the second project, I apply generalized linear models to evaluate exposure disparities by race/ethnicity for 143 chemicals in a representative sample of 38,080 US women. Compared to non-Hispanic White women, significant disparities are observed for non-Hispanic Black, Mexican

American, Other Hispanic, and Other Race/Multi-Racial women. These women have higher levels of pesticides, including 2,5-dichlorophenol and 2,4-dichlorophenol, compounds in personal care products, including parabens and mono-ethyl phthalate, and heavy metals, such as mercury and arsenic. These findings are being coupled with toxicological data to prioritize chemicals to evaluate their role in health disparities.

In the third project, I develop a framework using hierarchical clustering to characterize occupational exposures and physiological responses among 26,186 blue- and white-collar workers across 20 employment sectors for 108 chemicals and 27 physiological indicators. Blue-collar workers have higher levels of toxicants such as lead, cadmium, volatile organic chemicals, and polycyclic aromatic hydrocarbons compared to white-collar workers. Moreover, blue-collar workers exhibit higher levels of alkaline phosphatase (indicative of liver disease) and C-reactive proteins (indicative of inflammation). Together, these results suggest that blue-collar workers are exposed to higher levels of toxicants, which may induce physiological dysfunction.

In the final project, I implement 10-fold cross-validated regression models to characterize the linear and non-linear associations between all-cause mortality and 27 physiological indicators to identify directionalities indicative of increased mortality risk in a sample of 45,032 participants. Twenty-four out of 27 indicators show non-linear associations, while height, triglycerides, and 60-second pulse show linear associations. Cholesterol-related indicators and glomerular filtration rate unexpectedly show parabolic associations, implying that higher mortality risk is associated with measurements in either extreme of the distribution instead of in one extreme. These findings highlight a need to study associations between these indicators and other health endpoints to gain insights into the physiological profiles associated with adverse health outcomes.

Together, this thesis contributes to a better understanding of how chemical exposures can impact human health across multiple subpopulations. It also enables further exploration of how chemical exposures can perturb physiologic function conducive to increasing the risk for adverse health outcomes.

Chapter 1 Introduction

1.1 Background and Motivation

As early as the 1890s, factory workers were screened for lead in their blood and urine to prevent elevated levels from reaching the point of acute lead poisoning [1]. Here is one of the earliest examples of tracking biomarker levels to characterize the impact of chemical exposures. While this example pertains to occupational exposures, we are exposed to thousands of environmental substances everyday based on where we live, where we work, what we drink, what we eat, what we use, etc. Some of these substances may be toxic, but it is challenging to know which ones and the levels at which an adverse effect would occur.

In 2005, Christopher Wild termed the exposome as the totality of all exposures in an individual's lifetime and its impact on human health [2]. This totality of exposures can consist of, but is not limited to, ecosystem, lifestyle factors, social indicators, and physical-chemical factors. The concept of the exposome was developed as the complement to the genome, and that an integrated understanding of both the exposome and the genome would facilitate progress toward addressing chronic human health problems [3]. Genetics account for only about 10%-40% of the contribution to most diseases [4–6], suggesting that environmental exposures may play a greater role in understanding the etiology of disease compared to genetic factors alone.

The chemisome is defined as the chemical components of the exposome and how it affects human health [7]. Toxic chemicals can enter the body from exogenous sources such as air, water, and lifestyle behaviors such as dietary habits and medication use [8,9]. Toxicants can also be

generated by endogenous processes such as inflammation, oxidative stress, diseases, and infection [10,11]. As environmental insult can perturb physiological functions conducive to disease [12], characterizing biologically active chemicals inside the human body will promote the discovery of exposures responsible for chronic disease [5,13]. Thus, better understanding of how environmental pollutants relate to health problems requires human biomonitoring.

Human biomonitoring is the measuring of chemicals or their metabolites in the human body to ascertain the extent of human exposure to environmental toxicants [14]. A chemical biomarker is an indicator of environmental exposures quantified as the chemical, its metabolites, or the products of an interaction between the chemical and a biological target in an organism [15,16]. The largest continuous source of data on chemical biomarkers in the United States (US) is the National Health and Nutrition Examination Survey (NHANES) [17]. The Centers for Disease Control and Prevention (CDC) designed this survey to ascertain the nutritional and health status of children and adults in the US. NHANES deploys a complex, multistage, probability sampling design to ensure that the data is representative of the noninstitutionalized, civilian US population [18]. Since 1999, the CDC has conducted the continuous NHANES survey by inviting several thousands of volunteers every two years to participate in interviews, questionnaires, and examination. Biological samples are analyzed for indicators of disease, disorders, and chemical biomarkers. These data are publicly available online accompanied with data on demographics, questionnaire response, physiological measurements, and other laboratory tests [19].

In our current, cleaned NHANES dataset, we have blood, serum, and urinary measurements available for 411 chemical biomarkers to analyze exposures and 60 physiological indicators to analyze the physiologic condition of a representative US sample of 82,091 participants. The diverse suite of chemical biomarkers are from 17 different chemical classes: Acrylamide,

Brominated Flame Retardants (BFR), Phosphate Flame Retardants (PFR), Dioxins, Furans, Melamine, Metals, Other, Personal Care & Consumer Product Compounds, Pesticides, Phthalates & Plasticizers, Phytoestrogens, Polyaromatic Hydrocarbons (PAH), Polychlorinated Biphenyls (PCB), Per- and Polyfluoroalkyl Substances (PFAS), Smoking Related Compounds, and Volatile Organic Compounds (VOC). The physiological indicators comprise a wide range of human body systems including body composition, cardiovascular system, respiratory system, immune system, metabolic system, and nephrology. Such wealth of data is conducive to conducting exposome research to evaluate the impact of environmental insults on human physiology.

While exposomic data such as NHANES provides hope to better understand the environmental contribution to disease and how to prevent such diseases, addressing this problem requires addressing the scale of exposomic data. Measuring techniques such as omics and sensor technologies will generate heterogenous, massive, and high-velocity data [20]. Thus, there is a need to develop approaches to manage, analyze, and visualize these large-scale data to begin untangling the intertwining associations between genetics, exposures, and disease [3]. Inspiration can be drawn by the approaches implemented in genome-wide association study (GWAS), which involves screening across millions of genetic markers to identify those that can be used to predict the onset of a particular disease [21]. In fact, inspired by techniques used in GWAS, Patel et al. conducted an environment-wide association study (EWAS) on chemical, clinical, and questionnaire data from NHANES to evaluate the associations between 266 chemical biomarkers and type 2 diabetes mellitus and serum lipid levels, respectively [22]. Patel et al. observed strong associations between the risk of type 2 diabetes and exposures to heptachlor epoxide, g-tocophperol, b-carotenes and polychlorinated biphenyls. Such results suggest the importance of implementing an unbiased

approach to gain insights on the role of the environment in disease onset and ultimately lead to the prevention of chronic disease [3].

However, most current approaches tend to implement a one-chemical [23–27] or one-chemical-family [28–34] when studying exposures, which is not only limited in addressing the totality of exposures but also limited in prioritizing chemicals for further study. Furthermore, addressing the totality of exposures is further complicated by the dynamic nature of an individual's exposome [3], since it changes with demographic factors such as age, sex, race, and occupation and other lifestyle factors such as dietary habits, physical activities, etc. However, this one-chemical or one-chemical-family approach is commonly applied to one exposed subpopulation instead of multiple different subpopulations [35–38], thus hindering progress towards identifying populations vulnerable to chemical-mediated effects. While this approach is necessary for an in-depth, detailed analysis, this single-chemical or single-family approach can complement an unbiased approach that systematically identifies chemicals with high exposures in specific susceptible populations to provide wealth and depth in understanding the exposure profiles associated with vulnerable populations. Hence, I have organized sections 1.2, 1.3, and 1.4 to discuss limitations in current approaches for studying age-based trends, characterizing racial exposure disparities in women in the context of breast cancer, and evaluating the impact of occupational co-exposures on worker's health, respectively. For section 1.3, I focused primarily on breast cancer disparities due to interest of the Colacino Lab to prioritize toxicants for further experimental work to understand their role in breast cancer disparities.

There are also limitations in addressing how chemical exposures influence health. For ease of interpretation, most studies assume linearity when modelling the associations between chemical exposures and a given disease [22,39]. Linear models cannot detect whether participants with either

higher or lower exposures are at increased risk, as these models are limited to only detecting one subpopulation at risk. For example, lower levels of vitamin A are associated with increased risk for blindness, infection, rashes, and higher levels are associated with bone pain, liver damage, and increase pressure on the brain [40]. A linear model would suggest one of the following: 1) higher levels, 2) lower levels, or 3) no levels are associated with increased risk. This shows the limitations of the using a linear model to detect susceptible participants, especially in situations when participants with measurements in either extremes of exposures are at risk. Furthermore, while these studies provide evidence of environmental influence on human health [41–45], they do not provide insight on how exposure may perturb important physiological mechanisms conducive to increasing disease risk. Therefore, we first need to gain insights into understanding the physiological profiles associated with adverse outcomes [46–51]. In addition, there is a need to develop a framework to study the linear and non-linear relationship between physiological indicators and adverse health outcomes. This framework will not only be useful in identifying directionalities associated with increased disease risk, but it will also enable further exploration on how chemical exposures perturb physiological profiles in a manner that increases risk for a given adverse outcome. In section 1.5, I further discuss the limitations of the current approaches to modeling the associations between physiological indicators and health outcomes.

Figure 1.1 illustrates the contributions from studies published within the last two decades that characterized associations along the continuum of chemical indicators, physiological indicators, and health endpoints. These domains converge toward the high-throughput evaluation of chemical exposures to simultaneously prioritize chemicals and identify susceptible populations along with developing a framework to characterize linear and non-linear associations along this continuum.

1.2 Need to Systematically Identify Highly Exposed Chemicals in Children

Children are especially susceptible to toxicant exposures due to an assortment of physiological traits and normal development behaviors. For instance, children tend to have higher metabolic rate [52,53], which facilitates the absorption of toxicants. Infants often have higher body burdens of chemicals, with estimated half-lives of several chemicals being 3-9 times longer than those found in adults [54]. These higher half-lives may be due to childrens' developing metabolism pathways. Due their small body sizes, children are more susceptible to adverse effects if exposed to the same doses as adults [55]. Children are rapidly growing and developing, leading to changes in their organ system functioning, which may modify the effects from toxicant exposures [56]. For example, developing organ systems, such as the central nervous system [57] and respiratory system [58], are more susceptible to environmental insults in children than in adults. Furthermore, children are especially sensitive to adverse effect of toxic exposures as childhood development is a crucial window of susceptibility [56]. Children may also be susceptible to higher exposures via routes of exposures associated with behaviors linked to normal development such as crawling [59], mouthing [60,61], and playing [62]. Therefore, approaches to identify chemicals of higher biomarker levels in children will help prioritize chemicals for further toxicologic and epidemiologic assessment and implement policy or regulations to prevent such exposures and subsequent adverse effects.

Children are exposed to a wide ensemble of chemicals [63]. Commonly used as plasticizers, especially in toys, phthalates are found at higher concentrations in children with links to developmental and reproductive toxicity in animals [64]. Brominated flame retardants such as Polybrominated Diphenyl Ethers (PBDEs) are also found at higher levels in children [65] and have been shown to affect the developing brain and thyroid homeostasis [66], thus affecting the regulation

of many body functions due to irregular release of thyroid hormones in the bloodstream. Blood lead concentrations are declining [67] but are still found at levels associated with neurotoxicity and intellectual impairment, as no safe level of lead has been identified [68,69]. From a literature review, children tend to have higher biomarker levels of bisphenol A (BPA), phytoestrogens, perchlorate, and metabolites of PAHs and benzene [70]. From a range of many different chemicals, which of these chemicals should be prioritized for further risk assessment and consideration for implementing legislation to prevent such exposures and related adverse outcomes? While these studies are informative for detailed analysis, studying one chemical or one chemical family at a time is limited in addressing the totality of chemical exposures in children. Hence, there needs to be an untargeted approach to prioritize chemicals by exposures and severity of health problems in children.

Studying age-based trends is salient to applying an untargeted approach to identify a set of chemicals found at higher concentrations in children than in adults. Studying such trends are also informative into gaining insights on the history of exposures as age can be used as a surrogate for time. In addition, studying age-based trends enables us to observe the efficacy of legislation and policy on chemical productions and its subsequent effects [71,72]. Furthermore, identifying the drivers of age-based trends of chemical exposures will help prioritize susceptible populations and guide intervention strategies to prevent chemical-mediated adverse health outcomes. Several mechanistic models have been developed to study the relationship between chemical exposures and age for PCBs [73,74], dioxins [75], and selected PFASs such as PFOS and PFOA [76,77]. While these models have enabled the identification of biological half-lives, restriction dates, and change of intake as important drivers of age-based trends, these models have been limited to persistent and legacy chemicals such as PCBs and Dioxins and required a substantial amount of data to

parameterize. These stipulations create challenges in applying an unbiased approach to characterize drivers of age-based trends and identify high risk populations across a broad set of chemicals. Thus, using human biomonitoring data containing measurements for hundreds of chemicals will enable the characterization of age-based trends and their determinants to help prioritize chemicals in children.

1.3 Role of Environmental Insult on Breast Cancer Disparities

There are stark racial disparities in breast cancer among women in the US. For example, non-Hispanic Black women are 40% more likely to die from breast cancer compared to women of any other race [78]. Furthermore, the incidence of non-Hispanic Black developing triple negative breast cancer, the most aggressive subtype of breast cancer, is three times higher compared to that of non-Hispanic White women [79,80]. Therefore, understanding the etiology of these disparities is integral to design targeted interventions to ensure health equity.

While the mechanisms driving these breast cancer disparities are likely due to complex interactions between genetic and environmental factors, the contribution of genetic variations appears minor in explaining cancer disparities [81–83]. For instance, a meta-analysis found that genetic risk factors explain less than 5% of the variation in cancer disparities [84]. Furthermore, high penetrance inherited genes contribute to only 5-10% of breast cancer [85], leaving a substantial proportion to be possibly explained by the environment and hence preventable with targeted intervention [86,87].

Differences in chemical exposures have been hypothesized to be pertinent drivers in racial disparities of disease [78–83]. The role of chemical exposures in breast cancer disparities is not well studied nor understood. Thus, understanding the etiology of environmental insults on breast cancer disparities require comprehensive characterization of differences in chemical exposures.

Furthermore, systematic characterization of exposure differences across a diverse set of chemical indicators will enable the identification of toxicants showing the highest disparities and hence the prioritization of these toxicants for further experimental work to better understand the role of environmental insult in increasing the risk of breast cancer.

1.4 Susceptibility of Workers to Occupational Exposures and Related Effects

Workers are vulnerable to chemical-mediated effects as occupational exposures have been identified as causal factors in a wide variety of disorders and diseases. Prolonged exposures to hazardous chemicals even at lower doses have been identified as a causal factor in cancer [88,89]. According to the National Institute for Occupational Safety and Health (NIOSH), 13 million workers in the US are exposed to chemicals via dermal absorption, which can result in occupational skin diseases and systemic toxicity [90]. According to the United Nations, a worker dies from toxic occupational exposure every 30 seconds [91].

These statistics lend urgency to characterize the totality of exposures across different industries and occupations to identify workers most susceptible to chemical-mediated health effects. Despite these alarming statistics, most studies tend to characterize occupational exposures by implementing a one-chemical or one-chemical family approach in a selected group of workers [30,35,37]. Such approaches are limited in prioritizing chemicals for further analysis. They are also not representative of the totality of chemical exposures as individuals are exposed to multiple chemicals in an occupational setting. In addition, characterizing co-exposures is important in understanding the synergistic effect of chemical exposures on disease. For example, workers, who were exposed to asbestos but do not smoke, have a 70% higher risk of lung cancer compared to the reference group of non-smoking workers who were not exposed to asbestos. While an increased risk of 70% is substantial, this risk is minimal when compared to the case of workers, who are

exposed to asbestos and smoke, as the likelihood of having lung cancer for this group is 9 times higher compared to the reference group [92]. Despite these findings, studying co-exposures have been limited to one chemical family [93–95], but populations are not only exposed to one chemical class. Thus, there is a need to develop a framework to characterize co-exposures across multiple chemical classes. This is essential to identify workers, whose exposure profiles are characterized by elevated biomarker concentrations of the most toxic substances as they would benefit from targeted intervention.

Understanding how chemical exposures may elicit or diminish important physiological responses will be integral to better understand the role of occupational exposures in causing diseases. Physiological responses can be characterized by measurements of various biomarkers known as physiological indicators that characterizes the normal functioning or dysfunction of the different systems in the human body. While some have studied the associations between a chemical or chemical family and a physiological indicator [30,35,96], this is not representative of the totality of exposure on human health. Few have evaluated the impact of multiple chemical on physiological response, but this was limited to one physiological indicator [97]. Physiological function is not solely characterized by only one physiological indicator. Thus, evaluating the impact of chemical co-exposures on a diverse suite of physiological indicators is fundamental to better understand the influence of environmental insults on human health, especially in an occupational setting.

1.5 Characterizing Associations Along the Spectrum of Chemical Exposures to Physiological Indicators to Adverse Health Outcomes

While several studies have provided evidence of the influence of chemical exposures on disease onset [22,39,98], there are several limitations in these approaches. First, the sample size for some studies are low [99–102], leading to decreased statistical power in detecting significant

differences. In addition, these studies do not adjust for multiple comparisons, which may potentially lead to spurious findings [103]. Second, most studies assume a linear association when modeling the relationship between chemical exposures and disease risk [22,39,98]. Others do evaluate non-linear effects [104,105], but very few studies compare the prediction performance between linear and non-linear models to determine which type of association best describe the dose-response relationship [96,106]. Finally, while these studies provide evidence of environmental influence on human health, they do not provide insight on how exposure may perturb important physiological mechanisms conducive to increasing risk for the disease. Furthermore, we need to first evaluate the physiological profiles associated with adverse outcomes, so that we can have the foundation to understand how chemical exposures can modulate the associations between physiological function and adverse outcomes.

Understanding whether associations between physiological indicators and health outcomes are linear or non-linear will enable identification of populations prone to increase risk for adverse outcomes. Most association studies assume linearity [107–110] or non-linearity [111–115] separately. Furthermore, these studies do not quantitatively evaluate whether linearity or non-linearity better describes the associations between a given physiological indicator and a health outcome. While there are many studies that do compare the prediction performance of various models such as linear, quadratic, cubic, and logarithmic [116,117], these results are not validated on new data. Without any validation, the more flexible model, usually the non-linear model, is deemed better at characterizing the associations, which may reflect the actual relationship or show an overfitted model [118]. Thus, these limitations emphasize the need for a statistical framework to establish the appropriate model to best characterize the association between a physiological indicator and a health outcome while also evaluating for overfitting by using cross-validation and metrics such as

the AIC or Bayesian Information Criterion to penalize for added complexity in the non-linear models. This framework will enable identification of directionalities of the physiological indicator associated with increased disease risk. In addition, it will be applicable to characterize the associations between chemical exposures and physiological indicators and health outcome, respectively. Such applications will lead to the identification of the chemical doses conducive to elicit a negative physiological response leading to an adverse outcome.

1.6 Objectives and Specific Aims

The main objective of this thesis is to better characterize the totality of chemical exposures and to gain insights into understanding how chemicals exposures perturb physiological function in a manner that increase onset risks for adverse health outcomes. More specifically, to address the aforementioned limitations and challenges, I define the following four specific aims.

Specific Aim 1 (Chapter 2): Identify toxicants of higher biomarker levels in children by characterizing age-based trends in a representative sample of the US population. More specifically, this aim entails: 1) understanding the influence of temporal drivers, in particular time trends, biological half-lives, and restriction dates on age-based trends, 2) systematically defining an age-based pattern to identify chemicals with ongoing and high exposure in children, and 3) characterizing how age-based trends for six PFASs are changing over time to evaluate the criteria indicative of legacy exposures.

Specific Aim 2 (Chapter 3): Conduct a comprehensive analysis of racial disparities in chemical biomarker concentrations among US women to identify chemicals of higher exposures in Non-Hispanic Black women. More specifically, this aim involves: 1) evaluating chemical exposures disparities by race, 2) evaluating how these exposure disparities varies across the life stage, and 3) characterizing how these exposure disparities change over time.

Specific Aim 3 (Chapter 4): Characterize occupational exposures and physiologic dysfunction in a working US population comprised of NHANES participants from a broad range of industrial sectors and occupations. More specifically, this aim involves: 1) characterizing differences in chemical exposure and physiological response profiles across combinations of sectors and occupations, 2) identifying groups of workers with similar chemical exposure and physiological response profiles, and 3) characterizing the physiological dysfunction of workers who have similar chemical exposure profiles.

Specific Aim 4 (Chapter 5): Characterize the relationships between all-cause mortality and 27 physiological indicators in the US population. All-cause mortality was selected as it is the ultimate health endpoint. More specifically, this aim includes: 1) comparing the prediction performance of linear and different nonlinear models by applying a machine learning approach, 2) assessing the robustness of the models by observing changes in prediction performance when extreme measurements are excluded, 3) describing the associations between the physiological indicator and mortality as characterized by the most appropriate model(s), and 4) determining the relevance of the current clinical thresholds by evaluating whether these values are indicative of increased mortality risk.

1.7 Dissertation Outline

This dissertation is organized based on the aforementioned specific aims with Chapter 2-5 used to address each of the four specific aims. Chapters 2-5 are formatted as journal articles accompanied by supplemental information available in Appendices 1-4. Chapter 2 [119] and 3 [120] were published in *Environment International*. Finally, Chapter 6 concludes the dissertation by providing an overall discussion of the dissertation results and potential future directions of these results.

1.8 Figures

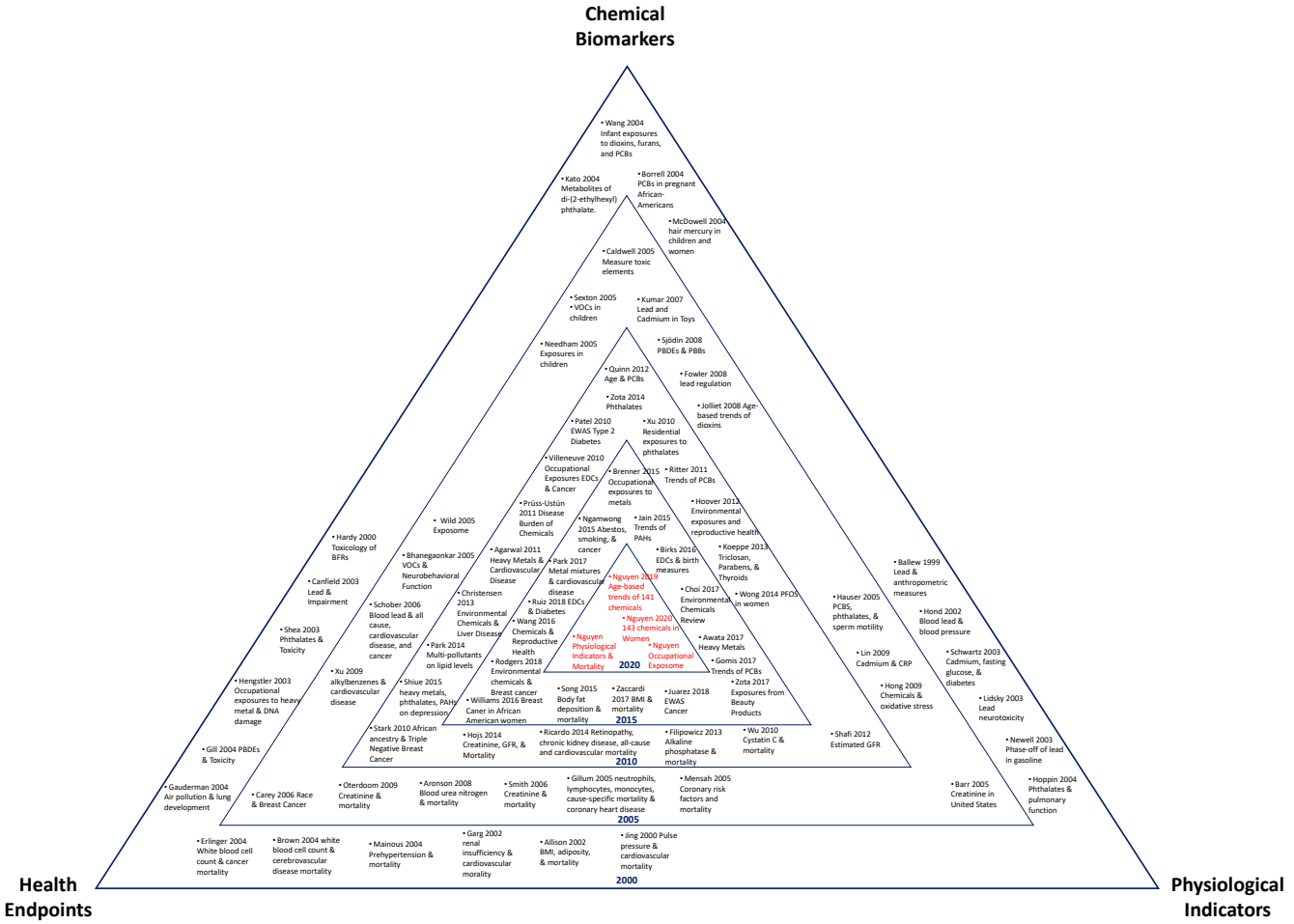

Figure 1.1. General structure of association studies along the spectrum of chemical biomarkers, physiological indicators, and health endpoints. BFRs, Brominated Flame Retardants; BMI, Body Mass Index; CRP, C-Reactive Proteins; GFR, Glomerular Filtration Rate; PAHs, Polycyclic Aromatic Hydrocarbons; PBDEs, Polybrominated Diphenyl Ethers; PBBs, Polybrominated Biphenyls; PCBs, Polychlorinated biphenyls; VOCs, Volatile Organic Compounds.

Chapter 2 Characterization of Age-Based Trends of Chemical Biomarker Levels

2.1 Abstract

Background: Chemical biomarker concentrations are driven by complex interactions between chemical use patterns, exposure pathways, and toxicokinetic parameters such as biological half-lives. Criteria to differentiate legacy from current exposures are helpful for interpreting variation in age-based and time trends of chemical exposure and identifying chemicals to which children are highly exposed. A systematic approach is needed to study temporal trends for a wide range of chemicals in the US population.

Objectives: Using NHANES data on measured biomarker concentrations for 141 chemicals from 1999-2014, we aim to 1) understand the influence of temporal determinants, in particular time trends, biological half-lives, and restriction dates on age-based trends, 2) systematically define an age-based pattern to identify chemicals with ongoing and high exposure in children, and 3) characterize how age-based trends for six Per- and Polyfluoroalkyl Substances (PFASs) are changing over time.

Methods: We performed an integrated analysis of biological half-lives and restriction dates, compared distributions of chemical biomarker concentrations by age group, and then applied a series of regression models to evaluate the linear (β_{age}) and nonlinear (β_{age^2}) relationships between age and chemical biomarker levels.

Results: For restricted chemicals, a minimum persistence of 1 year in the human body is needed to observe substantial differences between less exposed young population and historically

exposed adults. We define a metric ($\beta_{age^2}/\beta_{age} > \frac{1}{26.9}$) that identifies several phthalates, brominated flame retardants, pesticides, and metals such as lead and tungsten to reflect elevated and ongoing exposures in children. While a substantial reduction in children's exposures was reflected in PFOS and PFOA, levels of PFNA and PFHxS in children were higher in 2013-2014 compared to those in 1999-2000.

Conclusions: Integrating a series of regression models with systemized stratified analyses by age group enabled us to define an age-based pattern to identify chemicals that are of higher level in children.

2.2 Introduction

Characterizing an individual's exposome requires understanding their lifelong chemical exposures, including how chemical exposures change over time and by age. Studies using population-level chemical biomonitoring data have observed a variety of chemical-specific time and age trends. Persistent chemicals such as polychlorinated biphenyls (PCBs) tend to show a strong decline over time and differentiated exposure patterns across life stages, which are linked to chemical persistency and changes in legislation [73,121]. Relative to PCB exposures which derive mainly from the diet, characterizing exposures to chemicals in consumer products, such as phthalates, are more complex, since these chemicals are used in a range of products with varying usage patterns. As a result, very different age-based and temporal patterns can be observed even within the same chemical family. For instance, urinary concentrations of mono-ethyl phthalate, mono-n-butyl phthalate, mono-benzyl phthalate, and metabolites of di(2-ethylhexyl) phthalate showed a decline, whereas mono-isobutyl phthalate, mono(3-carboxypropyl) phthalate, mono-carboxyoctyl phthalate, and mono-carboxynonyl phthalate increased from 2001 to 2010, implying that the latter phthalates may be substitutes for the former [28]. Similar trends can be observed in

biomonitoring data for per- and polyfluoroalkyl substances (PFASs), such as perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA), where differences in population concentrations manifested following restrictions in 2000-2002 [122,123]. The exposure patterns of other PFASs from treated consumer products [124], water [125,126], and food contamination [127] are not as well understood and evoke the need to study how age-based trends for these substances are changing over time [76]. While many studies have used biomonitoring data to identify a variety of chemical-specific time and age trends, expanding these analyses to a broader set of chemicals and chemical classes will enable us to understand the drivers behind these age-based trends.

To better understand the relationship between chemical biomarker levels and age, several mechanistic models have been developed to investigate the potential determinants. These models have studied age relationships for specific chemical classes such as PCBs Ritter et al. 2011), dioxins [75], and selected PFASs [76,77], with most considering dietary exposure pathways. These models have enabled the identification of key potential determinants such as biological half-lives, restriction dates, and change of intake with age as important factors in understanding age-based trends. However, such models have mostly been applied to dietary exposures for persistent chemicals and require substantial amount of data on age-based exposure patterns, chemical properties, and chemical usage. Such stipulations make a systematic application across a broad set of chemical classes and exposure pathways complex and challenging. Thus, an overarching statistical approach anchored in biomonitoring data would complement mechanistic approaches by allowing us to screen age-based trends and main determinants across a larger number of chemicals, chemical classes and (even unknown) product usage, to identify subpopulations at risk of high exposure.

Compared to adults, children are particularly susceptible to toxicant exposures due to factors such as higher metabolic rate [52,53], rapid growth, development of organs and tissues [56], and behaviors associated with normal development such as crawling [59], mouthing [60,61], and playing [62].. For example, higher concentrations of polybrominated diphenyl ethers (PBDEs) in younger individuals were attributed to lifestyle and activity differences [65]. Due to their increased susceptibility, it is imperative to identify chemicals to which children are highly exposed. Comparing geometric means of chemical levels across age groups enables the identification of chemicals that are higher in children. Such approaches do not account for confounders, however, nor do they inform the influence of potential determinants on age-based trends [128–130]. There is a need integrate data on biological half-lives and restriction dates with cross-sectional biomonitoring data to understand age patterns and systematically identify ongoing exposures in children.

While progress has been made to characterize temporal trends for a few chemical classes, an overarching screening approach has yet to be developed to systematically study age-based and temporal trends of biomarker data in context with temporal determinants such as half-lives and restriction dates for a wide range of chemicals in the US population. In this study, we therefore applied a systematic approach through a series of regression models to characterize chemical specific age-based patterns and identify highly exposed subpopulations for a broad set of 141 chemical biomarkers from a 1999-2014 sample of the US population. More specifically, our objectives were to 1) understand the influence of temporal determinants on age-based trends, in particular time trends, biological half-lives, and restriction dates, 2) systematically define an age-based pattern of concern to identify chemicals of ongoing and high exposures in the younger population, and 3) conduct a targeted analysis of six PFASs to characterize how age-based trends of these substances are changing over time.

2.3 Material and Methods

The approach integrates four types of data: a large dataset of biomarker concentrations for multiple chemicals in a large sample of the US population, the corresponding demographic factors for the studied population, a dataset of human biological half-lives for the observed chemicals, and a dataset describing the year and type of restrictions imposed on the production, emission, sale or use of products containing these substances, if applicable.

2.3.1 Study Population

Since 1999, the Centers for Disease Control (CDC) has conducted the continuous National Health and Nutrition Examination Survey (NHANES) to collect cross-sectional data on demographic, socioeconomic, dietary, and health-related characteristics in the US population. For this analysis, we combined data from the chemical biomarker and demographic datasets between years 1999-2014 for an initial number of 82,091 participants. We then excluded participants for which corresponding data on chemical biomarkers do not exist (n = 7,149), resulting in a sample size of 74,942 study participants. On a chemical specific basis, we also excluded participants with missing information on any of the following covariates: age, NHANES cycles, sex, race/ethnicity, poverty income ratio, cotinine levels, and urinary creatinine. These exclusion and inclusion criteria are detailed in **Figure 2.1**.

2.3.2 Chemical Biomarker Measurements

We define chemical biomarker as an indicator of environmental exposure that can be measured in blood, serum, or urine. We replaced all measurements below the limit of detection (LOD) with the LOD divided by the square root of 2 [131], as recommended by the CDC to produce reasonably unbiased means and standard deviations [132]. At times, NHANES identified a problem of interference from molybdenum oxide that resulted in corrected concentration of urinary

cadmium recorded as 0 ng/mL (NCHS, 2005a, NCHS, 2005b). Log-transforming such data would be undefined, therefore such measurements were replaced with the LOD divided by the square root of 2 if the participant's urinary cadmium level was under the LOD or otherwise excluded. We calculated detection frequencies for each chemical biomarker and excluded biomarkers with detection frequencies of 50% or less (n = 173). Across the NHANES cycles, improvements in laboratory technology can change the LOD and thus influence changes in detection frequencies by NHANES cycle. To prevent such influence, we calculated detection frequencies by NHANES cycle for each chemical biomarker and excluded measurements that showed drastic changes in the LOD and detection frequencies over time. For instance, percentages of participants with PCB 196 measurements above LODs for Cycle 2 and Cycle 3 are 37.8% and 86.7%, respectively, and the LOD for Cycle 2 and Cycle 3 were 10.50 ng/g and 0.40 ng/g, respectively. As such, measurements from Cycle 2 for PCB 196 were excluded (**Figures A1.1** and **A1.2**). Measurements from given cycles for all PCBs, Dioxins, and Furans along with 2-(N-methyl-PFOSA) acetate, 2,4-D, Paranitrophenol, and 1-pyrene (*n* = 134,453) were therefore also excluded based on these criteria (**Table A1.1**). We also excluded biomarkers that are not indicative of exposure (*n* = 30). We preferred lipid adjusted measurements for biomarkers indicated by 7- or 8-letter NHANES codename ending in "L" or "LA," respectively, for which NHANES provided both lipid-adjusted and non-lipid adjusted measurements, and excluded non-lipid adjusted chemical biomarkers (*n* = 79). Finally, transition from the early to recent NHANES cycles resulted in differences in NHANES chemical codenames, which we corrected to reflect a unique codename for each biomarker (*n* = 22). The final dataset for analysis consisted of 141 chemical biomarkers from 16 different classes.

2.3.3 Half-Lives of Organic and Inorganic Substances in Humans

The biological half-life of a chemical is an important factor to explain differences in chemical biomarker levels across the life-stages [73]. To determine a set of relevant half-lives, we first developed a table of NHANES codenames and corresponding CAS No. for each chemical biomarker (**Table A1.2**). We then matched metabolite biomarkers to their corresponding parent compounds. For biomarkers that are metabolites of several parent compounds, we developed the composite half-life by summing the half-life of the metabolite with the maximum half-life of the corresponding parent substances. This assumes the parent substance or compartment with the highest persistence drives the persistency of the metabolic biomarker. We searched a database of empirically-based whole body elimination half-lives and identified 39 chemicals on the list [133]. For an NHANES chemical biomarker that is a mixture of two substances, i.e. m-/p-Xylene, we applied the average of the substance's half-life. Thirty nine of the 118 organic chemicals in this study have empirically-based whole body elimination half-lives available in the OECD QSAR ToolBox (https://www.qsartoolbox.org/). Since estimated persistency of PFASs showed high variability with estimates up to 220 years, empirically based half-lives were selected from literature for this chemical class (**Text A1.1** and **Table A1.3**). For organic chemicals that are not in the empirical database, the total elimination (intrinsic) half-life was predicted using a screening-level Quantitative Structure-Activity Relationship (QSAR) [133]. The model is a fragment-based QSAR that was developed and validated following OECD QSAR guidance [134,135]. Since these QSARs are only applicable to organic substances, we identified the half-lives of inorganic substances in humans through a review (**Table A1.4**). In selecting literature half-lives, we preferred 1) human half-lives over those from animals, 2) half-lives from animal species that are anatomically similar to humans if human data were not available, and 3) slower elimination kinetics over rapid kinetics.

We selected the maximum half-life for 1) inorganic chemicals that have multiple half-lives for a given biological compartment, and for 2) chemicals with half-lives available for multiple biological compartments, *e.g*. body, bones, blood, or lungs. **Table A1.5** tabulates the methods used to find or estimate half-life for each chemical biomarker.

2.3.4 Restriction Dates

It has been suggested that the time-lapse between a chemical's restriction date and sample collection date is an important contributor to biomarker concentration time trends and age-based differentiations [73]. To investigate this, we developed a database of restriction dates (years) in US commerce through an extensive review (**Table A1.6 and A1.7**). Some chemicals have several reported restriction dates, in particular those that were restricted from different products in different years, such as lead. Note that some chemicals were restricted in certain applications but not in others. For instance, the use of lead was banned in paint [136] and gasoline [137], but it is still used in cosmetic products [138] and plumbing [139,140]. Also, some chemicals have been gradually phased out over several years, such as PFASs. For chemicals with dates recorded as a range, and for which we were unable to determine the relative importance of a given year, we applied the mean year. When there are several dates associated with a chemical biomarker, we applied the latest date to represent the most recent period that the substance was banned or phased out.

2.3.5 Statistical Analysis

We performed all analyses using R version 3.5.1. We first defined 11 different age groups to compare chemical biomarker differences by age and then partitioned the distribution of each chemical biomarker by age group and NHANES cycle. To aid data visualization, such as in **Figure 2.5**, we adjusted concentrations of urinary chemical biomarkers by urinary creatinine levels [141].

For a given biomarker, we used ANOVA to test for differences among geometric means of chemical concentration across age groups.

In NHANES, deliberate oversampling was commonly employed to detect susceptible subpopulations at risk for exposures and/or disease [142]. As such, generalizing the results to the US population requires the application of survey weights to account for the sampling design, but this decreases statistical power in identifying associations within the susceptible and oversampled subpopulations [143]. We applied the survey weights in our statistical models for a few chemical biomarkers and identified minor differences between the weighted and unweighted regression coefficients for age. Due to this minimal influence, survey weights were not included in our statistical analyses.

We used multivariate regression models to evaluate the influence of age and time on the chemical biomarker concentrations in blood and urine after log-transformation of these data. We included log-transformed levels of cotinine as a covariate to represent smoking (Benowitz, 1999), and creatinine levels to adjust for urine dilution and flow differences (Barr et al., 2005). We modeled poverty income ratio (PIR), *i.e.,* the ratio of household income and poverty threshold adjusted for family size and inflation, as a surrogate variable for socioeconomic status. First, we examined the influence of age and time on chemical biomarker concentrations by performing a series of chemical-specific regression models with the main predictors of age centered at $\overline{X_{age}}$ (continuous), survey cycle (continuous), sex (categorical), race/ethnicity (categorical), PIR (continuous), and cotinine (continuous) as described in Equation 1 without the term for age squared:

$$\begin{aligned} log10(X_{Chemical\ Concentrations}) = {} & \beta_{age}(X_{age} - \overline{X_{age}}) + \\ & \beta_{age^2}(X_{age} - \overline{X_{age}})^2 + \\ & \beta_{cycle}X_{cycle} + \end{aligned}$$

$$\beta_{sex}X_{sex} + \beta_{race/ethnicity}X_{race/ethnicity} + \beta_{PIR}X_{PIR} + \beta_{cotinine}X_{cotinine} + \beta_{creatinine}X_{creatinine} + \alpha, \quad [1]$$

where $X_{Chemical\ Concentrations}$ is the log-transformed, unadjusted chemical biomarker concentration for all participants, X_i, where $i \in \{age, age^2, cycle, sex, race/ethnicity, PIR, cotinine, creatinine\}$, is the i covariate for all participants, β_i is the linear regression coefficient for the i covariate, and α is the intercept. For urinary chemical biomarkers, we further corrected the regression models by adjusting for urinary creatinine levels (continuous). For cotinine, the regression models were not corrected for cotinine. Age coefficient (β_{age}^{linear}) and cycle coefficient (β_{cycle}) are interpreted as the change in log-transformed chemical biomarker concentration due to a one-year increase in age or a one-survey-cycle increase in time, respectively. To account for multiple comparisons, we used a False Detection Rate (FDR) method on the *p*-values of the linear regression age-coefficients [103].

To evaluate nonlinear relationships between chemical biomarker levels and age, and systematically identify chemicals that are of higher concentrations in children, we included age centered at $\overline{X_{age}}$ squared as another main predictor as shown in Equation 1. Age was centered at $\overline{X_{age}}$ to reduce the collinearity between the linear and quadratic age predictors to assess the separate contribution of these terms. We denote the age coefficient of the nonlinear regression models as β_{age} to differentiate it from that of the linear models, β_{age}^{linear}. It is interpreted as the change in log-transformed chemical biomarker concentration due to a one-year increase in age. β_{age^2} is interpreted as the change in the slope relationship between chemical concentrations

and age for a one-year increase in age. We assumed a quadratic association between chemical biomarker levels and age to capture how biomarker levels may change non-monotonically across the life stage and for ease of interpretation.

Using β_{age} and β_{age^2}, we defined a metric ($s_{children}$) to rank the chemicals from most concerning to least concerning for children as described in Equation 2:

$$s_{children} = \beta_{age}(X_{age} - \overline{X_{age}}) + \beta_{age^2}(X_{age} - \overline{X_{age}})^2 \quad [2]$$

where X_{age} is designated to 5 years old for this analysis (**Table A1.8**). A more positive $s_{children}$ is indicative of higher chemical biomarker levels in children followed by a downward, convex trend across the older age groups. $10^{s_{children}}$ is interpreted as the fold difference in chemical biomarker levels between a child of 5 years and adult of 31.88 years. To define a boundary line to differentiate the chemicals of higher exposures in children compared to those in adults, we solved for $\beta_{age^2}/\beta_{age}$ when $s_{children} = 0$ from Equation 2 and calculated the slope of the boundary line to be $\frac{1}{\overline{X_{age}} - X_{age}} = \frac{1}{26.9}$. Any chemical with $s_{children} = 0$ implies that biomarker levels are the same between a child of 5 years and adult of 31.88 years. Using the regression coefficients, we predicted the log-transformed chemical biomarker levels for all participants with complete data on age, cycle, sex, race/ethnicity, PIR, and cotinine. Predictions are not available for children between one to two years of age, since measurements for blood cotinine in this age group were missing.

To understand how differences in chemical biomarker concentrations between young and older individuals change over time, *i.e.*, how age-based trends are changing over time, we conducted stratified analyses by NHANES cycle. We first partitioned life-stage changes in chemical biomarker concentrations by NHANES cycles and fitted these cycle-specific concentrations with smooth curves through LOESS (locally weighted scatterplot smoothing) [144]. Then for each cycle with measurements, we performed a chemical-specific linear regression with

age (continuous) as the main predictor while adjusting for sex (categorical), race/ethnicity (categorical), PIR (continuous), and smoking (continuous) described in Equation 3:

$$log10(X_{Chemical\ Concentrations}[Cycle = k]) = \beta_{age,k}X_{age}[Cycle = k] + \beta_{sex,k}X_{sex}[Cycle = k] + \beta_{race/ethnicity,k}X_{race/ethnicity}[Cycle = k] + \beta_{PIR,k}X_{PIR}[Cycle = k] + \beta_{cotinine,k}X_{cotinine}[Cycle = k] + \beta_{creatinine,k}X_{creatinine}[Cycle = k] + \alpha_k, \quad [3]$$

where k is the available cycle number that can range from 1 to 8, $X_{Chemical\ Concentrations}[Cycle = k]$ is the log-transformed, unadjusted chemical biomarker concentrations of participants in the *kth* cycle, $X_m[Cycle = k]$, where $m \in \{age, sex, race/ethnicity, PIR, cotinine, creatinine\}$ is the m covariate for all participants in the *kth* cycle, $\beta_{m,k}$ is the linear regression coefficient for the m covariate in the *kth* cycle, and α_k is the intercept for the *kth* cycle. The linear regression age coefficient ($\beta_{age,k}$) is interpreted as the change in log-transformed chemical biomarker concentration due to a one-year increase in age for a given *kth* cycle.

2.4 Results

2.4.1 Study population

Table 2.1 presents population characteristics for the 74,942 NHANES participants from 1999-2014. The mean age ($\overline{X_{age}}$) is 31.88 (SD 24.28) with approximately 42.1% of the population being 18 years old or younger. This indicates children are oversampled, since according to the US Census, 26% of the US civilian noninstitutionalized population are below 19 years of age. The number of participants across the cycles does not vary drastically. The population is evenly distributed by sex with approximately 51% of the population being female. All race/ethnicity were oversampled, except for Non-Hispanic Whites, since according to the US Census, the proportions

of Hispanics, Non-Hispanic Blacks, and Other Race are 17.8%, 13.3%, and 9.8%, respectively [145]. The mean of PIR is 2.301 (SD 1.59). The means of cotinine and creatinine levels were 38.39 (SD 103.80) ng/mL and 130.3 (SD 81.98) mg/dL, respectively.

2.4.2 Age-Based Trends, Half-Lives, and Restriction Dates

Figure 2.2A shows the number of biomarkers for each chemical class, and **Figure 2.2B** shows the range of log-transformed half-lives for each chemical class with a dashed line representing one year (**Table A1.5**). Chemical classes with half-lives in the range of 1 to 100 hours include Phthalates, Acrylamide, Other, Smoking Related Compounds (SRCs), Phytoestrogens, Polycyclic Aromatic Hydrocarbons (PAHs), Personal Care and Consumer Product Compounds (PCCPCs), Volatile Organic Compounds (VOCs), and Melamine, while classes with more persistent chemicals include Brominated Flame Retardants (BFRs), PFASs, PCBs, Dioxins, and Furans. Chemicals from the Metals and Pesticides classes demonstrate a wide range of persistency in the human body.

Figure 2.2C shows ranges of $\beta_{age}^{linear\prime}s$ for each chemical class. These values are interpreted as the log change in chemical concentration for a one-year increase in age. The majority of chemicals from PCBs, Furans, Dioxins, Melamine, Metals, and Pesticides along with a single BFR (2,2',4,4',5,5'-hexabromobiphenyl) have high positive ranges of $\beta_{age}^{linear\prime}s$, indicating higher concentrations in the older population. In contrast, most of the phthalates, SRCs, and BFRs along with a few VOCs, PCCPCs, PAHs, and phytoestrogens have negative $\beta_{age}^{linear\prime}s$, reflecting higher concentrations in younger individuals. The majority of chemical biomarkers have $\beta_{age}^{linear\prime}s$ between -0.01 and 0.01, suggesting small or no differences in chemical biomarker levels across the life-stages.

Figure 2.2D shows the proportions of unrestricted or restricted chemicals for each class, and **Table A1.7** tabulates the restriction dates. Since the latest data were from 2013-2014, chemicals with restriction dates after 2014 are categorized as having no restriction. Chemical classes with higher proportions of unrestricted chemicals include Acrylamide, Other, SRCs, VOCs, and Melamine, and these have limited $\beta_{age}^{linear}{}'s$. In contrast, PCBs, Dioxins, and Furans show higher proportions of historically restricted chemicals and have the highest $\beta_{age}^{linear}{}'s$ and high half-lives. The majority of BFRs, PCCPCs, and PFASs have been restricted more recently and have limited $\beta_{age}^{linear}{}'s$ despite PFASs having high half-lives. The Metal and Pesticides classes demonstrate a wide variety of restriction types, with most of the persistent chemicals in these classes having been restricted before the turn of the century.

Figure A1.3 and **Text A1.2** further analyze changes in biomarker levels over the NHANES cycles, demonstrating a decrease in chemical biomarker levels over time for the majority of pesticides and PFASs show, while a few pesticides, phthalates, and PAHs have increasing time trends.

2.4.3 Influence of Temporal Determinants on Linear Age-Based Trends

To understand the influence of chemical persistence in the body, time trends, and restriction dates on differences in chemical biomarker concentrations across the life-stages, we examined the association between the $\beta_{age}^{linear}{}'s$ and human whole body elimination half-lives for all chemical biomarkers, color-coded by 1) restriction dates (**Figure 2.3**) and by 2) time trend trajectories (**Figure A1.4**). Chemical biomarkers with half-lives less than one year have $\beta_{age}^{linear}{}'s$ ranging from -0.01 to 0.01, indicating limited variation across life-stages. For these chemicals, cross-sectional biomonitoring data is primarily reflective of present exposures in different age groups or populations [73]. In contrast, chemical biomarkers with half-lives greater than one year demonstrate

more variation across life-stages and show a positive association between the $\beta_{age}^{linear\prime s}$ and half-lives. The majority of these persistent chemicals were banned or phased out between the 1970s and 1999 (blue markers in **Figure 2.3**). This implies exposures of the younger population have been strongly reduced, and that higher concentrations observed in the older population are likely due to historical exposures and long biological half-lives. Despite the long half-lives of BFRs and PFASs, $\beta_{age}^{linear\prime s}$ of these chemical classes are substantially lower than those of other persistent substances with similar half-lives. The lower $\beta_{age}^{linear\prime s}$ with age may be explained by the fact that these chemicals have been recently restricted or are still in use (red and yellow markers in **Figure 2.3**) and that current exposures remain higher than exposures to legacy pollutants that were banned earlier. Of special concern are chemicals with negative $\beta_{age}^{linear\prime s}$, since these chemicals are of higher levels in the younger population compared to the aged population. Most of these chemicals are unrestricted (red markers in **Figure 2.3**) and demonstrate an increasing or stable time trend (red and orange markers in **Figure A1.4**).

2.4.4 Nonlinear Age-Based Pattern of Higher Levels in Children

Since a linear relationship between age and log-transformed biomarker levels may not be representative for chemicals that display a nonlinear relationship with age, we refined the chemical-specific regression models to have age squared centered at X_{age} as another main predictor to better characterize this relationship. **Figure 2.4** summarizes the results for age from the quadratic regression model by presenting the association between β_{age^2} and β_{age} for all chemical biomarkers. The chemical classes are indicated by difference shapes, while the colors show the different categories of fold difference in chemical biomarker levels between a child of 5 years and adult of 31.88 years. For instance, mono-benzyl phthalate levels in 5-years-old children are on average 2.598 times higher compared to those for 31.88-years-old adults. A positive β_{age^2}

indicates a convex (or u-shaped) relationship between log-transformed chemical biomarker levels and age, while a negative β_{age^2} indicates a concave or (n-shaped) relationship.

Chemicals in the upper left quadrant are of interest, since these are higher in 5-years-old children compared to 31.88-years-old adults by more than a factor of 2. Most of these chemicals are metals, pesticides, and phthalates used in building materials and articles. Based on Equation 2, the boundary line corresponds to equal biomarker levels for a child of 5 years and an adult of 31.88 years. Chemicals above and to the left of the boundary line ($\beta_{age^2}/\beta_{age} > \frac{1}{\overline{X_{age}}-X_{age}} = \frac{1}{26.9}$) have a downward and convex trend across age groups, implying the highest biomarker levels for the youngest participants. The highest levels in children compared to adults of average age are observed for mono-benzyl phthalate, O-Desmethylangolensin (O-DMA), mono-(3-carboxypropyl) phthalate, 2-amnothiazolne-4-carbxylic acid and tungsten. **Table A1.8** provides a detailed list of chemicals ranked from highest to lowest relative value between children and adults of average age.

To further compare chemical biomarker distributions across the different life-stages and identify linear and nonlinear age-based trends, we stratified these distributions into 11 age groups (**Figure 2.5**) and selected example chemicals to represent the specific age-group trends within a chemical class. The geometric mean of the measured chemical biomarker levels for each age group is represented by a gray circle, while the geometric mean of the predicted chemical biomarker levels is indicated by a brown triangle. Outliers are represented by dash marks outside of the distributions of chemical biomarker concentration. A residual standard error (RSE) of 0 implies the model perfectly predicts the log-transformed biomarker levels. Geometric mean of predicted chemical biomarker levels for the 1-2 age group is unavailable, since cotinine was not measured in these participants.

Overall, the nonlinear regression models predicted the measured geometric means fairly well, particularly for children, middle-aged adults, and the elderly. The models, however, overestimated biomarker levels for tungsten, phthalates, and parabens in the adolescent age group, and underestimated lead in the toddler age group, indicating the need for a higher order polynomial model rather than a parabolic regression model for these specific cases.

The following section analyzes in further detail these age trends by chemical class and type of usage.

2.4.5 Age-Based Trends by Chemical Class

2.4.5.1 PCBs, Dioxins, and Furans

For PCBs, there are three main age patterns: a slight downward and convex trend, a steep upward and concave trend, and no trend across the life stage (**Figure 2.4**). The PCBs with half-lives less than one year (**Figure 2.3**) showed little or no variation in chemical biomarker concentrations by age. PCB 49 (**Figure 2.5A**) and PCB 44 are the only two PCBs for which the youngest participants have the highest biomarker levels, with their negative β_{age} and positive β_{age^2} characterizing a slight downward and convex trend across the age groups. This might indicate children are exposed through a pathway specific to these two congeners. In contrast, the more persistent PCBs, dioxins and furans have higher concentrations in the older population, except for 1,2,3,4,6,7,8-Heptachlorodibenzofuran (half-life of 3.58 years), which has a β_{age} of -0.0021. This may indicate ongoing exposure despite this chemical having been banned much earlier. With the highest β_{age} of 0.037 and a β_{age^2} of -0.00033, PCB 194 illustrated well (**Figure 2.5B**) a steep upward and concave trend across the age groups with the oldest participants having at most a 100-fold difference in biomarker levels compared to the youngest age group. This tendency is confirmed in the pooled serum concentrations observed for four age groups (12-19,

20-39, 40-59, 60+) in 2005-2008 by different race-sex combinations (details in Text S3, Section 4). Age is a good predictor of biomarker levels for the more persistent PCBs, with an adjusted correlation coefficient (R^2) ranging between 0.37 and 0.72 for most PCBs, with the exception of PCB 28 ($R^2 = 0.035$), PCB 44 ($R^2 = 0.036$), and PCB 49 ($R^2 = 0.041$).

2.4.5.2 PFASs

PFASs are also highly persistent, but their $\beta_{age}{}'s$ do not vary as substantially as those of PCBs. Most of the PFASs have $\beta_{age}{}'s$ close to 0, indicating there is little to no difference by age and implying ongoing exposures. PFNA shows little to no variation across age groups (**Figure 2.5C**). On the other hand, PFOA and PFOS show a slight upward and convex trend across the age groups (**Figure A1.5** and **2.5D**). This is confirmed by a β_{age} of 0.0029 and a β_{age^2} of 3.27E-05 for PFOS, and by a β_{age} of -5.80E-05 and a β_{age^2} of 2.97E-05 for PFOA. Since PFOS and PFOA were phased out in 2002 [146–148], differences across the age groups are substantially smaller than those observed in PCB 194, which was banned in 1979. Such differences across the life-course, however, are expected to increase in the future as articles and materials containing PFASs will reach the end of their usable life. A specific trend analysis is presented in the next section for these PFASs to illustrate how the age-based trends vary across the different cycles.

2.4.5.3 Metals

Another class of highly persistent chemicals is the Metals. Although many of the metals demonstrate a stable trajectory over time (yellow markers in **Figure A1.4**), there are high variations in chemical biomarker levels across the life-stages, with three different types of age group patterns evident (**Figure 2.3** and **2.4**). Cadmium demonstrates higher urinary concentrations in the older population with a β_{age} of 0.014 and a β_{age^2} of -9.22E-05, denoting a slight upward and concave trend across the age groups (**Figure A1.6**). Lead is one of the few chemicals with

measurements in children 1 to 4 years old. Although the β_{age} of lead (0.0039) is not as high as that of cadmium, the convex trend of lead (β_{age^2} = 8.72E-05) across the age groups indicates the youngest and oldest age groups have the highest biomarker concentrations compared to the other age groups (**Figure 2.5F**). Although tungsten has similar persistency to cadmium and lead, it has a β_{age} of -0.0069 and a β_{age^2} of 0.00015, indicating a downward, convex trend across the age groups (**Figure 2.5E**). This is also indicative of high and ongoing exposures in the younger population.

2.4.5.4 Phthalate and Parabens

Most phthalates are used as plasticizers. These phthalates show a similar age group pattern to that of mono-(3-carboxypropyl) phthalate - a metabolite of mono-n-butyl phthalate, di-n-butyl phthalate, mono-n-octyl phthalate, and di-n-octyl phthalate (**Figure 2.5G**), and mono-benzyl phthalate (**Figure 2.5H**), with the highest concentration apparent in the youngest age group, a decrease during adolescence and young adulthood, and then stabilization for older age groups. In contrast, mono-ethyl phthalate is mostly used in cosmetics and demonstrates a very different age group pattern from those in its chemical family (**Figure 2.5I**). It has a similar age group pattern to chemicals used in cosmetics such as methyl paraben (**Figure 2.5J**). Methyl paraben has a slight upward and concave trend across the 5-12, to 13-18, and 19-28 years-old participants. Its levels peak for the mature adults and show a slight decrease in older age groups.

2.4.6 Change in Age-Based Trends of PFASs over Time

To determine how the age-based trends are changing over time, we fitted smooth curves to the life-stage changes in chemical biomarker concentration for each available NHANES cycles and conducted a series of linear regression models stratified by cycle to extract the $\beta_{age,k}$*'s*. The $\beta_{age,k}$*'s* shows the overall difference in chemical biomarker concentration between the young and

aged populations for a given NHANES *kth* cycle. Understanding how these $\beta_{age,k}$*'s* change over time provides insight on how the difference between the youth and elderly changes across the cycles. In addition, these $\beta_{age,k}$*'s* will help determine how long a time lapse must occur between the restriction date and sample collection date in order to observe these life-stage differences.

PFASs were further analyzed, since some have been recently phased out and have measurements spanning over six or more cycles. PFOS, Perfluorohexane sulfonic acid (PFHxS), PFOA, Perfluorodecanoic acid (PFDA), PFNA, and 2-(N-methyl-PFOSA) acetate were selected due to their high detection frequencies for each available cycle. **Figure 2.6** presents how the age-based trends are changing over time. Each curve represents the variation in chemical biomarker concentration by age for a given NHANES cycle. A vertical shift in an *ith* cycle curve indicates how the chemical biomarker concentrations have increased or decreased compared to those in the *(i-1)th* cycle. The steepness of the curve shows the rate at which the log-transformed chemical concentration is changing with each one-year increase in age for a given cycle, providing insight on how the differences in chemical biomarker levels between the youth and elderly are changing over time. An increase in the $\beta_{age,k}$*'s* indicates the difference between the young and aged populations are expanding over time.

Between 1999-2000, the $\beta_{age,1}$ for PFOS is 6.51E-4 (p-value = 0.025). If we assume chemical biomarker concentrations change linearly with age, then this value implies a 1.13-fold difference ($10^{80\times 6.51E\text{-}4} = 1.13$) in chemical concentration between an 80-year old participant and a newborn (aged 0). Between 2013-2014, the $\beta_{age,8}$ is 6.49E-3 (p-value = 2.63E-59) with a 3.3-fold difference. This suggests that a decade after the phase-out of PFOS, the aged population has approximately a 3-fold difference in PFOS levels compared to the youth (**Figure 2.6A** and **2.6G**). As the biomarker levels of PFOS and 2-(N-methyl-PFOSA) acetate decrease across the cycles,

illustrated by the downward shifts in the concentration-age curves, the difference between the youth and elderly increases. This is evidenced by the increasing steepness of these curves (**Figure 2.6A** and **2.6F**) and the upward trend in $\beta_{age,k}$*'s* (**Figure 2.6G**). These patterns imply the use of PFOSA stopped around the time of the restrictions on PFOS and PFOA in 2002. A similar pattern can be observed with PFHxS and PFOA concentrations, but the differences between the young and aged populations for these chemicals do not change as drastically they do for PFOS and PFOSA (**Figure 2.6B, 2.6C,** and **2.6G**). These patterns suggest that as the time increases between the restriction and sample collection dates, the differences by age will become more prominent.

On the other hand, PFHxS, PFDA, and PFNA display different trends over time. For instance, biomarker levels for PFHxS initially decrease during 1999-2006, increase in 2007-2008, and then decrease again for the more recent NHANES cycles. For PFDA, biomarker levels increase from 1999 to 2006 and then decrease afterward. Biomarker levels of PFNA increase during the early NHANES cycles and then decrease after 2009-2010, but the PFNA levels for 2013-2014 are on average higher than those between 1999-2000 especially for children. The cycle-specific age coefficients for PFNA fluctuate during the early NHANES cycles but then show a strictly increasing trend after 2007-2008 (**Figure 2.6E** and **2.6G**). These fluctuations in biomarker levels suggest PFHxS, PFDA, and PFNA may have been used as substitutes for PFOS and PFOA and reflect ongoing exposure throughout the population.

2.5 Discussions

In this article, we present a comprehensive analysis of age-based and time trends in chemical biomarker concentrations in the US population. We have accounted for biological half-lives of chemicals, type of usage, and historical events, *i.e.,* dates of chemical bans and phase-outs, which are expected to influence population-level exposures. These results provide insight on

population exposure trajectories. They are also informative for differentiating legacy exposures from current exposures and for identifying chemicals of higher levels in the younger population.

For restricted chemicals, our data confirm that a minimum persistence of 1 year in the human body is necessary to observe substantial differences between the young population and historically exposed adults. Biological half-life is not the only determinant of high chemical biomarker levels in the aged population, however. Studies on age-based and time trends of biomonitoring data have suggested the potential influences of bans, phase-outs, bioaccumulation, metabolic rates, and consumer product usage on such trends [28,65,121–123], with the most influential determinant for simulated longitudinal data being the time lapse between the peak of emission and the sample collection [73]. Thus, elevated concentrations in the elderly population are primarily due to a combination of past exposure and slow elimination. Using measured biomonitoring data for a wide range of chemicals, we confirm that chemicals with high age coefficients primarily have a biological half-life longer than 1 year and have been banned or phased out for longer than the chemical's half-life. This evidences the efficacy of public health interventions, such as the International Stockholm Convention on Persistent Organic Pollutant, to reduce or prevent high exposures and associated health outcomes for the younger population effects [71,72].

PFASs are also persistent, with half-lives ranging from 1.6 years to 7.3 years, yet show minimal differences in biomarker concentrations across the life-stages. These substances demonstrate contrasting age-based patterns even within the same family. For instance, we observed a substantial reduction in PFOA and PFOS levels in children, but levels of PFNA and PFHxS in children during 2013-2014 are still higher or equal than those in the earlier NHANES cycles. This indicates ongoing and higher exposures for the younger population. Such exposures to PFASs may occur through breastfeeding [149–151] or drinking contaminated water [125,126]. In

addition, this pattern could be due to the fact that some of these chemicals were recently phased-out, or due to the short time lapse between the emission peak and the sample collection. The time lapse of a decade for PFASs is shorter than the time lapse of almost 30 years for PCBs. Thus, it can be inferred that as this time lapse increases, especially if it exceeds the half-life of the substance, the difference in PFASs concentrations by age will continue to increase (Quinn and Wania 2012, Gomis et al. 2017).

While cadmium levels are lower in the younger population, this is not the case for lead and especially tungsten for which the younger population has surprisingly higher biomarker levels. Higher lead levels have been attributed to consumer products usage, such as toys and children jewelry [62,152], exposures to dust and soil [153,154], and exposures via maternal transfer in utero or during breastfeeding [155,156]. For the older participants, high lead levels may be due to leaded gasoline combustion before tetraethyl lead in gasoline was banned [137]. High tungsten levels in children may be due to exposures to contaminated soil, articles from parents' workspace, and electrical devices [157,158] (ATSDR 2005; Kampmann et al. 2002). The overall trend of higher levels in the 5-12 years old followed by a downward, convex trend across the older age groups suggests exposures in children may be driven by factors specific to this susceptible population. Hence, further research is necessary to elucidate potential reasons for higher exposures in children.

For several less-persistent chemicals, such as phthalates that are widely used in consumer products, younger individuals seem to have been highly exposed, in addition to some persistent chemicals such as the BFRs, PFASs and lead. Our results suggest age-based trends in biomarker levels reflect product usage trends. Most phthalates show a plasticizer age pattern with higher concentrations in the 5-12 years old age group followed by a downward, convex trend across older age groups, which is quantified by a positive β_{age^2} and negative β_{age}. Children may be highly

exposed to these chemicals through frequent contact with flooring materials [59,61,159] and toys [160], which are products that typically have high levels of plasticizers. Also, children may more readily absorb these compounds [161]. Metabolic rate is known to vary across age with a peak occurring during childhood and then stabilizing or decreasing during the senior years [52,53]. In contrast, mono-ethyl phthalate is used in personal care products [162] and shows a different concave pattern similar to other personal care products such as parabens and triclosan. The increase in exposure from children to teenagers may be explained by a greater use of cosmetic and/or skin care products during the teenage years [163–165]. Comparing chemical levels by age group and quantifying trends across the age groups with β_{age^2} and β_{age} enables us to identify two interesting clusters: 1) a cluster of phthalates used as plasticizers and 2) a cluster of chemicals used in personal care products. Mono-ethyl phthalate was shown to cluster with PCCPCs instead of with those in its chemical family. These age-based clusters of chemicals with similar product usage suggest a possibility to develop product-specific archetypes of intake pattern with *e.g.* a concave age curve for personal care products versus a convex age curve for plasticizers in articles and building materials. These archetypes could then be used to help extend mechanistic modelling approaches to predict direct exposures to chemicals used in consumer products.

The present study has a number of limitations. By comparing chemical biomarker levels by age group, we have identified several chemicals, such as lead, tungsten, and phthalates, to be of higher concentrations in the younger population than in the older population. Although we have identified a number of potential reasons for higher exposures in children, we have not accounted for differences in metabolic rate within our models. Future extensions could determine surrogate variables to develop a scoring system to quantitatively represent metabolic rate and understand how it could confound age and chemical biomarker concentrations. As another limitation, we

assumed a quadratic relationship between age and chemical biomarker levels, but we did not test for other higher order polynomials that may better describe the changes in chemical biomarker levels across the life stage. Thus, future studies can compare the cross-validated prediction performance between a quadratic and cubic polynomial and/or even other higher order polynomials to determine whether a higher order polynomial would substantially improve the fit of the model to the data. Finally, while we demonstrated an overarching, statistical approach to identify chemicals that are of higher concentrations in children, there is a need to understand toxicological effects of these chemicals along with identifying sources and pathways of exposures to prevent elevated chemical levels and the onset of adverse health effects [71,72,166].

Though we defined an age pattern of concern for children, quantifying exposure for young children, especially those below the age of 4, was limited to a few chemical biomarkers such as lead, manganese, cadmium, methyl mercury, cotinine, acrylamide, and glycideamide. Due to the ban of legacy chemicals such as PCBs, Furans, and Dioxins, fewer measurements are available in children, and children have substantially lower levels of these persistent substances. This creates challenges in understanding sources of exposures for these legacy chemicals in children. Thus, future studies can use the 2001-2016 biomarker data of PCBs for pooled samples to characterize differences by age group to gain more insights on sources of exposures in children. However, a challenge of using pooled data is that demographic and questionnaire data are limited. As shown with lead, predictions were unavailable for children below the age of 2, since cotinine was not measured for these participants. Thus, when more measurements for children become available, future extensions could incorporate such data to better quantify and predict exposure for this susceptible population as well as understand sources of such exposures.

Geographical location has been identified as a confounder of chemical exposure disparities, particularly for heavy metals, but we did not consider this as a covariate in this study. Future studies could consider geospatial variations in chemical biomarker concentrations to systematically address geographical location as a confounder.

For lipophilic chemicals, we preferred the lipid-adjusted measurements, since these measurements were normalized to the blood lipid content of the participants. In addition, adipose content tends to increase as a person age, which can potentially lead to higher concentrations of more lipophilic compounds in the aged population. Even though BMI could modulate the concentration-age associations, we did not consider it as a covariate, since we wanted to study the BMI mediated effect of age on chemical biomarker levels. Future extensions could further explore the confounding nature of BMI on age-based and time trends of chemical biomarker levels.

2.6 Conclusions

This study presents a framework for systematically analyzing and interpreting biomonitoring data, to better understand chemical biomarker differences across the life-stages. We suggest different criteria for determining which chemicals are reflective of legacy exposures vs. current exposures and identify an age pattern of concern when longitudinal data are unavailable or incomplete. We confirm the criteria indicative of legacy exposure as follows: 1) biological half-life of at least one year, 2) decreasing average biomarker concentration over time due to the chemical being banned or phased out, and 3) the time lapse between emission peak and the sample collection exceeding the human elimination half-life. For chemicals below the one-year half-life mark, cross-sectional biomonitoring data mostly reflect recent intake rates. In addition to confirming the criteria for legacy versus relevant exposures, the complementary analysis combining a series of regression models with systemized stratified analyses by age group helped

us define an age-based pattern for identifying chemicals of higher and ongoing exposures in children. This is especially evident when a chemical biomarker has an increasing or stable time trajectory, demonstrates a convex relationship with age, and is of higher concentration in the younger population. The presented framework can be used to help facilitate risk stratification and guide targeted interventions.

2.7 Figures

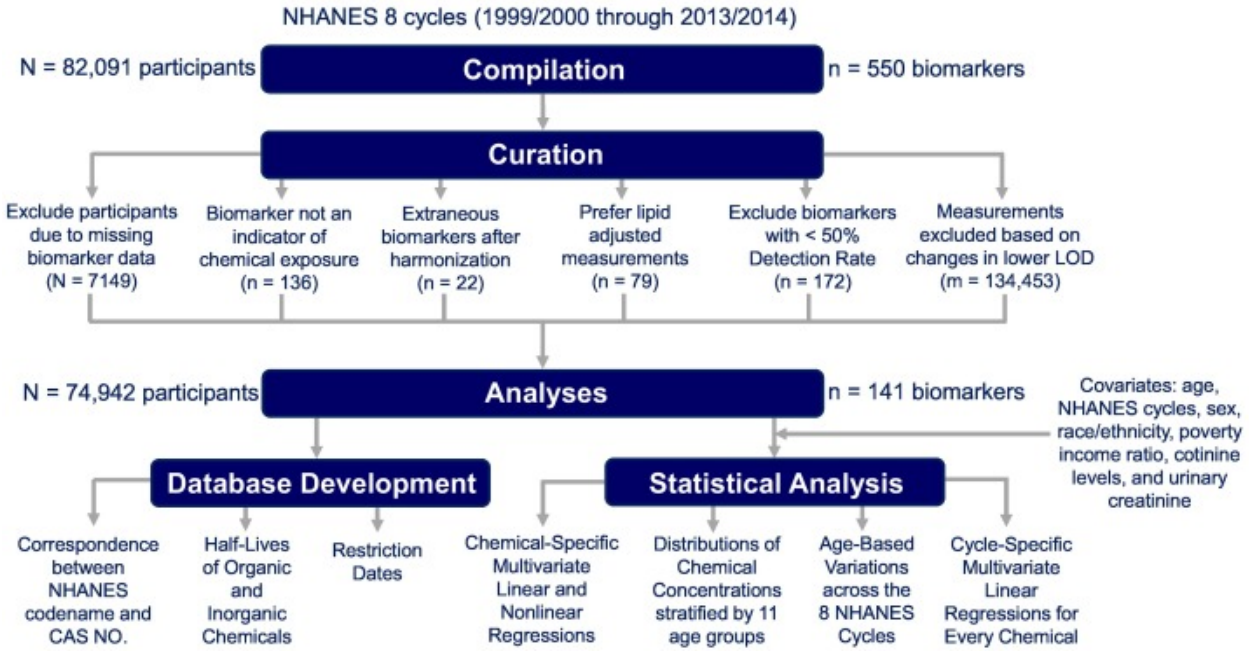

Figure 2.1. Schematic description of the process to curate chemical biomarker measurements and of the analytical methods used to identify temporal variations in biomarker levels.

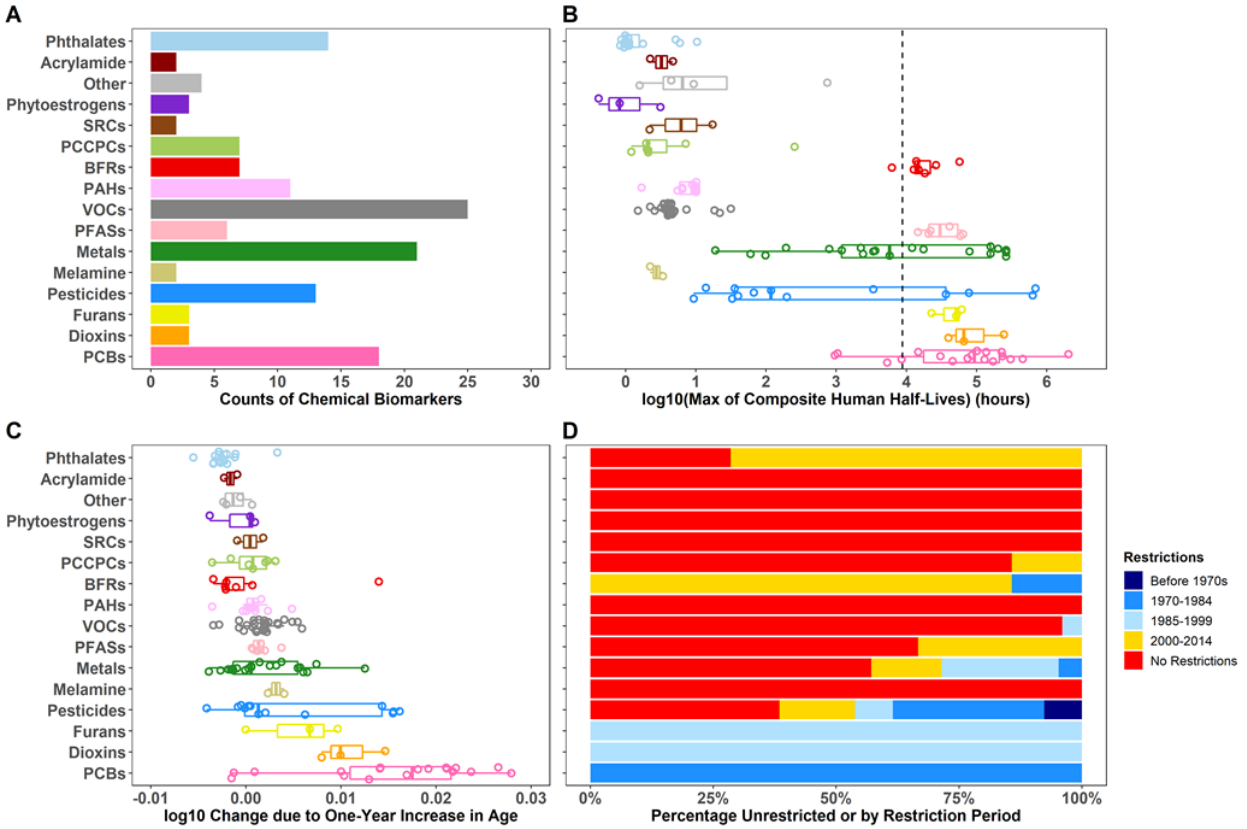

Figure 2.2. Characteristics of the 141 NHANES chemical biomarkers for 16 classes, including (A) the number of chemical biomarkers for each colored-specific chemical class, (B) ranges of log-transformed composite half-lives in hours, (C) ranges of linear age coefficients ($\beta_{age}^{linear'}$ s), defined as the log change in chemical concentration due to a one-year increase in age, and (D) percentage of unrestricted or restricted chemicals per class. Colors of the restriction types only applied to (D) and are also used in Figure 2.3. BFRs, Brominated Flame Retardants; SRCs, Smoking Related Compounds; PAHs, Polycyclic Aromatic Hydrocarbons; PCCPCs, Personal Care and Consumer Product Compounds; VOCs, Volatile Organic Compounds; PFASs, Per- and Polyfluoroalkyl substances; PCBs, Polychlorinated Biphenyls. Models were adjusted for age centered at ($\overline{X_{age}}$), survey cycle, sex, race/ethnicity, poverty income ratio, blood cotinine concentrations, and urinary creatinine concentrations.

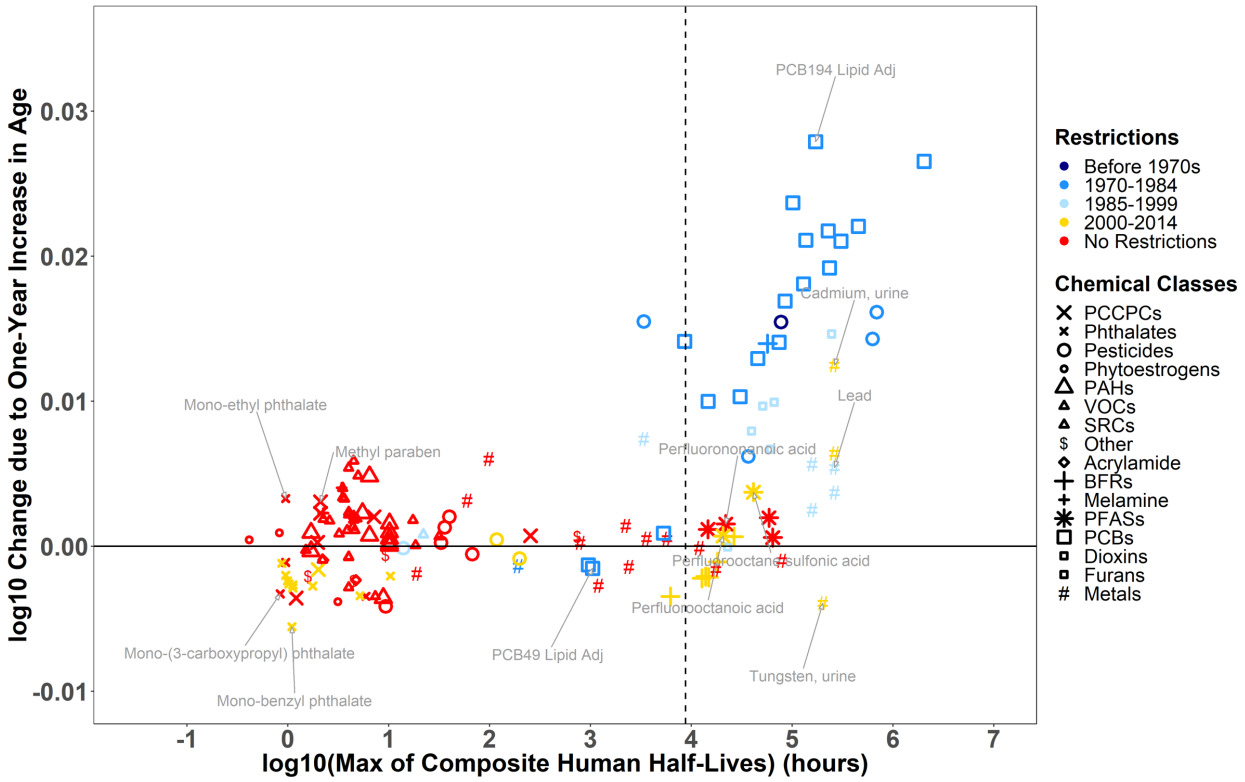

Figure 2.3. Association between linear age coefficients ($\beta_{age}^{linear's}$) and chemical persistency in the human body for 141 substances with symbols indicating the different chemical classes. The colors indicate the time period during which the compound was restricted (same as Figure 2.2D). Models are adjusted for age centered at $\overline{X_{age}}$, survey cycle, sex, race/ethnicity, poverty income ratio, blood cotinine concentrations, and urinary creatinine concentrations. See Figure 2.2 for abbreviations.

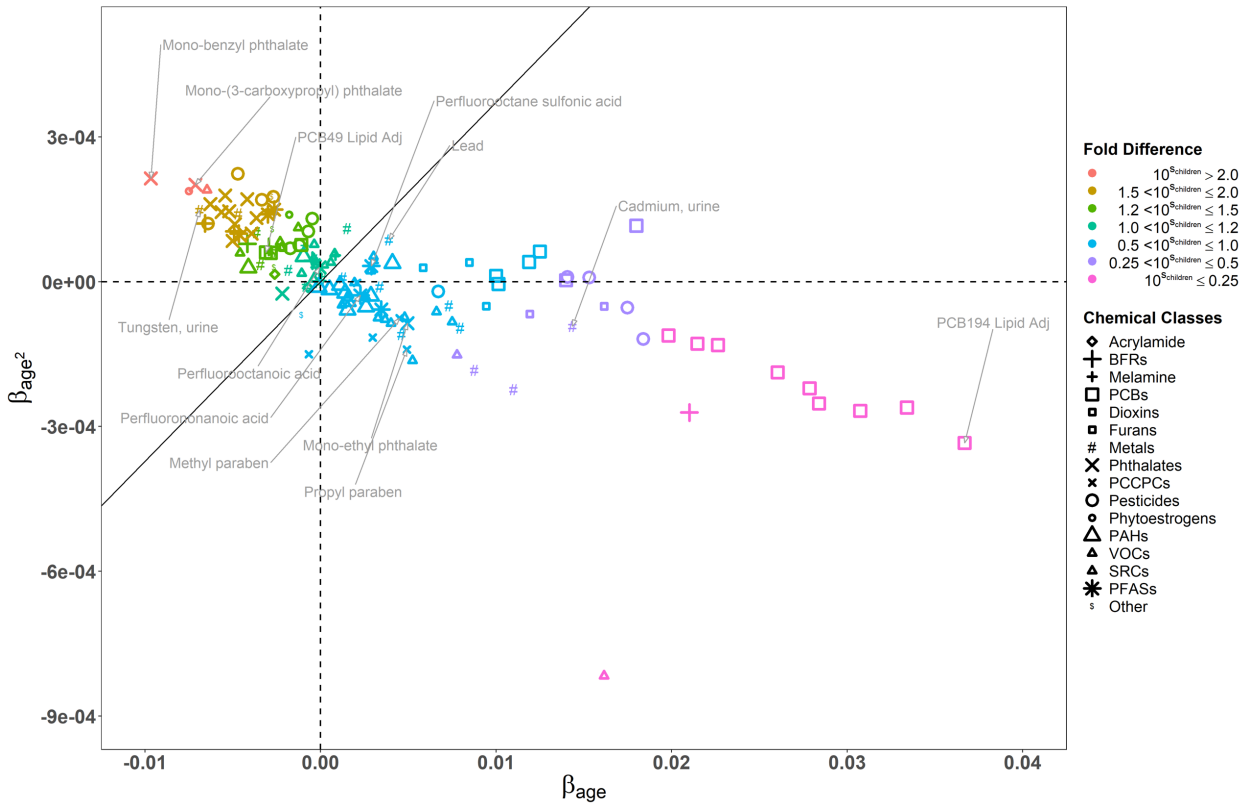

Figure 2.4. Association between β_{age^2} and β_{age} for 141 substances with symbols indicating chemical classes and colors indicating categories of fold difference in biomarker levels between a child of 5 years and adult of 31.88 years. The boundary line $\beta_{age^2}/\beta_{age} > 1/26.9$ differentiates chemicals of higher levels in children from those of higher levels in the older population. Models were adjusted for age centered at $\overline{X_{age}}$, age centered at $\overline{X_{age}}$ squared, survey cycle, sex, race/ethnicity, poverty income ratio, blood cotinine concentrations, and urinary creatinine concentrations. See Figure 2.2 for abbreviations.

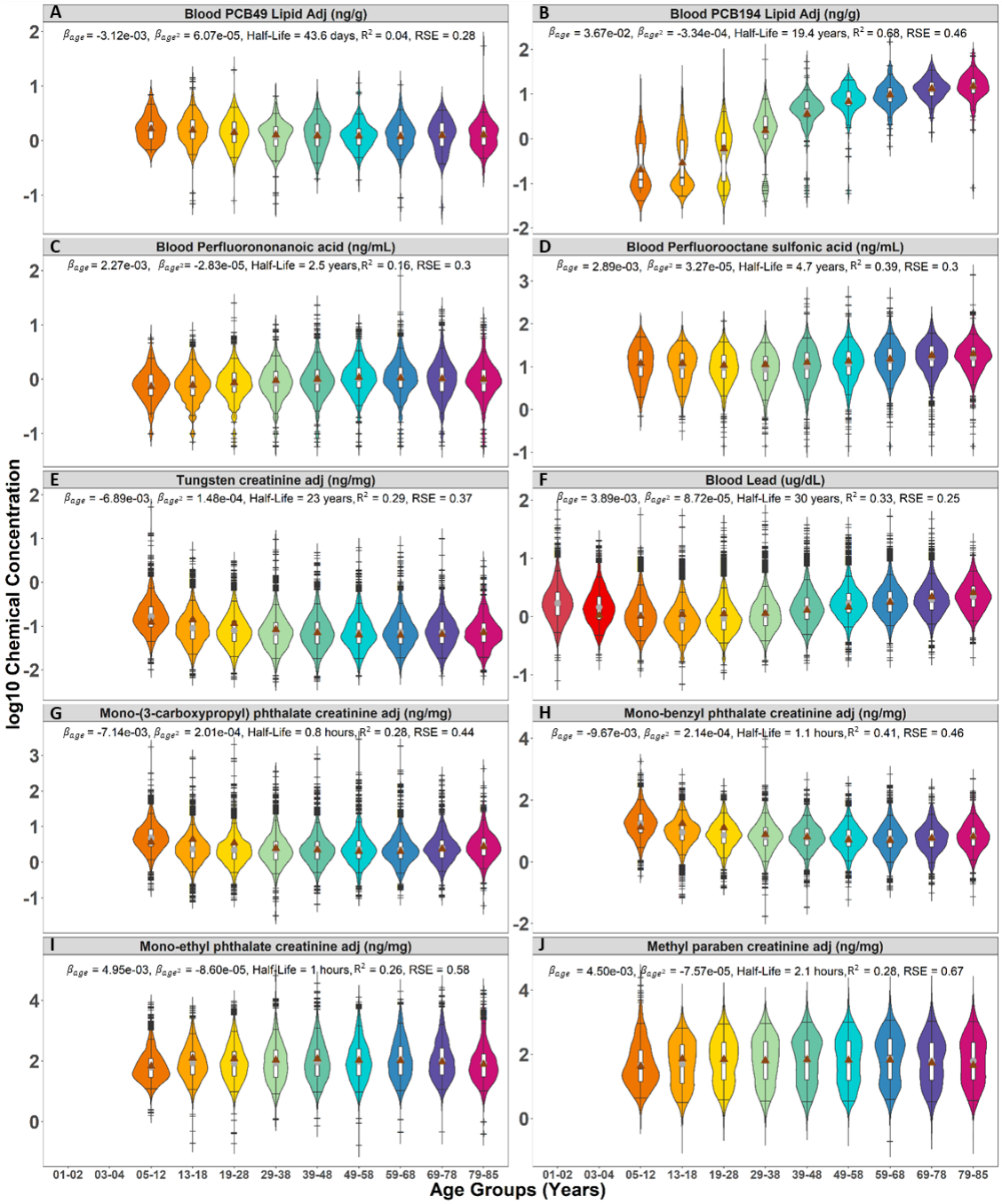

Figure 2.5. Violin plots of chemical biomarker concentrations partitioned by age group to display the 5th, 25th, 50th, 75th, and 95th percentiles, indicated by the superimposed boxplot. The frequency of chemical biomarker levels are represented by the width of the violins for (A) PCB 49, (B) PCB 194, (C) PFNA, (D) PFOS, (E) Tungsten, (F) Lead, (G) Mono-(3-carboxypropyl) phthalate, (H) Mono-benzyl phthalate, (I) Mono-ethyl phthalate, and (J) Methyl paraben. (●) geometric mean of measured data. (▲) geometric mean of predicted chemical biomarker levels. Colors differentiate age groups. Models were adjusted for age centered at $\overline{X_{age}}$, age centered at $\overline{X_{age}}$ squared, survey cycle, sex, race/ethnicity, poverty income ratio, blood cotinine concentrations, and urinary creatinine concentrations.

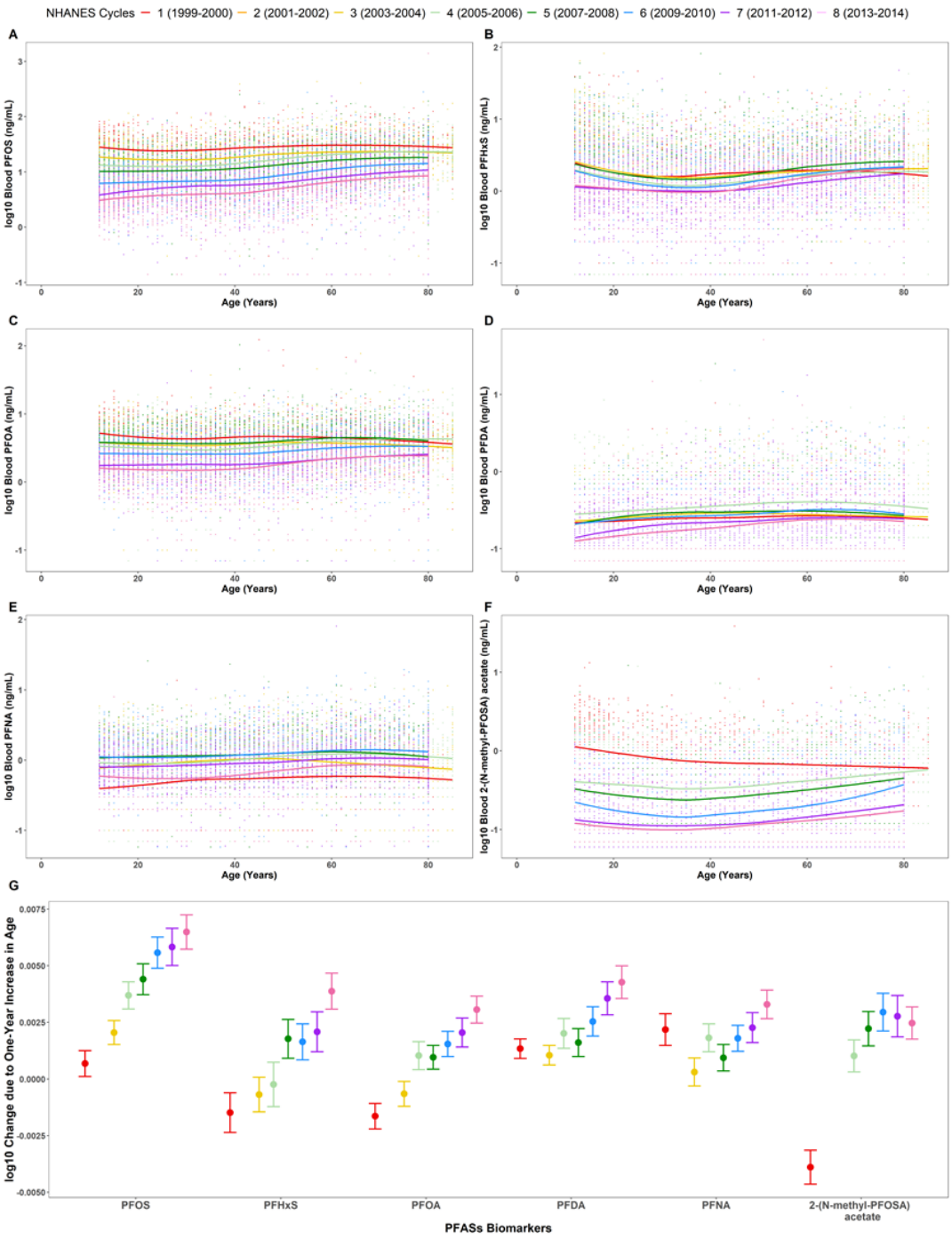

Figure 2.6. Chemical biomarker concentrations across the life-stages stratified by NHANES cycles for (A) PFOS, (B) PFHxS, (C) PFOA, (D) PFDA, (E) PFNA, and (F) 2-(N-methyl-PFOSA) acetate. (G) 95% confidence intervals for the cycle-specific age coefficients for PFOS, PFHxS, PFOA, PFDA, PFNA, and 2-(N-methyl-PFOSA) acetate. The cycle-specific age coefficients (βage,k's) with age shows the adjusted rate at which the chemical concentration is changing for a one-year increase in age for a particular cycle. Models were adjusted for age, sex, race/ethnicity, poverty income ratio, blood cotinine concentrations, and urinary creatinine concentrations.

2.8 Tables

Table 2.1. Characteristics of the study population of 74,942 participants.

CATEGORICAL					
Age Groups	N (%)	Cycle	N (%)	Sex	N (%)
1-2	4714 (6.29)	1999-2000 (Cycle 1)	8832 (11.79)	Male	36941 (49.29)
3-4	3307 (4.41)	2001-2002 (Cycle 2)	9929 (13.25)	Female	38001 (50.71)
5-12	12741 (17.01)	2003-2004 (Cycle 3)	9179 (12.25)	Race/Ethnicity	
13-18	10793 (14.40)	2005-2006 (Cycle 4)	9440 (12.60)	Mexican Americans	17199 (23.95)
19-28	8391 (11.20)	2007-2008 (Cycle 5)	9307 (12.42)	Other Hispanics	5580 (7.45)
29-38	7129 (9.51)	2009-2010 (Cycle 6)	9835 (13.12)	Non-Hispanic Whites	28555 (38.10)
39-48	7168 (9.56)	2011-2012 (Cycle 7)	8956 (11.95)	Non-Hispanic Blacks	18055 (24.09)
49-58	6209 (8.29)	2013-2014 (Cycle 8)	9464 (12.62)	Other Races	5553 (7.41)
58-68	6528 (8.71)				
69-78	4676 (6.24)				
79-85	3286 (4.38)				

CONTINUOUS					
	N (%)	5th	Median	Mean (SD)	95th
Age (years)		2	25	31.88 (24.28)	77
PIR (-)	68192 (90.99)	0.30	1.82	2.301 (1.59)	5.00
Cotinine (ng/mL)	54513 (72.74)	0.011	0.066	38.39 (103.80)	282.00
Creatinine (mg/dL)	63457 (84.67)	26	116	130.3 (81.98)	284

Chapter 3 Racial Disparities in Chemical Biomarker Concentrations in United States Women

3.1 Abstract

Background: Stark racial disparities in disease incidence among American women remain a persistent public health challenge. These disparities likely result from complex interactions between genetic, social, lifestyle, and environmental risk factors. The influence of environmental risk factors, such as chemical exposure, however, may be substantial and is poorly understood.

Objectives: We quantitatively evaluated chemical-exposure disparities by race/ethnicity, life stage, and time in United States (US) women (n=38,080) by using biomarker data for 143 chemicals from the National Health and Nutrition Examination Survey (NHANES) 1999-2014.

Methods: We applied a series of survey-weighted, generalized linear models using data from the entire NHANES women population along with cycle and age-group stratified subpopulations. The outcome was chemical biomarker concentration, and the main predictor was race/ethnicity with adjustment for age, socioeconomic status, smoking habits, and NHANES cycle.

Results: The highest disparities across non-Hispanic Black, Mexican American, Other Hispanic, and Other Race/Multi-Racial women were observed for pesticides and their metabolites, including 2,5-dichlorophenol, o,p'-DDE, beta-hexachlorocyclohexane, and 2,4-dichlorophenol, along with personal care and consumer product compounds, including parabens and mono-ethyl phthalate, as well as several metals, such as mercury and arsenic. Moreover, for Mexican American, Other Hispanic, and non-Hispanic black women, there were several exposure disparities that persisted across age groups, such as higher 2,4- and 2,5-dichlorophenol

concentrations. Exposure differences for methyl and propyl parabens, however, were the starkest between non-Hispanic black and non-Hispanic white children with average differences exceeding 4-fold. Exposure disparities for methyl and propyl parabens are increasing over time in Other Race/Multi-Racial women while fluctuating for non-Hispanic Black, Mexican American, and Other Hispanic. Differences in cotinine levels are among the highest in Non-Hispanic White women compared to Mexican American and Other Hispanic women with disparities plateauing and increasing, respectively.

Discussion: We systematically evaluated differences in chemical exposures across women of various race/ethnic groups and across age groups and time. Our findings could help inform chemical prioritization in designing epidemiological and toxicological studies. In addition, they could help guide public health interventions to reduce environmental and health disparities across populations.

3.2 Introduction

The stark racial disparities in disease incidence and health outcomes among American women remain a persistent public health challenge. For example, preterm birth incidence is approximately 60% higher in non-Hispanic Black women relative to non-Hispanic White women [167]. Non-Hispanic Black and Hispanic women are at increased risk of being diagnosed with developing dysglycemia [168] and diabetes [169], relative to non-Hispanic White women. Non-Hispanic Black women are also 2-3 times more likely to develop the most aggressive subtype of breast cancer, triple negative, compared to non-Hispanic White women [79,80]. Furthermore, relative to non-Hispanic White women, non-Hispanic Black women are also 2.4 times more likely to die of breast cancer after being diagnosed with the pre-invasive lesion, ductal carcinoma *in situ* [170]. Recent statistics from the American Cancer Society also show variation in trends in breast cancer

incidence rates by race/ethnicity in US women from 2005-2014. Specifically, breast cancer incidence rates are increasing in Asian (1.7% per year), non-Hispanic Black (0.4% per year), and Hispanic (0.3% per year) women, while rates remain stable in non-Hispanic White and American Indian/Alaska Native women [171]. Dementia rates also vary by race/ethnicity, with rates highest in non-Hispanic black women, followed by American Indian/Alaskan native, Latina, Pacific Islander, non-Hispanic White, and lowest in Asian American women [172]. These rates vary 60% between African American and Asian American women. Reproductive outcomes are also significantly different by race/ethnicity, with studies reporting increased incidence of gestational diabetes in South and Central Asian American women [173] and Black and Hispanic women [174]. Collectively, these findings suggest profound racial disparities in disease outcomes that manifest throughout the life course. Understanding the etiological factors driving these health disparities is essential for informing public health interventions seeking to promote health equity.

While health disparities are likely due to complex interactions between genetic, social, and lifestyle factors, the impact of genetic variation on disease disparities appears to be minor [81–83]. For example, a meta-analysis of genetic factors underlying racial disparities in cardiovascular disease failed to identify heterogeneity of genetic risk factors by race/ethnicity [175]. These findings of a modest genetic impact on differential cardiovascular disease risk by race/ethnicity are consistent with genome-wide association studies. A study found that variants with the strongest association with blood pressure explain, in aggregate, less than 5% of the phenotypic variance [176]. Moreover, a meta-analysis of genetic risk factors and cancer disparities reported similar findings, with almost no heterogeneity in cancer risk alleles by race/ethnicity [84].

The limited impact of genetic risk factors in explaining health disparities points towards environmental risk factors as major determinants. Indeed, estimates of environmental factors on

chronic disease suggest than 70-90% of risk is due to environmental exposures [12,177]. A mechanistic understanding of racial disparities in disease therefore requires a characterization of differences in environmental risk factors. In particular, differences in chemical exposures have been hypothesized to be important etiologic factors in racial disparities in disease rates [178–182].

To investigate the influence of environmental risk factors on health disparities, the goal of this study was to conduct a comprehensive analysis of racial disparities in chemical biomarker concentrations among US women. To accomplish this, we leveraged data from the National Health and Nutrition Examination Survey (NHANES), an ongoing population-based health study conducted by the US Centers for Disease Control and Prevention (CDC). Additionally, we developed visuals to highlight differences in biomarker concentrations across races, age groups, and time, by defining the relative magnitude of exposure disparities for individual chemicals and chemical families.

3.3 Methods

3.3.1 Study Population

NHANES is a cross-sectional study designed for collecting data on demographic, socioeconomic, dietary, and health-related characteristics in the non-institutionalized, civilian US population. For this analysis, we used the continuous NHANES data on chemical biomarkers and demographics, which were collected from 1999-2014 with 82,091 participants initially. We excluded participants for not having any data on chemical biomarkers (N = 7,001), resulting in a sample size of 75,090 study participants. Since this analysis is focused on measuring chemical disparities in US women, we excluded male participants (N = 37,010), leading to a final sample size of 38,080 female participants. For a given chemical, we also excluded participants with missing data on any of the following covariates: race/ethnicity, age, NHANES cycles, poverty

income ratio, cotinine levels, and urinary creatinine. These exclusion and inclusion criteria are delineated in **Figure 3.1**.

3.3.2 Chemical Biomarker Measurements

This section along with **Figure 3.1** delineates the curation process for selecting chemical biomarkers to include for analysis. First, we excluded biomarkers that are not indicative of chemical exposures (n = 99). Next, we corrected for differences in chemical codenames by using a unique codename for each biomarker (n = 36). We gave preference to lipid-adjusted data and therefore excluded non-lipid adjusted chemical biomarkers (n = 79) when both types of data were provided for a given chemical. We replaced all measurements below the limit of detection (LOD) with $LOD/\sqrt{2}$ as recommended by the CDC [131]. This approximates a lognormal distribution, so that reasonably unbiased means and standard deviations are produced [132]. There were also instances in which urinary cadmium concentrations were recorded as 0 ng/mL due to interference with molybdenum oxide (NCHS, 2005a, NCHS, 2005b). We replaced such values with $LOD/\sqrt{2}$ if the participant's urinary cadmium level was under the LOD or otherwise excluded. We calculated detection frequencies for each chemical biomarker and excluded biomarkers with detection frequencies of 50% or less (n = 182) across all study participants. As we have reported previously, across the NHANES cycles, improvements in laboratory technology can change the LOD and thus lead to differences in detection frequencies by NHANES cycle (Nguyen et al. 2019). To limit bias from these changing LODs over time, we calculated detection frequencies by NHANES cycle for each chemical biomarker and excluded measurements where the LOD changed by over an order of magnitude and detection frequencies over time (Nguyen et al. 2019). For instance, percentages of participants with PCB 199 measurements above LODs for Cycle 2 and Cycle 3 are 36.2% and 84.9%, respectively, and the LOD for Cycle 2 and Cycle 3 were 10.50 ng/g and 0.40 ng/g,

respectively. As such, measurements from Cycle 2 for PCB 199 were excluded. Measurements from given cycles for several PCBs, Dioxins, Furans, Phytoestrogens, and VOCs along with Paranitrophenol, 2-napthol, 1-pyrene and 9-pyrene ($m = 449{,}396$ total data points) were therefore also excluded based on these criteria (**Table A2.1**). The final dataset for analysis consisted of 143 chemical biomarkers from 16 different chemical classes (**Table A2.2**). These chemical classes were: Acrylamide, Brominated Flame Retardants (BFR), Phosphate Flame Retardants (PFR), Dioxins, Furans, Metals, Other, Personal Care & Consumer Product Compounds, Pesticides, Phthalates & Plasticizers, Phytoestrogens, Polyaromatic Hydrocarbons (PAH), Polychlorinated Biphenyls (PCB), Per- and Polyfluoroalkyl Substances (PFAS), Smoking Related Compounds, and Volatile Organic Compounds (VOC).

3.3.3 Statistical Analysis

We performed all analyses using R version 3.6.0. Given the NHANES complex sampling design, we applied appropriate survey weights in our statistical models to produce estimates representative of the non-institutionalized, civilian US population. Applying the appropriate survey weights involved selecting the weights of the smallest analysis subpopulation [183]. For example, there are two types of survey weights that can be used for a conducting an analysis with total arsenic: WTSA2YR or WTMEC2YR. WTMEC2YR is the NHANES codename for survey weights for all participants whose measurements were taken in a Mobile Exam Center (MEC). WTSA2YR is similar to WTMEC2YR with the exception that the survey weights are only for participants who belong to subsample A, which is a smaller subpopulation with measurements for total arsenic. Since WTSA2YR pertains to the smaller analysis subpopulation, WTSA2YR is the appropriate survey weight to apply for an analysis on total arsenic. The appropriate survey weights are listed in the original NHANES files containing the measurements for the chemical biomarker.

If the survey weights are not listed for a given chemical, then we use WTMEC2YR as the default. Conducting an analysis across several NHANES cycles may require the use of different survey weights [183]. For instance, a statistical model for total arsenic requires using only one type of survey weights (WTSA2YR). But a statistical model for triclosan requires three different types of survey weights (WTSA2YR, WTSB2YR, WTSC2YR), since in each NHANES cycle, a different subpopulation of NHANES is measured for triclosan. To account for NHANES sampling design and use the appropriate survey weights, we developed two databases. The first was a database of codenames indicating the appropriate survey weights for each chemical biomarker and NHANES cycle (**Table A2.3**). For several of the Per- and Polyfluoroalkyl Substance (PFAS), there were two different type of survey weights available within the same cycle (one for children aged 3-11 and the other for participants aged 12 and older). Therefore, we developed another database of codenames indicating which additional survey weights to use when generalizing these results for PFASs (**Table A2.4**).

Using multivariate regression models, we evaluated differences in biomarker concentrations in blood and urine by race after log-transforming the data. We included log-transformed levels of cotinine as a covariate to represent smoking (Benowitz, 1999), and creatinine levels to adjust for urine dilution and flow differences (Barr et al., 2005). We modeled poverty income ratio (PIR) as a surrogate variable for socioeconomic status. PIR is the ratio of household income and poverty threshold adjusted for family size and inflation. First, we examined the racial differences in chemical biomarker levels by performing a series of chemical-specific regression models with the main predictor being race/ethnicity (categorical), adjusting for age (continuous), NHANES cycle (continuous), PIR (continuous), and cotinine (continuous) as described in Eq. (1):

$$Log10(X_{Chemical\ Concentration}) = \beta_{race/ethnicity,j}(X_{race/ethnicity,j}) +$$

$$\beta_{age}(X_{age}) +$$
$$\beta_{cycle}(X_{cycle}) +$$
$$\beta_{PIR}(X_{PIR}) +$$
$$\beta_{cotinine}(X_{cotinine}) +$$
$$\beta_{creatinine}(X_{creatinine}) +$$
$$\alpha \qquad [1]$$

Here, $X_{Chemical\ Concentration}$ is the log-transformed, unadjusted chemical biomarker concentration for all participants, X_i, where $i \in \{race/ethnicity, j,\ age,\ sex,\ cycle,\ PIR,\ cotinine,\ creatinine\}$, is the i covariate for all participants, β_i is the linear regression coefficient for the i covariate, and α is the intercept. $X_{race/ethnicity,j}$, where $j \epsilon \{Mexican\ Americans,\ Other\ Hispanics,\ Non\text{-}Hispanic\ Black,\ Other\ Race/Multiracial\}$ for 1999-2014, is the race covariate for comparing the jth race to the reference group of Non-Hispanic Whites. For chemical biomarkers which were measured in urine, we further corrected the regression models by adjusting for urinary creatinine levels (continuous). For the analyses where cotinine concentration was the outcome, the regression models were not further corrected for smoking. Prior to 2011, Asian Americans were categorized in Other Race/Multi-Racial category. Accordingly, to evaluate chemical exposure disparities in Asian American women, we also applied Eq. 1 to the 2011-2014 data. Then to determine whether racial disparities are driven by differences in socioeconomic status, we conducted a sensitivity analysis to observe how the race coefficients change with and without adjustment for PIR in the regression models. The coefficient for jth race represents the difference in log-transformed chemical biomarker concentration between the jth race and the reference group of Non-Hispanic Whites. As we are making multiple comparisons across chemical biomarkers and races, we have an increased probability of detecting false positives, i.e. a high false

positive rate. To identify significant comparisons while maintaining a lower false positive rate, we used the False Detection Rate (FDR) method on the p-values of the linear regression race-coefficients [103].

We are interested in understanding how the racial disparities are influenced by reproductive and nutritional factors such as parity, breastfeeding, menopause/hysterectomy status, and iron deficiency, which may impact toxicant absorption and excretion. We used the reproductive health questionnaire on reasons for having period irregularities and another questionnaire on having regular periods in past 12 months to classify participants as having menopause/hysterectomy, otherwise with irregular periods, or otherwise with regular periods (reference). Parity (continuous) is defined from the questionnaire on the number of pregnancies resulting in live births. We used parity and a questionnaire asking whether the mother breastfed her children for at least a month to categorize participants into three categories: breastfed, did not breastfeed her children, and did not breastfeed as she does not have children (reference). We used sex-specific thresholds for hemoglobin levels (hemoglobin <13.5 g/dL for men and <12 g/dL for women) [184] to classify participants as iron deficient or not (reference). To determine whether the frequency distribution of these variables differ by race, we conducted a Chi-Square test on the categorical variables and a one-way ANOVA test on parity. To characterize how the estimates of racial disparities change when a given reproductive or nutritional factor is included in the regression model, we conducted a series of regression models with the outcome variable as chemical concentrations and the main predictor as race/ethnicity (categorical) while adjusting for age (continuous), NHANES cycle (continuous), PIR (continuous), cotinine (continuous), creatinine (continuous), and an fth reproductive or nutritional factor from the set of

{parity (continuous), breastfeeding (categorical), menopause/hysterectomy (categorical), iron deficiency (categorical)} as described in Eq. (2).

$$Log10(X_{Chemical\ Concentration}) = \beta_{race/ethnicity,j}(X_{race/ethnicity,j}) + \beta_{age}(X_{age}) + \beta_{cycle}(X_{cycle}) + \beta_{PIR}(X_{PIR}) + \beta_{cotinine}(X_{cotinine}) + \beta_{creatinine}(X_{creatinine}) + \beta_{factor,f}(X_{factor,f}) + \alpha \quad 2]$$

For example, if $f = parity$, the $X_{factor,f}$ refers the number of pregnancies resulting in live birth for all participants. $\beta_{factor,f=parity}$ is interpreted as the change in log-transformed chemical biomarker concentration for every successful pregnancy. Now, if $f = iron\ deficiency$, then $X_{factor,f}$ indicates whether a participant is iron deficient or not for all participants. $\beta_{factor,f=iron\ deficiency}$ is interpreted as the difference in log-transformed chemical biomarker concentration between iron deficient participants and the reference group of participants who are not. The interpretations for menopause/hysterectomy status and breastfeeding status are similar to that of iron deficiency.

To evaluate how these racial differences in chemical exposures differ by age group, we conducted stratified analyses by age groups in the 1999-2014 data. We defined 4 age groups: 0-11, 12-19, 20-50, and 51-85. For each age group with chemical biomarker measurements, we performed a chemical specific linear regression with the main predictor as race/ethnicity

(categorical) and adjusted for age (continuous), NHANES cycle (continuous), PIR (continuous), cotinine (continuous), and creatinine (continuous), stratified by age group described in Eq. (3).

$$
\begin{aligned}
Log10(X_{Chemical\ Concentration}[age\ group = k]) = \ & \beta_{race/ethnicity,j,k}\big(X_{race/ethnicity,j}[age\ group = k]\big) + \\
& \beta_{age,k}\big(X_{age}[age\ group = k]\big) + \\
& \beta_{cycle,k}\big(X_{cycle}[age\ group = k]\big) + \\
& \beta_{PIR,k}(X_{PIR}[age\ group = k]) + \\
& \beta_{cotinine,k}(X_{cotinine}[age\ group = k]) + \\
& \beta_{creatinine,k}(X_{creatinine}[age\ group = k]) + \\
& \alpha
\end{aligned}
$$

Here, k is an available age group from the set of $\{0\text{-}11, 12\text{-}19, 20\text{-}50, 51\text{-}85\}$, $X_{Chemical\ Concentration}[age\ group = k]$ is the log-transformed, unadjusted chemical biomarker concentration for all participants with ages in the kth age groups, $X_{i,k}[age\ group = k]$, where $i \in \{race/ethnicity, j,\ age,\ sex,\ cycle,\ PIR,\ cotinine,\ creatinine\}$, is the i covariate for all participants with ages with the kth age group, $\beta_{i,k}$ is the linear regression coefficient for the i covariate and kth age group, and α is the intercept. $X_{race/ethnicity,j,k}$, where $j \epsilon \{Mexican\ Americans,\ Other\ Hispanics,\ Non\text{-}Hispanic\ Black,\ Other\ Race/\ Multiracial\}$, is the race covariate for comparing the jth race to the reference group of Non-Hispanic Whites in the kth age group. To account for multiple comparisons, we used an FDR method on the p-values of the linear regression race-coefficients across all age groups (Benjamini and Hochberg, 1995).

To evaluate how racial disparities in chemical exposures changes over time, we conducted stratified analyses by NHANES cycles. We denote a cth cycle as from a study period from the set

of

{1 (1999-2000), 2 (2001-2002), 3 (2003-2004), 4 (2005-2006), 5 (2007-2008), 6 (2009-2010), 7 (2011-2012), 8 (2013-2014)}. We excluded measurements from Cycle 2 for Blood Bromodichloromethane, Cycle 2 for Blood Chloroform, and Cycle 6 for Blood Toluene due to having an error in the statistical program R when accounting for the sampling design. We excluded measurements from Cycle 2 for 3-fluoranthene as there is only one person with measurements for this chemical in Cycle 2. For each NHANES cycle with chemical biomarker measurements, we performed a chemical specific linear regression with the main predictor as race/ethnicity (categorical) and adjusted for age (continuous), PIR (continuous), cotinine (continuous), and creatinine (continuous) as described in Eq. (4). $[cycle = c]$ denotes the inclusion of participants who have chemical biomarker measurements in the cth cycle. We used an FDR method on the p-values of the coefficients to account for multiple comparisons across all races and NHANES cycles.

$$
\begin{aligned}
Log10(X_{Chemical\ Concentration}[cycle = c]) \quad = \quad & \beta_{race/ethnicity,j,c}\left(X_{race/ethnicity,j}[cycle = c]\right) + \\
& \beta_{age,c}\left(X_{age}[cycle = c]\right) + \\
& \beta_{PIR,c}(X_{PIR}[cycle = c]) + \\
& \beta_{cotinine,c}(X_{cotinine}[cycle = c]) + \\
& \beta_{creatinine,c}(X_{creatinine}[cycle = c]) + \\
& \alpha
\end{aligned}
\qquad 4]
$$

3.4 Results

Table 3.1 displays demographic characteristics of the study population. The study population includes 38,080 female study participants of ages 1-85 years, with a median age of 26. Using a series of covariate adjusted regression models, we first calculated the fold-difference in chemical biomarker concentrations by race across the entire study population. These regression results are presented in graphical format in **Figure 3.2**, where the letters in the plot reflect the fold-difference in chemical biomarkers for each race/ethnicity, relative to non-Hispanic White women, who made up the largest portion of the study population. Pesticides and pesticide metabolites, including 2,5-dichlorophenol, o,p'-DDE, beta-hexachlorocyclohexane, and 2,4-dichlorophenol had amongst the highest average fold difference across non-Hispanic Black, Mexican American, Other Hispanic, and other race/multiracial women. On average, large differences by race are also apparent for personal care and consumer product compounds including methyl paraben, propyl paraben, monoethyl phthalate and metals, such as mercury and arsenic. Conversely, cotinine, PBDE-153, PBB-153, Equol, DEET, and bisphenol F were among the chemicals of which non-Hispanic White women had the highest levels.

In order to more clearly visualize the differences in chemical biomarkers by race/ethnicity, we generated volcano plots, which are displayed in **Figure 3.3**. The x-axis of these plots depicts the fold difference in average chemical biomarker concentration between each race/ethnicity and non-Hispanic White women. The y-axis depicts statistical significance, as reflected in the negative $\log_{10}$ transformation of the FDR-adjusted p-value from the regression analysis for that chemical biomarker, where chemicals with larger values on the y-axis are more statistically significant. As shown in **Figure 3.3A**, non-Hispanic black women have biomarker concentrations that are more than twice those of non-Hispanic White women for multiple chemicals, including 2,5-

dichlorophenol, 1,4-dichlorobenzene, methyl paraben, monoethyl phthalate, 2,4-dichlorophenol, and propyl paraben. The heavy metals, mercury and lead, are also significantly higher in non-Hispanic Black women. Conversely, levels of benzophenone-3, a UV blocker used in sunscreen, are significantly higher in non-Hispanic White women. In general, concentrations of PCBs tend to be modestly elevated in non-Hispanic Black women, while volatile organic compounds (VOCs) and phytoestrogen concentrations are higher in non-Hispanic White women. **Figure 3.3B** shows relative differences in chemical biomarker concentrations between Mexican American and non-Hispanic White women. Pesticides, including 2,5-dichlorophenol, beta-hexachlorocyclohexane, and 2,4-dichlorophenol, along with the polycyclic aromatic hydrocarbon 2-napthol were on average higher in Mexican American women. Conversely, the smoking biomarker, cotinine is significantly lower in Mexican American women. Exposure patterns comparing Other Hispanic and non-Hispanic White women, displayed in **Figure 3.3C**, showed some similarities, with pesticides 2,5-dichlorophenol and p,p'-DDE elevated in Other Hispanic women. Multiple PFASs, including PFOS, PFHxS, and 2-(N-methyl-PFOSA) acetate, as well as cotinine, are significantly lower in Other Hispanic women. **Figure 3.3D** shows a distinct exposure pattern in women of other race/ethnicity or multiracial women. Here, levels of heavy metals, including cadmium, mercury, and multiple arsenic biomarkers, are significantly elevated relative to non-Hispanic White women. Conversely, the smoking biomarkers, NNAL and cotinine, are significantly lower.

To understand whether socioeconomic status is a driver of racial disparities in chemical exposures, we generated a series of correlation plots, comparing how the differences in chemical biomarker concentrations by race/ethnicity change with the inclusion and exclusion of PIR in the regression models (**Figure A2.1**). For many of the chemicals, the fold differences for comparing chemical biomarker levels by race did not change drastically when including PIR as a covariate in

the regression models, implying that socioeconomic status is not the primary driver in explaining differences in chemical exposures. However, for cotinine, PCB 194, and several chemicals used in personal care products, the relative differences changed by greater than 25% when PIR was included as a covariate in the regression models. This suggests that either exposure differences between races for these chemicals are mediated by PIR, and/or exposure differences are explained by interactions between race and socioeconomic status. To visualize differences in chemical biomarker concentrations by race across a gradient of income for a few selected biomarkers, we generated violin plots of the chemical biomarker distribution stratified by categories of PIR for each race/ethnicity (**Figure A2.2**). For benzophenone-3 and cotinine (**Figure A2.2A and A2.2B**), the trends of biomarker concentrations across the PIR categories and the average concentrations within the same PIR categories differ by race. This is similar for ethyl paraben (**Figure A2.2C**), but differences are not as drastic. On the other hand, mercury (**Figure A2.2D**) along with other remaining chemicals demonstrated a very different pattern from those of the previously mentioned substances. Across all races, the trends across PIR categories are similar for mercury, but within the same PIR category, there are differences in biomarker concentrations by race, suggesting that many chemical exposures disparities by race are independent of PIR.

We also characterized how reproductive and nutritional factors such as parity, breastfeeding, menopause/hysterectomy status, and iron deficiency may influence racial disparities in chemical exposures, as well as to account for racial variations in these factors in our models (Aliyu et al. 2005; Henderson et al. 2008; Zakai et al. 2008). There are significant differences by race in all studied reproductive and nutritional factors with the p-values listed as the following: parity (p-value = 6.59e-95), breastfeeding (p-value = 4.71e-121), menopause/hysterectomy status (p-value = 6.14e-148), and iron deficiency (p-value = 9.24e-322).

Contingency tables are provided to show the frequency distribution of these variables by race in **Tables A2.5-A2.8**. In addition, we compared how differences in chemical biomarker concentrations by race/ethnicity change with and without accounting for these factors by generating a series of correlation plots comparing regression coefficients from models that include or exclude these factors (**Figure A2.3-A2.6**). Adjusting for either iron deficiency or menopause/hysterectomy status resulted in four chemicals that have the relative differences changed by greater than 25%, implying that neither iron deficiency nor menopause are primary driver in explaining racial differences in exposure. On the other hand, adjusting for breastfeeding and parity showed 25 and 16 chemicals, respectively, with changes greater than 25%, which implies that racial disparities in chemical exposure are better explained by these factors. For cotinine, the changes in relative differences by race was among the highest at approximately two-fold different when either menopause or breastfeeding was considered in the regression models.

Starting in 2011, more detailed information on NHANES study participant race/ethnicity were collected, including specifically identifying individuals who report Asian ethnicity. To understand whether the results presented in **Figure 3.3D** predominantly reflect results in Asian women, who prior to 2011 were categorized in other race/multi-racial category, we assessed exposure disparities specifically in the Asian population. These results, presented in **Figure 3.4A**, show that, on average, multiple heavy metal biomarkers are more than 2-fold higher relative to non-Hispanic White women, including cadmium, mercury, lead, and arsenics. Additionally, the PFAS compound PFDA is significantly higher in Asian women, while cotinine and biomarkers of phosphate flame retardants (Bis(1,3-dichloro-2-propyl) phosphate, Dibutyl phosphate, Diphenyl phosphate) are significantly lower. We also calculated whether there were significant disparities in chemical biomarker concentrations in women of other or multi-race after excluding Asian

women. **Figure 3.4B** suggests relatively few differences in this regard, confirming that the other race effect in **Figure 3.3D** is indeed associated with Asian women.

We have previously shown dramatic differences in the chemical "exposome" by age in NHANES study participants, not stratified by gender or race [185]. Here, we tested for differences in chemical biomarkers by race, after stratifying by age group. **Figure 3.5** displays these results across the entire study population from 1999-2014. Blue colors reflect chemicals where levels are higher in non-Hispanic White women, while red colors reflect chemicals that are of higher concentration in women of the labeled race/ethnicity. Here, there appear to be exposure disparity patterns that persist across age groups – such as higher 2,4- and 2,5-dichlorophenol concentrations in Mexican American, Other Hispanic, and non-Hispanic black women. Differences in 1,4-dichlorobenzene concentrations are consistent across age groups, although this biomarker was not measured in the youngest individuals. Heavy metal concentrations are elevated in women of other race across age groups. Some exposure patterns differ by age, however. For example, differences in methyl and propyl paraben are most apparent between young non-Hispanic black and non-Hispanic White women less than 12 years old. Increased levels of phosphate flame retardants and the insect repellent DEET in non-Hispanic White women are the most evident in women less than 12 years of age. Reduced levels of brominated flame retardants (PBDE's) in levels in non-Hispanic White women are emphasized for adolescents, age 12-19 (all other races are higher, in red). Similarly, higher relative concentrations of benzophenone-3, bisphenol A, and bisphenol F occur in non-Hispanic White women less than 12. Elevated PCB levels in non-Hispanic black women shown in **Figure 3.3A** are most evident in women greater than 51 years of age. Overall, these results highlight racial exposure disparities that are either stable or that vary across age groups.

To characterize how racial disparities in chemical exposures changes over time, we conducted analyses stratified by NHANES cycles (**Figure 3.6**). There are exposure disparities patterns that persist across time for arsenics and its metabolites, chemicals used in personal care products, cotinine, 2,5−dichlorophenol, 2,4−dichlorophenol, and 1,4−dichlorobenzene. Differences in 2,5−dichlorophenol concentrations are consistent higher in Non-Hispanic Black women over time, while differences peaked at the turn of the century and towards the end of 2000s, respectively, for Mexican American and Other Hispanic women. Arsenic and its metabolite arsenobetaine show the highest disparities in Other Race/Multi-Racial women with the disparities peaking in 2007-2010. Disparities patterns for methyl and propyl parabens fluctuate over time in Mexican American, Other Hispanic, and Non-Hispanic Black women, while differences for these substances are slightly increasing in Other Race/Multi-Racial women. For mono-ethyl phthalate, a metabolite of DEP used in personal care products, differences over time are consistent in Mexican Americans, fluctuating in Other Hispanics, and increasing in non-Hispanic Black women. Differences in cotinine levels are among the highest in Non-Hispanic White women compared to Mexican American and Other Hispanic women with disparities plateauing and increasing, respectively

3.5 Discussion

Based on population based chemical biomonitoring generated as part of the 1999-2014 NHANES, we performed a comprehensive analysis of racial disparities in biomarker concentrations of 143 chemicals in 38,080 participants. Specifically, we quantified the relative magnitude of racial disparities for individual chemicals and chemical families while utilizing appropriate regression weightings. This helped ensure that the results were as generalizable to the entire US population. These results highlighted striking differences in chemical biomarker

exposure patterns by race/ethnicity, independent of other demographic factors such as socioeconomic status and independent factors such as menopause/hysterectomy status, parity, breastfeeding, and iron deficiency. In particular, exposure patterns of pesticides, heavy metals, tobacco smoke associated compounds, and chemicals found in personal care products are found to be most disparate across race/ethnic groups. Stratified analyses revealed exposure patterns that persisted across age groups. For example, this was apparent in heavy metals exposure for women who identify as other race or multiracial, as well as in age-specific exposure patterns, such as elevated PCB, dioxin, and dibenzofuran exposure in older non-Hispanic black women. In some cases, average differences in chemical biomarker concentrations between race/ethnic groups exceeded 400%, such as for urinary propyl or methylparaben concentrations between the youngest non-Hispanic Black and non-Hispanic White women. Racial disparities were attenuated or emphasized after adjusting for reproductive and nutritional factors. For example, when adjusting for breastfeeding, the average differences in cotinine biomarker levels were attenuated two-fold between Hispanic and non-Hispanic White women, whereas accounting for variation in iron deficiency resulted in average differences in cotinine levels increasing by two-fold when comparing between non-Hispanic Whites and non-Hispanic Black women. Since parity, breastfeeding, menopause, and iron deficiency may influence the absorption or elimination of chemicals, and the rates of these potential confounders differ by race, these factors may lead to further attenuation or amplification of racial disparities in chemical biomarker concentrations. These findings contextualize racial disparities in chemical exposures across US women and highlight the vast differences in chemical exposomes between demographic groups with well characterized disparities in health outcomes.

Environmental injustice is the disproportionate exposure of individuals of color, lower socioeconomic status, or other politically disadvantaged groups to toxic chemicals in food, air, consumer products, at the workplace, or in their communities [186]. Disproportionate chemical exposures have been hypothesized to be important drivers of health disparities, including obesity and neurodevelopmental outcomes [187]. While the primary goal of this study was to quantify and compare chemical exposure disparities across racial/ethnic groups, independent of income, others have evaluated combined income and race related disparities in exposure. For instance, one analysis compared geometric mean concentrations of 228 chemical biomarkers between six groups stratified by income and race in NHANES and identified 37 chemicals as likely contributing to environmental justice [188]. Some of these chemicals, including cotinine, lead, 2,4- and 2,5-dichlorophenol, methyl paraben, and propyl paraben, were associated with the highest disparities across race/ethnic group in the present study. We also compared chemical exposures disparities across racial/ethnic groups with and without adjustment for income and found that cotinine, PCB 194, methyl mercury, and chemicals used in personal care products such as benzophenone-3, the parabens, and triclosan show disparities across both race and socioeconomic status. However, for most of the studied chemicals, differences in chemical exposures were not driven by socioeconomic status but were instead primarily associated with race/ethnicity. Furthermore, a study of racial and social disparities in exposure to BPA and PFAS examined differences in biomarker concentrations in NHANES study participants [189]. The concentrations of the four PFAS chemicals examined, PFOA, PFOS, PFNA, and PFHxS, were inversely associated with household income, while BPA concentrations were higher in individuals who reported low food security [189]. Here, we identified that, independent of socioeconomic status, as assessed by poverty-income ratio, non-Hispanic White women had the highest concentrations of PFOA, while non-Hispanic

Black and other race/multiracial women had the highest concentrations of PFDA. Major routes of exposure to PFAS compounds include contaminated drinking water [190], diet [127], and occupational routes [191]. BPA concentrations were not strikingly different by race in our study, but non-Hispanic Black women had, on average, 93% higher BPS concentrations than non-Hispanic White women. Common routes of exposure to BPA and other bisphenol analogues are diet, thermal paper, and personal care products [192]. Further research is necessary to identify the major routes of exposure which are driving racial disparities in PFAS and bisphenol chemicals biomarker concentrations.

The findings of highly elevated monoethyl phthalate and methyl and propyl paraben concentrations in the non-Hispanic Black women are consistent with a personal care product route of exposure. A study assessing the chemical composition of hair products used by Black women consistently identified high levels of cyclosiloxanes, parabens, and the fragrance carrier diethyl phthalate [193]. In our study, the concentrations of the diethyl phthalate metabolite monoethyl phthalate were approximately 78% higher on average in non-Hispanic black women of all ages relative to non-Hispanic White women, and 122% higher in non-Hispanic black women less than 12 years of age. This is concerning, since urinary concentrations of monoethyl phthalate have been positively associated with odds of developing breast cancer in a case-control study of women from Northern Mexico [194]. Differences in concentrations of methyl and propyl paraben biomarkers were among the highest observed in this study, particularly for the youngest non-Hispanic Black women. These differences were observed to remain consistently higher across the NHANES cycles in Mexican Americans, Other Hispanics, and non-Hispanic Blacks. These chemicals have been used as preservatives in personal care products, pharmaceuticals, and food additives, and have been found to promote cell growth through multiple mechanisms, including estrogenicity [195–197] and epidermal growth factor receptor signaling [198]. Particularly relevant to our findings of the

greatest methyl and ethyl paraben disparities in the youngest non-Hispanic Black women was the finding that early life paraben exposures can alter developing mammary gland morphology and induce gene expression that resembles an early cancer-like state [199]. Use of hair products has been identified as a potential risk factor for breast cancer in non-Hispanic Black women [200]. When we adjusted for breastfeeding in the regression models, parabens levels are higher in Non-Hispanic White women compared to Mexican American, Other Hispanics, and Other Race/Multi-Racial women. This implies that women of other races are breastfeeding more often [201], and/or they are eliminating parabens from their body but exposing their infants via breast milk [202]. Further research is needed to determine whether early-life exposure to potentially estrogenic compounds, like parabens, can induce biological alterations that increase risk of estrogen receptor negative breast cancers.

One of the most apparent disparities in chemical biomarker concentrations by race was with the compounds 2,4-dichlorophenol, 2,5-dichlorophenol, and 1,4-dichlorobenzene. 1,4-dichlorobenzene is used as a disinfectant, pesticide, and deodorant. 2,5-dichlorophenol is a metabolite of 1,4-dichlorobenzene, while 2,4-dichlorophenol is a metabolite of the antimicrobial triclosan or other pesticides. Elevated concentrations of these chemicals in non-Hispanic Black individuals has been noted previously [188,203] The concentrations of these three chemicals were up to 350% higher on average in non-Hispanic Black women, relative to non-Hispanic White women, and also elevated in Mexican American and Other Hispanic women. Importantly, these exposure disparities were consistent across all age groups. While 2,4-dicholorophenol concentrations were significantly elevated in non-Hispanic Black and Hispanic women, urinary triclosan levels were not significantly different by race/ethnicity. This suggests that either triclosan is not the main chemical exposure that explains the differences in concentrations of 2,4-dichlorophenol or that

there are differences in metabolism and excretion rates by race, which is less likely. 2,5-dichlorophenol is of particular interest, since after adjusting for breastfeeding, the exposure disparity between Mexican American and non-Hispanic White women was further emphasized. In a study measuring environmental phenols in milk of lactating North Carolina women, 2,5-dichlorophenol was undetectable in all milk samples [202]. This suggests that either 2,5-dichlorophenol is not absorbed into breast milk or it is hindering lactation, and therefore excretion through this pathway. 1,4-dichlorobenzene exposure has been associated with altered thyroid biomarkers in NHANES [204], altered immunologic and liver function parameters in occupationally exposed workers [205], and altered sperm production and increased prostate weight in exposed rats [206]. Understanding and mitigating exposure to these chemicals is therefore of importance to reduce disparate risk of these health outcomes.

Heavy metals were among the chemicals most consistently different across racial/ethnic groups. In particular, women who identified as other race or multiracial had the highest concentrations of multiple metals, including cadmium, mercury, arsenics, lead, and manganese. Focusing on data from NHANES 2011-14, we identified that these elevated metals concentrations were restricted to women who identified as Asian. This is consistent with a previous finding of increased concentrations of a subset of these metals in Asian NHANES participants [36]. Furthermore, elevated levels of mercury, lead, and arsenics were also identified in non-Hispanic Black women, relative to non-Hispanic White women. Mexican American women had elevated levels of uranium, lead, mercury, arsenics, and cadmium, while Other Hispanic women had higher concentrations of mercury, arsenics, and cadmium than non-Hispanic White women. Non-Hispanic White women, however, had higher concentrations of urinary barium. In our temporal analysis, differences in biomarker levels between Other Race/Multi-Racial and Non-Hispanic

White women are increasing for cadmium, mercury, and dimethylarsonic acid. While studies have identified the interaction between iron deficiency and biomarker levels of heavy metals [207,208], we observed in our study that accounting for iron deficiency in the regression models did not influence the racial disparities for heavy metals. Previous research has linked diet, occupation, education level, and smoking status to elevated metals exposure [36], in addition to housing [209], air pollution [210], and contaminated water [211]. The well characterized toxicity of heavy metals exposure, even at low doses, makes identifying and ameliorating heavy metal exposures a top priority for addressing environmental health disparities.

The oldest non-Hispanic Black women in our study had consistently higher concentrations of persistent organic pollutants, including dioxins, dibenzofurans, PCBs, and DDT metabolites. In addition, these stark disparities for the persistent organic pollutants were plateauing or increasing over time in non-Hispanic Black women when these chemicals were last measured in NHANES. This is consistent with a previous report of non-Hispanic Black individuals having an increased risk of having multiple persistent organic pollutants detectable their blood [212] or higher average levels of PCBs [121]. In our study, biomarker levels of most PCBs were higher in women who were iron deficient. The less persistent PCBs were also of lower concentrations in women who breastfed or have higher parity, suggesting that the depuration of PCBs occurs through these excretion mechanisms [213]. However, biomarker levels were shown to be higher for the more persistent PCBs. Biomarkers of persistent organic pollutants were quantified on an individual (non-pooled) basis in the 1999-2004 NHANES cycles. Elevated concentrations of these pollutants, such as the DDT metabolite, DDE, have been associated with an increased risk of breast cancer [214]. Interestingly, racial differences in DDE further increased when either parity or breastfeeding was accounted in the regression models, suggesting that environmental insult from this substance may perturb

pathways associated with the reproductive system. A lack of disparities, and decreasing concentrations of these chemicals in younger individuals over time, generally reflect a public health success in decreasing population exposures to these toxic compounds [185]. The long half-life of these chemicals suggests that the detected biomarkers predominantly reflect historical exposures. This could, however, be of substantial importance for children of non-Hispanic Black women, who could have been exposed to disproportionately high levels of these chemicals in the womb or early in childhood. For example, *in utero* exposure to the pesticide, DDT, has been associated with an increased risk of breast cancer in adulthood. Specifically, women in the highest quartile of *in utero* DDT exposure were found to have a 3.7-fold increased risk of developing breast cancer relative to women in the lowest quartile of exposure [215]. Prenatal exposure to organochlorine compounds has also been associated with decreased lung function later in life [216], risk of infection in childhood [217], attention deficit hyperactivity disorder [218], and obesity [219]. If these effects of elevated early life persistent organic pollutant exposure last throughout the life course, there could be continued adverse health consequences that manifest in those exposed for the foreseeable future.

While differences in chemical exposures likely explain most of the variations observed in the human body, these variations in biomarker levels may also be a result of differences in physiological processing. Specifically, genetic factors may control the metabolism of these chemicals. For example, the enzyme, cytochrome P450 2A6 (CYP2A6) controls the metabolism of nicotine into cotinine and subsequently into 3-hydroxycotinine. In a diverse cohort of current smokers, CYP2A6 activity was found to be the highest in Latinos, followed by White Americans, African Americans, Native Hawaiians, and Japanese Americans with relevant alleles varying significantly by race [220]. Furthermore, lower CYP2A6 activity was associated with lower

concentrations of nicotine metabolites and subsequentially lower risk of lung cancer, which may explain the relatively lower risk of lung cancer found in Japanese smokers [220]. As another example, single nucleotide polymorphisms (SNPs) in the arsenic (III) methyltransferase gene AS3MT have been shown to influence the metabolism of inorganic arsenic into monomethylarsonate and dimethylarsinate [221,222]. A SNP in AS3MT was found to be associated with a decrease in inorganic arsenic in Non-Hispanic Whites but an increase in African Americans, Hispanics, and Chinese American. However, these associations were not significant [223], which is likely due to low sample size by race/ethnicity [223]. These findings emphasize the need to conduct a GWAS on a larger multiethnic cohort to better understand how genetic variants mediate racial differences in metabolic activity. These examples elucidate a limitation of using NHANES, since data on chemical exposures are not available. Having both exposure and biomarker measurements for the same chemicals would enable the comparison of metabolic activity by race to understand whether chemical biomarker levels are better explained by the environment or by genetics. In addition, these findings also showcase the need for joint studies that incorporate both genetics and the environment to comprehensively characterize the genetic and environmental contribution to disease.

While there are substantial health disparities between races, there may be variations within the same race. For example, the frequency of triple negative breast cancer (TNBC) was highest in West African/Ghanaian and African American women and lowest in Ethiopian and White American women [224]. In addition, highest prevalence of ER-negative disease were found among African Americans and patients born in West Africa, while the lowest were seen among White Americans and patients born in East Africa [225]. In another study, the highest prevalence of triple-negative tumors was found in West African/Ghanaian, followed by African Americans, and then

White Americans [226]. Furthermore, women with TNBC tend to be of West African ancestry compared to women who do not have TNBC [226]. These results may highlight a gradient of risk for TNBC due to increased proportion of ancestry attributed to West Africa. Moreover, the DARC/ACR1 allele (commonly known as the "Duffy-null" variant) was associated with western sub-Saharan African ancestry and independently associated with likelihood of TNBC [226]. This variant was selected due to its role in resisting malaria, and thus this mutation was largely fixed in regions of African affected by malaria [227]. However, this variant has been identified as the genotype responsible for lower average white blood cell counts observed in patients with African ancestry and thus potentially also responsible for disparities in transplant rejection and other health conditions [228,229]. This genetic predisposition combined with environmental insults, whether due to chemical exposures and/or social factors such as long-term stress, health access, etc., may exacerbate the risk for having TNBC, thus showing the consequences of forced migration on descendent of West Africans even after several generations have passed. Hence, these examples highlight the needs to conduct studies that integrate genetic and environmental data to identify population, who have a genetic and environmental predisposition to disease. Such results also highlight a limitation of using NHANES as there is no greater resolution on the race/ethnicity variable to study susceptibility due to differing degree of ancestry. Hence, future studies can incorporate the restricted genetic data from NHANES as well as use other datasets such as the UK Biobank to better understand the genetic and environmental contribution to disease that are attentive to race.

Our study has important limitations. First, the cross-sectional nature of NHANES only allows biomarker measurement at one time point per individual. In addition, there are no available data on season, but should such data become available, then it would be interesting to study

chemical exposure trends across the seasons. Moreover, since the half-lives of the biomarkers assessed in this study are highly variable [185], the precision of estimates of long-term exposure largely varies across chemical family. Additionally, this study was not able to assess geographic variation in exposure. Others have identified that persistent organic pollutant exposures in the NHANES cohort varies geographically, with higher DDT metabolite concentrations in individuals residing in the West, and elevated PCB concentrations in individuals residing in the Northeast [230]. Future work is needed to precisely characterize exposure "hot spots," in order to design intervention studies to reduce exposure disparities. Our study also focused on identifying average differences in biomarker concentrations. By ignoring the extremes of these distributions, we have likely not considered individuals at greatest risk of developing adverse health outcomes. Similarly, our analyses were limited by low detection rates, with 182 chemicals not meeting our inclusion threshold of at least 50% detection in the study population. A more in-depth analysis of differences in detection frequency by race/ethnicity could identify additional chemicals with significant racial disparities. For chemical biomarkers measured in urine, variations in the concentration of urinary creatinine, used as a correction factor for urine dilution, potentially confounds our comparison of exposures between individuals of different races. This is because increased average concentrations of urinary creatinine have been identified for non-Hispanic Black individuals, relative to Mexican American and non-Hispanic White individuals [231]. While we adjusted for urinary creatinine as a covariate in our regression models, the still may be residual confounding. The large number of chemicals assessed also precluded an in-depth characterization of the various routes of exposure of individual chemicals – this is undoubtedly an essential future direction of research to develop strategies to eliminate exposure disparities. Finally, while we performed analyzed all chemical biomarkers available from NHANES 1999-2014, these chemicals only represent a small

proportion of the over 80,000 chemicals estimated to be used in commerce in the United States. Future studies could benefit from an unbiased metabolomics approach to identity disparities in chemical exposures which are not captured in NHANES.

3.6 Conclusions

The persistent health disparities between women of different races/ethnicities make understanding the etiological drivers of these disparities a pressing public health issue. A recent commentary highlighted a lack of knowledge regarding the molecular underpinnings of health disparities. It described how the vast majority of genome sequencing data had been generated in populations of European ancestry [232]. Environmental exposures, however, are hypothesized to be the major driving risk factors for a vast suite of complex diseases [12]. Even when genetic data has been generated in an equitable fashion, understanding gene-environment interactions and complex disease phenotypes will still require in-depth quantification of environmental exposures. In this study, we have comprehensively identified differences in biomarker of chemical exposure across women of various race/ethnic groups and across age groups. These findings can guide future efforts to understand chemical impacts on health disparities by helping to prioritize chemicals for assessment in epidemiological studies. Additionally, chemicals as identified as highly disparate here can be further prioritized for toxicological assessment relevant to disease outcomes of interest. Finally, these findings can inform public health interventions designed to reduce chemical disparities and promote health equity across the population.

3.7 Figures

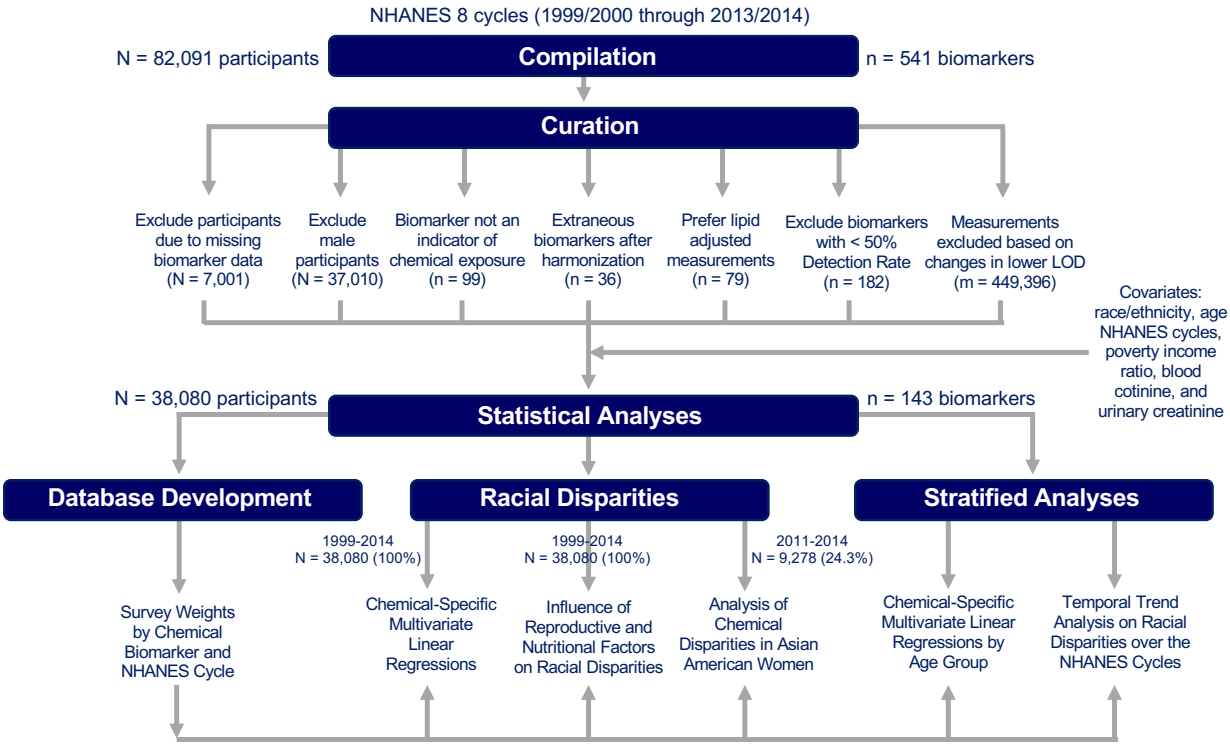

Figure 3.1 Dataset compilation and cleaning workflow.

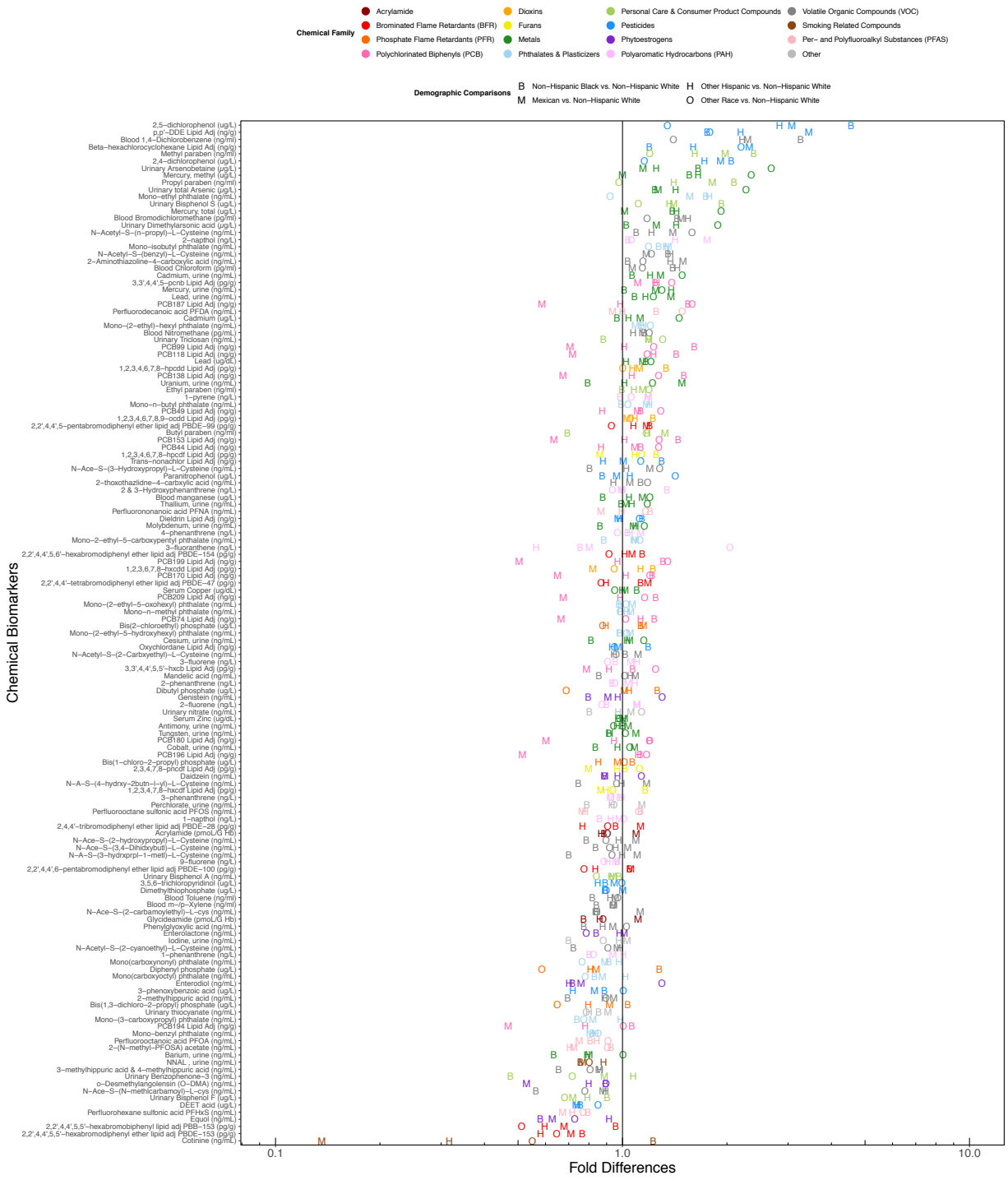

Figure 3.2. Alphabet soup plot displaying the covariate adjusted fold differences in chemical biomarker concentration by race, ranked by the average difference with non-Hispanic White individuals. Colors represent the chemical families. Shapes represent the comparison between a given race and non-Hispanic White individuals.

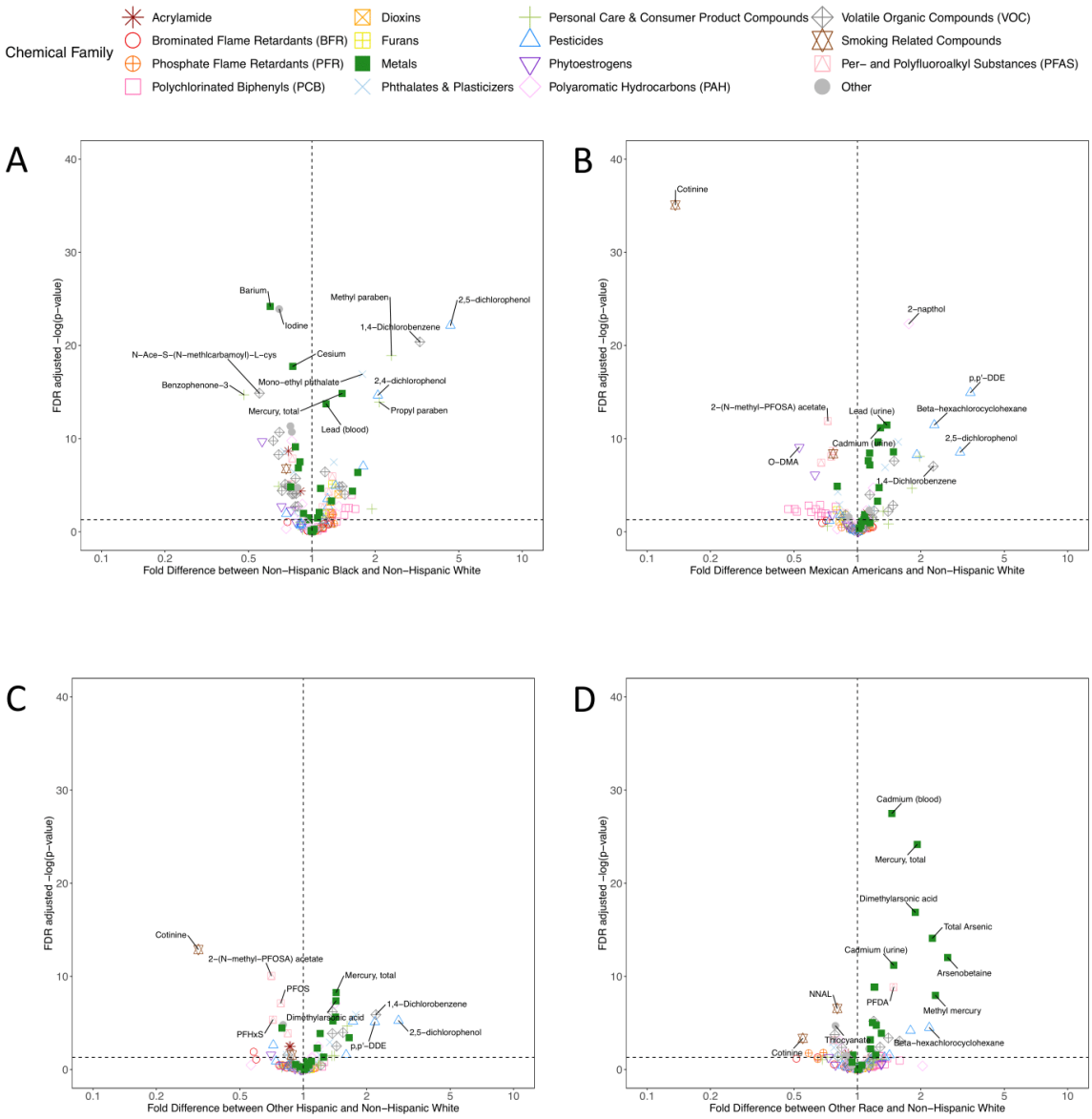

Figure 3.3. Volcano plots representing the significance of the covariate-adjusted differences in chemical biomarker concentrations between non-Hispanic White women and (A) non-Hispanic Black women, (B) Mexican American women, (C) Other Hispanic women, and (D) Other race/multiracial women. Color and shapes represent the chemical families.

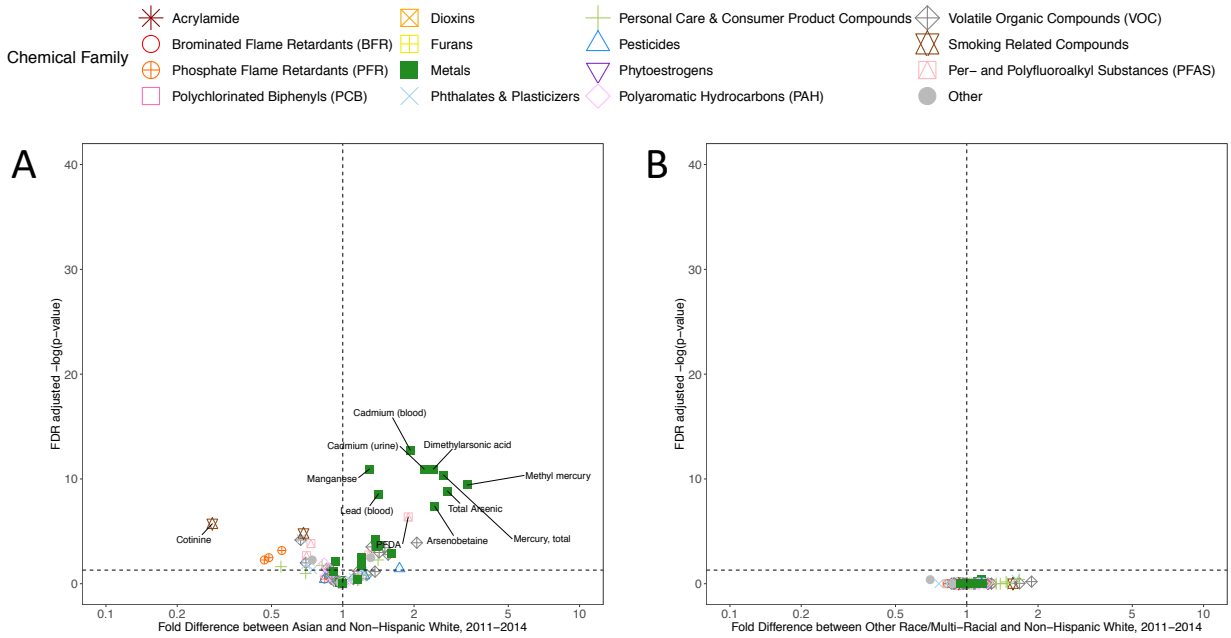

Figure 3.4. Volcano plots representing the significance of the covariate-adjusted differences in chemical biomarker concentrations between non-Hispanic White women and (A) Asian women, and (B) Other Race /Multi-Racial women in NHANES 2011-2014. Colors and shapes represent the chemical families.

Figure 3.5. Heatmap displaying covariate adjusted fold differences in chemical biomarker concentrations by race, relative to non-Hispanic White women, stratified by age group and chemical family. Color reflects the log2 fold difference in chemical biomarker concentration. Biomarkers in grey color were not measured in that age group.

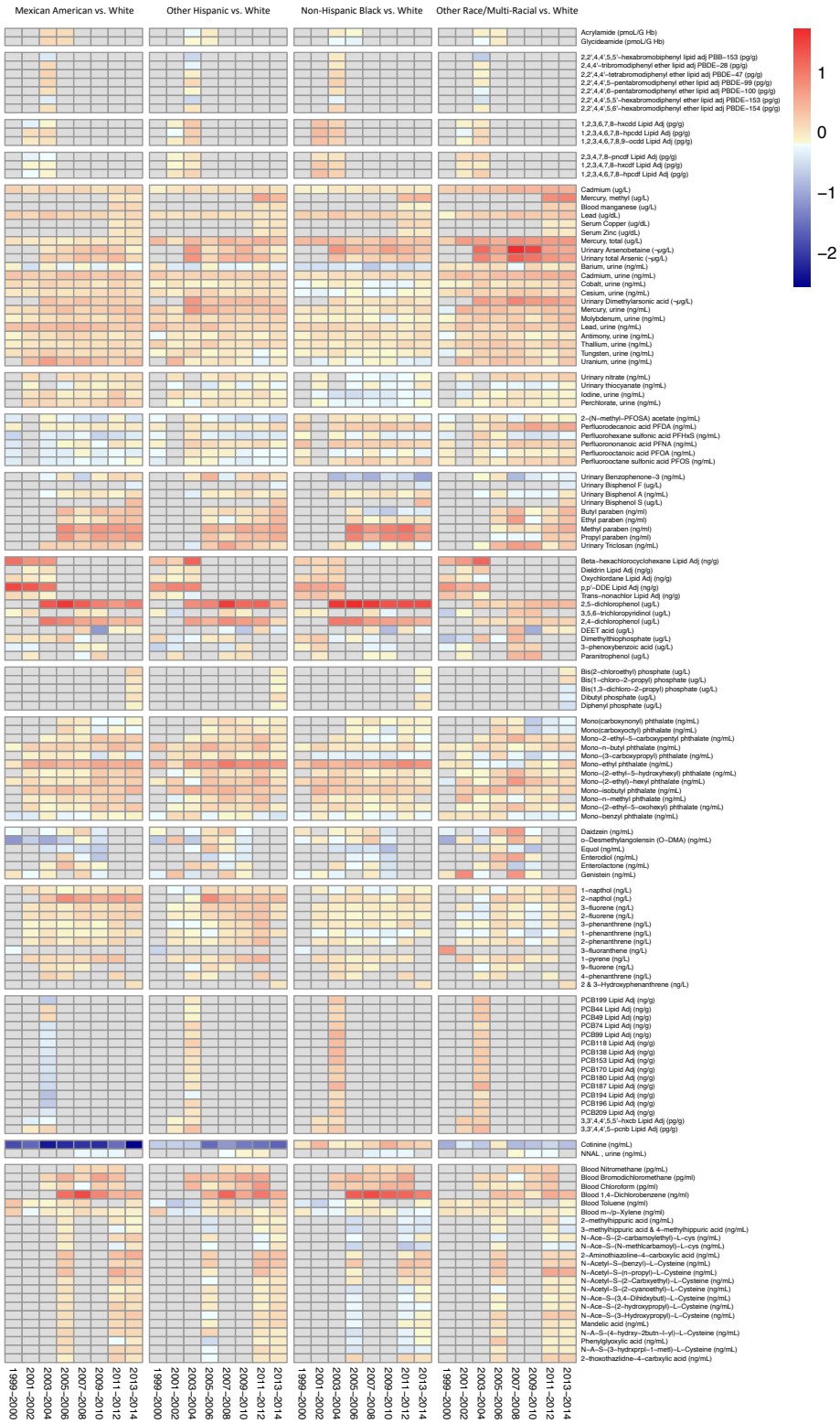

Figure 3.6. Heatmap displaying covariate adjusted fold differences in chemical biomarker concentrations by race, relative to non-Hispanic White women, stratified by study period and chemical family. Color reflects the log2 fold difference in chemical biomarker concentration. Biomarkers in grey color were not measured in that study period.

3.8 Tables

Table 3.1. Demographic characteristics of the study population.

CATEGORICAL					
Age	N (%)	Cycle	N (%)	Race/Ethnicity (%)	N (%)
0-11	9392 (24.66)	1999-2000 (Cycle 1)	4535 (11.91)	Mexican American	8760 (23.00)
12-25	9555 (25.09)	2001-2002 (Cycle 2)	5127 (13.46)	Other Hispanic	2949 (7.74)
26-50	9330 (24.50)	2003-2004 (Cycle 3)	4732 (12.43)	Non-Hispanic White	14384 (37.77)
51-85	9803 (25.74)	2005-2006 (Cycle 4)	4834 (12.69)	Non-Hispanic Black	9116 (23.94)
		2007-2008 (Cycle 5)	4628 (12.15)	Other Race	2871 (7.54)
		2009-2010 (Cycle 6)	4946 (12.99)		
		2011-2012 (Cycle 7)	4493 (11.80)		
		2013-2014 (Cycle 8)	4785 (12.57)		

CONTINUOUS					
	N measured (% of population)	5^{th} %tile	Median	Mean (SD)	95^{th0}%tile
Age (years)	38080 (100)	2	26	32.1 (24.2)	77
PIR (-)	34968 (91.83)	0.29	1.73	2.2 (1.6)	5.00
Cotinine (ng/mL)	31699 (83.24)	0.011	0.045	29.9 (91.4)	245.00
Creatinine (mg/dL)	32314 (84.86)	22.00	102.00	115.9 (76.6)	263.00

Chapter 4 Biomarker-Based Occupational Exposome

4.1 Abstract

Background: According to the World Health Organization, occupational exposures to hazardous chemicals may be responsible for over 370,00 premature annual deaths. Hence, this lends urgency to characterize the totality of exposures across different industries and occupations to identify workers most susceptible to chemical-mediated health effects.

Objectives: We 1) characterized differences in chemical exposures across 108 toxicants and physiological responses for 27 physiological indicators in a US sample of 26,361 blue- and white collar workers across 20 industrial sectors, 2) identified groups of workers with similar chemical exposure and physiological response profiles, and 3) evaluated differences in physiological measurements among the groups with similar chemical exposure profiles.

Methods: We applied a series of generalized linear models with the outcome as chemical biomarker concentrations or physiological measurements and the main predictor as sector-collar combinations with adjustment for age, sex, and race to characterize differences in chemical exposures and physiological responses. We applied hierarchical clustering with Pearson's correlation-based distance on the normalized regression coefficients to identify groups of sector-collar combinations with similar chemical exposures and physiological response profiles.

Results: White-collar workers had higher exposures to heavy metals such as mercury, arsenic, and their metabolites, along with a biomarker of sunscreen use, Benzophenone-3. Most blue-collar workers, especially those working in professional services, technical services, and

retail trade, have higher biomarker levels of heavy metals such as lead and cadmium, PAHs, PFASs, and VOCs such as toluene and benzene compared to their white-collar counterparts. Moreover, in blue-collar workers from the previously mentioned sectors along with agriculture, forestry, fishing and education services, alkaline phosphatase (biomarker of liver disease or bone disorder), white blood cell count (biomarker of inflammation), and glomerular filtration rate (biomarker of kidney function) showed higher values compared to white-collar workers. These results suggest that certain occupational exposures can diminish important physiological functions in blue-collar workers.

Discussion: We systematically characterized similarity in chemical exposures and physiological response profiles in a population of blue- and white-collar workers from a wide range of industrial sectors. Our findings could guide efforts to design targeted interventions to reduce health disparities in susceptible occupations.

4.2 Introduction

The exposome is the totality of all exposures in a lifetime and its impact on human health [2]. While environmental exposures have been heavily studied to characterize an individual's exposome, there needs to be more studies conducted on occupational exposures to help provide a more holistic characterization of an individual's exposome. Data from the World Health Organization suggest that exposures to hazardous chemicals in an occupational setting are responsible for over 370,000 premature annual deaths [177,233]. Despite this alarming statistic, a limited number of studies have characterized occupational exposures within the manufacturing sectors [30,35,37,234]. Many of these studies have applied a one-chemical or one-chemical-family approach to evaluate occupational exposures [30,35,37]. Few studies, however, aimed to analyze exposure to multiple chemicals across different industries and job titles, but they are based on

estimates of exposures [30] or air measurements at the workplace [234,235]. These studies are limited in addressing the extent of exposures within the human body. Thus, there is a need for a comprehensive, untargeted approach to study occupational exposures for a wide range of chemicals across a variety of occupations.

Since populations are not exposed to one chemical at a time but are exposed to several toxicants, there is a need to evaluate co-exposures of multiple chemicals. The biomarker dataset from the National Health and Nutrition Examination Survey (NHANES) has a wealth of blood, serum, and urine measurements on chemical contaminants, as well as data on occupation, but its sampling procedure creates challenges in quantifying co-exposures. Not all chemical biomarkers are measured in all study participants, rather different chemical biomarkers are quantified in different population subsamples [141]. For instance, in 2003-2004, participants in subsample A were measured for heavy metals and Per- and Polyfluoroalkyl Substances (PFAS) but were not measured for chemicals from other classes [236,237]. This means that analyses to quantify co-exposure in subsample A will be limited to only these two chemical classes. Furthermore, analyses will be limited to participants with measurements for all toxicants within in these two chemical classes. Due to this sampling procedure, NHANES is non-randomly sparse. This makes it especially challenging to use machine learning techniques to identify co-exposures as no participants have measurements for all chemicals. Imputation techniques are not applicable as these techniques work on the assumption that the dataset is randomly sparse [238]. Hence, studying co-exposures have been limited to one chemical family [93–95], but populations are not only exposed to one chemical class. Some studies even selected a candidate set of chemicals without assessing if this combination of chemicals is relevant in a susceptible population [239,240]. Very few studies, however, have managed to characterize co-exposure across different chemical classes [241,242]. Thus, characterizing co-

exposures across a wide range of toxicants is essential to better understand the chemical exposure profiles associated with vulnerable populations.

To begin understanding how occupational exposures may lead to premature death [177], we first need to understand the associations between chemical exposures and physiological responses. Physiological responses can be characterized by measurements of various biomarkers known as physiological indicators that reflect function of the different human body systems. Several studies have evaluated differences in a physiological response between worker populations [243–247], but they do not link such a response to chemical exposures. Some have studied the associations between a chemical or chemical family and a physiological indicator [30,35,96], but this is limited in addressing the "totality" of exposure on human health. Others have studied the impact of co-exposures on physiological response, but such approaches have been limited to only one physiological indicator [97]. Human health is not well-described by a single physiological indicator, rather it should be described by many different indicators to provide a holistic view of physiological function. Thus, evaluating the influence of chemical co-exposures on several physiological indicators is integral to gain insights on the influence of exposure on human health, especially in an occupational setting.

Overall, our goal is to characterize the chemical exposure and physiologic dysfunction in a working population comprised of NHANES participants from a wide range of industrial sectors and occupations. More specifically, our objectives are to 1) define differences in chemical exposure profiles based on sector-collar combinations, 2) group occupations based on their similarities in chemical exposure profiles, and 3) evaluate differences in physiological measurements among the clustered groups to identify workers at risk of physiologic dysfunction.

4.3 Methods

4.3.1 Study Population

Since 1999, the Centers for Disease Control (CDC) has conducted NHANES to collect cross-sectional data on demographic, socioeconomic, dietary, and health-related information in the US population. For this analysis, we combined data from the chemical biomarker, demographic, and occupational datasets between years 1999-2014 for an initial sample of 82,091 participants with 411 chemical biomarkers and 60 physiological indicators measured. We categorized participants as white- or blue-collar workers in their corresponding industrial sector by using the publicly available industrial and occupational codes and the US Department of Labor definition of blue-collar [248]. Blue-collar workers are defined as workers who perform repetitive tasks with their hands, physical skill, and energy. We tabulate the job occupation description and the collar category in **Table A3.1**. We then excluded participants for which corresponding data on chemical indicators (N = 7,001) and occupational descriptions (N = 48,729) were not available. We also excluded participants (N = 175) from the following sector-collar combinations as the median sample size across the included chemicals is less than 10 participants: blue- and white-collar workers from armed forces and white-collar workers from private household, mining, and utilities. Thus, the sample size of our studied population is 26,186 participants. These exclusion and inclusion criteria are further detailed in **Figure 4.1**.

4.3.2 Chemical Biomarkers of Occupational Exposures

We defined chemical biomarker as an indicator of environmental exposure that can be measured in blood, serum, or urine. We replaced all measurements below the limit of detection (LOD) with the LOD divided by the square root of 2, as recommended by the CDC [14] to produce reasonably unbiased means and standard deviations [132]. At times, NHANES identified a problem

of interference from molybdenum oxide that resulted in corrected concentration of urinary cadmium recorded as 0 ng/mL [249,250]. Log-transforming such data would be undefined, therefore such measurements were replaced with the LOD divided by the square root of 2 if the participant's urinary cadmium level was under the LOD or otherwise excluded. We excluded smoking and soy metabolites (**Table A3.2**) as biomarker levels of these substances are driven primarily by smoking or dietary behaviors instead of from occupational exposures (c = 8). We preferred lipid adjusted measurements for biomarkers indicated by 7- or 8-letter NHANES codename ending in "L" or "LA," respectively, for which NHANES provided both lipid-adjusted and non-lipid adjusted measurements. Therefore, we excluded non-lipid adjusted chemical biomarkers (c = 79). We calculated detection frequencies for each combination of chemical biomarkers and sector-collars. Using these detection frequencies, we also calculated the fraction of sector-collars with detection frequency above 50% for each chemical and excluded chemical biomarkers where this fraction was less than 0.5 (c = 142). Then, we calculated the sample size for each combination of chemical and sector-collar, determine the median sample size across all sector-collars, and excluded chemical biomarkers with a median sample size of less than 90 participants (c = 174). We chose 90 participants as the cut-off based on a power calculation to have a 70% chance of detecting a significance difference with a p-value of less than 5% and with a regression coefficient of 0.37 (**Table A3.3**). The final dataset for analysis consisted of 108 chemical biomarkers from 11 classes (**Figure A3.1**).

4.3.3. Indicators of Physiological Response

We identified 60 biomarkers and anthropomorphic measures that characterize physiological response. We excluded those with measurements in fewer than six NHANES cycles (p = 10) and with a sample size of less than 9,000 participants (p = 21). We also excluded

physiological indicators that are categorical (p = 2). We calculated the mean measurement for participants with two or more measurements for systolic and diastolic blood pressure. The final dataset for analysis consisted of 27 physiological indicators. Descriptions, clinical thresholds, and sample size of each physiological indicator are provided in **Table A3.4**.

4.3.4 Statistical Analysis

The non-random sparsity of the chemical biomarker dataset in the worker population creates challenges in applying machine learning techniques, such as hierarchical clustering to group individual workers together based on similarity in chemical exposure profiles. Applying machine learning techniques requires a complete dataset [238]. However, as no worker has data available for all studied toxicants (**Figure A3.2**), we cannot characterize the chemical profile for each individual worker. Instead, we addressed this sparsity issue by characterizing the profiles for each sector-collar combination to identify clusters of the sector-collar combinations with similar chemical exposure profiles. To further highlight the sparsity of the NHANES chemical biomarker dataset, we tabulated that number participants with a given number of measured chemical biomarkers in **Table A3.5**.

We performed all analyses using R version 3.6.0. We used multivariate regression models to evaluate differences in the chemical biomarker levels and physiological indicators across the sector-collar combinations. We conducted a series of generalized linear regression models with the log10 transformed chemical measurements or physiological measurements as the outcome variable and the main predictor as the sector-collar combination with the reference group as white collars from public administration. The selection of the reference group was based on *a priori* assumption that white collars from public administration would be exposed to toxicants at low doses. We adjusted for age (continuous), sex (categorical), and race (categorical). We assume age

to have a linear association with a given chemical biomarker levels due to ease of interpretation. For ease of interpretation, the regression coefficients for the sector-collar combinations were converted to percent differences $[10^{\text{coefficient}} - 1] \times 100\%$. To show exposures and response differences across the sector-collar combinations, we visualized the results of the regression models with two heatmaps of percent differences for all 108 chemicals and all 27 physiological indicators. To identify significant comparisons while maintaining a lower false positive rate, we used the False Detection Rate (FDR) method on the p-values of the regression coefficients pertaining to the sector-collar combinations (Benjamini and Hochberg, 1995).

We performed two series of hierarchical agglomerative clustering analyses to identify groups of sector-collar combinations with similar chemical exposures and physiological response profiles. First, we normalized the regression coefficients across a given chemical biomarker or physiological indicator with the mean equal to 0 and the standard deviation equal to 1. Normalizing will make the chemicals or physiological indicators comparable to each other. Normalizing also prevent features, e.g. chemical biomarker or physiological indicator, with the highest variation from driving the clustering [251]. Second, we used Pearson's correlation-based distance as our distance measure [252,253]. This distance measure calculates the similarity between the chemicals or physiological indicators of every pair of sector-collar combinations by using Pearson's correlation. We preferred Pearson's correlation-based distance to Euclidean distance as we are interested in identifying cluster of sector-collars with similar overall profiles [251] in chemical exposures or physiological response. Using Euclidean distance would cluster the sector-collar combinations based on magnitude of the regression coefficients, e.g. sector-collars with the highest chemical biomarker levels would cluster together. Third, we needed to decide which linkage method to use in order to best determine how the sector-collars combinations should cluster together based on

similarity in chemical exposures or physiological responses [251]. Hence, we tested four difference linkage functions: single, complete, average, and McQuitty's linkage method. We provided description of each linkage method in **Table A3.6**. Fourth, we selected the linkage method that generated the dendrogram that best preserves the dissimilarity between the sector-collar combinations. We evaluated this performance with the cophenetic correlation coefficient (**Table A3.7-A3.8**), which is a correlation coefficient between the original distance matrix and the distance matrix generated by the clustering configuration [254]. A cophenetic correlation coefficient equal to 1 implies that the clustering configuration perfectly preserves the dissimilarity between the objects, e.g. sector-collar combinations. Finally, we conducted hierarchical agglomerative clustering with 1500 multiscale bootstrap replicates to calculate the approximately unbiased (AU) p-values [255]. An AU p-value of 100 implies that all bootstrap replicates support the cluster, implying that the cluster is strongly supported by the data [255]. We visualized the clusters of sector-collar combinations with dendrograms. For ease of interpretation, we used Euclidean distance to cluster the chemical indicators based having similar profiles on occupational groups.

We performed two types of analysis to understand the impact of occupational chemical exposures on physiological function in the workers population. In the first method, we compare the chemical exposures profiles with the physiological response within the same sector-collar combination or clustered group. If the same sector-collar combinations clustered together due to having similar profiles in both chemical exposures and physiological responses, then this may indicate that occupational chemical exposures are drivers leading to physiological dysfunction. In the second method, we defined clustered occupational groupings based on similarity in chemical exposure profiles. We defined a clustered group based on the sample size ≥ 400 or AU p-value ≥ 80. To link occupational co-exposures to physiological responses, we used these chemical

exposure-based clustered groups as the main predictor in a series of generalized linear regressions with the outcome variable as a given physiological indicator and the covariates as age (continuous), sex (categorical), and race (categorical). We defined the reference group to include the white collars from public administration and other sector-collar combinations who have a similar chemical exposures profile. Hence, the reference group is comprised of white collars from the following sector-collar combinations: Manufacture: Durable Goods; Arts, Entertainment, Recreation; Information; Professional, Technical Services; Education Services; Finance, Insurance, Real Estate, Rental; Health Care, Social Assistance; and Public Administration.

4.4 Results

4.4.1 Study Population

Table 1 presents population characteristics for the 26,186 NHANES participants from 1999-2014. **Figure A3.3** presents the sample size for each sector-collar combination. **Figure A3.4**, **A3.5**, and **A3.6** show the percentage of categories for sex, race, and poverty income ratio, respectively, for each sector-collar combination. **Figure A3.4** shows that blue-collar jobs are primarily occupied by males, while females tend to work in private household, health care, and education. **Figure A3.5** shows that white-collar occupations are predominantly comprised of Non-Hispanic White participants, while blue-collar occupations in agriculture, fishing, forestry, construction, and mining blue collars are mostly comprised of Mexican Americans. **Figure A3.6** shows a socioeconomic gradient with white-collar workers having higher poverty income ratio compared to blue-collar workers

4.4.2 Chemical Exposure Profiles

Figure 4.2 displays a heatmap of the percent differences of chemical biomarker levels relative to white collars from public administration to show differences in chemical biomarker

profiles across the sector-collar combinations. The dendrogram of the sector-collars is defined based on using the average linkage function with Pearson's correlation-based distance. The white blocks indicate that chemical biomarker levels are the same between a given sector-collar combination and the reference group of white collars working in public administration. Red blocks represent higher positive percent differences to indicate higher biomarker levels in a given sector-collar combination compared to the reference group, while blue blocks represent higher negative percent differences to indicate lower biomarker levels in a given sector-collar combination or higher biomarker levels in the reference group. For instance, the percent difference of 3-fluorene between blue-collar workers in mining and the reference group is 213%, which implies that on average, biomarker levels of 3-fluorene in blue-collar workers in mining are more than three times higher compared to those of white-collar workers in public administration.

This heatmap also shows co-exposure patterns by sector-collar combination. For example, compared to white-collar workers in public administration, blue-collar workers in professional or technical services tend to have higher levels of some of the most toxic chemicals such as cadmium, lead, several polyaromatic hydrocarbons (PAHs), several phthalates such as metabolites of Di-2-ethylhexyl phthalate (DEHP), and volatile organic chemicals (VOCs), including toluene, benzene, and 2,5-dimethylfuran. On the other hand, this occupational group, on average, have lower levels of arsenic and mercury metabolites and benzophenone-3 (BP-3), a UV blocker used in sunscreen.

This heatmap also enables the comparison of chemical exposure profiles across the different sector-collar combinations. The chemical exposure profiles of blue-collar workers are similar to the profiles of other blue-collar workers than to their white-collar counterparts. Blue-collar workers from professional services, technical services, and retail trade have some of the highest biomarker levels of heavy metals, such as cadmium and lead, phthalates, PFASs, PAHS,

and VOCs, including xylene, ethylbenzene, toluene, benzene, and 2,5-dimethylfuran, but have lower levels of arsenic and mercury metabolites. Within the "Food Services" cluster, which includes blue collars from transportation, warehousing, other services, accommodation, and food services along with white collars from agriculture, forestry, fishing, accommodation, and food services, the phthalates signal is stronger in the workers from transportation, warehousing, accommodation, and food services, suggesting that these workers are relatively more exposed to phthalates. White-collar workers from health care and social assistance have the most similar chemical exposure profiles to that of white-collar workers from public administration as the percent differences across all studied chemicals are near 0, i.e. the blue and red boxes are faded.

4.4.3 Heavy Metals

Within the same family, the heavy metals display different exposure patterns. Lead and cadmium, on average, are higher in most blue-collar workers compare to the other white-collar workers, i.e. the red blocks are darker for right half of the heatmap compared to the left half. When we compared the distribution of blood cadmium concentrations across the sector-collar combinations and the NHANES population (**Figure 4.3**), we also confirmed that cadmium levels are predominately and significantly higher in most blue-collar workers. In contrast, several white-collar workers have lower cadmium levels that are even below the average cadmium levels in the NHANES population, which is labeled with a yellow boxplot. We observed a similar pattern in blood lead (**Figure A3.7**).

In contrast, metabolites of mercury and arsenic display the opposite exposure patterns to those of lead and cadmium. White-collar workers tend to have higher biomarker levels of mercury and arsenic metabolites, whereas blue-collar workers have lower levels of these metabolites. **Figure 4.4** shows that the total blood mercury levels of most blue-collar are substantially and

significantly lower compared to those of white-collar workers. Blue-collar workers from mining have the lowest levels of total mercury with their average being slightly lower than the average of the NHANES population. In **Figure A3.8**, total urinary arsenic levels are also substantially lower for blue-collar workers from utilities, mining, information, and public administration compared to white-collar workers but not significant.

4.4.4 Polyaromatic Hydrocarbons (PAHs)

The chemical class of PAHs show very similar exposure patterns to each other and with lead and cadmium. Blue-collar workers, on average, have significantly higher levels of PAHs as show with percent differences ranging from 0 to 213.1% (**Figure 4.2** and **4.5**). Blue-collar workers from mining show that highest difference of 1-napthol and 3-fluorene at 213.1%, suggesting that levels of 3-fluorene are over three times as high as biomarker levels found in white-collars working in public administration. Average 3-fluorene levels are highest in blue-collar workers, who levels are higher compared to average biomarker levels in the NHANES population (**Figure 4.5**). In contrast, most white-collar workers, on average, have some of the lowest biomarker levels of 3-fluorene with their levels being under the average of the NHANES population.

4.4.5 Benzophenone-3 (BP-3)

Based on the linear regression results, biomarker levels for BP-3, a biomarker of sunscreen and cosmetic use, are among the three highest in the reference group of white-collar workers from public administration compared to all other sector-collar combinations. But upon examining the distributions of BP-3 (**Figure 4.6**), we observed that the mean measurements of white-collar workers from public administration are lower compared to nine other sector-collar combinations. Biomarker levels of BP-3 for these sector-collar combinations are also higher compared to the average for the NHANES population. On the other hand, blue-collar workers, especially those in

mining, manufacture of durable goods, professional, and technical services, have some of the lowest mean and median concentration of BP-3, well below the average of the NHANES population.

4.4.6 Per- and Polyfluoroalkyl Substances (PFASs)

Like the heavy metals, there are different exposures patterns within the chemical class of PFASs. Within blue-collar workers in professional services, technical services, and retail trade, the highest percentage differences are found in perfluorobutane sulfonic acid (PFBS) and perfluorododecanoic acid (PFDA), while slightly higher differences are found in perfluoroheptanoic acid (PFHpA), perfluorooctane sulfonic acid (PFOS), and perfluorooctane sulfonamide (PFOSA). These chemicals are clustered in the center of **Figure 4.2**. Blue collars from utilities, agriculture, forestry, and fishing have a similar exposure profile to the previously mentioned blue-collar workers but at lower magnitudes. Differences between the previously mentioned sector-collar combinations and white-collar workers from public administration are minimal for perfluoroundecanoic acid (PFUnA), perfluorodecanoic acid (PFDA), perfluorohexane sulfonic acid (PFHxS), and perfluorooctanoic acid (PFOA), which are clustered at the bottom of **Figure 4.2**. Interestingly, biomarker levels of perfluorononanoic acid (PFNA) are higher in the reference group. In **Figure 4.7**, blue collar workers in professional services, technical services, retail trade, agriculture, forestry, and fishing, on average, have substantially and significantly higher average biomarker levels of PFDA compared to the reference group and the NHANES population.

4.4.7 Phthalates

Even within the same chemical family, individual phthalates metabolites show very different exposure patterns. Metabolites of DEHP are higher in blue collars working in

professional services, technical services, and retail trade, who show the highest positive percent differences. On the contrary, other phthalates are lower in the same occupational groups with mono-isononyl and mono-isobutyl phthalates having negative percent differences (**Figure 4.2**). This suggests that biomarker levels of mono-isononyl and mono-isobutyl phthalates for these occupational groups are lower than those found in the reference group. All phthalate metabolites are generally higher in blue- and white-collar workers from accommodation and food services compared to the reference group. The exposure difference for mono-(2-ethyl)-hexyl phthalate (MEHP) is substantial and only significant for blue-collar workers in professional, technical services at 61% (p-value = 3.8e-02) and retail trade at 38.2% (p-value = 3.3e-02). In **Figure 4.8**, average MEHP levels were the highest for blue-collar workers in private household, professional and technical services, and utilities along with white-collar workers in arts, entertainment, and recreation.

4.4.8 Differences in Physiological Stress Response

Figure 4.9 displays the percent differences of physiological indicators to show differences in physiological response profiles across the sector-collar combinations. The interpretation is similar to that for **Figure 4.2**.

This heatmap enables the comparison of physiological response profiles among the sector-collar combinations. Like the chemical exposure profiles, hierarchical clustering nearly split the sector-collar combinations into two large groups: one comprised mostly of blue-collar workers on the right of **Figure 4.9** and the other of most white-collar workers on the left. This implies that workers within the same collar class have similar physiological profiles. The profiles of the blue-collar workers are predominantly characterized by lower C-reactive protein levels, which is indicative of decreased inflammation. The exception is found in blue collars working in mining,

transportation, and warehousing as the percent difference in C-reactive proteins levels are +57% and +6.3%, respectively, relative to the reference group of white collars from public administration. Compared to the reference group, most blue-collar workers have lower values in glucose control indicators such as Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) and the ratio of insulin to glucose, higher measurements of alkaline phosphatase (a biomarker of liver disease or bone disorder), and higher measurements of glomerular filtration rate, which is indicative of increased kidney filtration function. In addition, most blue-collar workers have lower values of markers describing body fat deposition including subscapular skinfold, triceps skinfold, relative fat mass index, waist circumference, body mass index (BMI), and weight, which may be indicative of their physical fitness or the physical demand of their jobs.

The physiological profiles of the white-collar workers differ even within the same collar class. Values of HOMA-IR and ratio of insulin to glucose are lower or on par with those found in white collars from public administration. This contrast with white-collar workers from agriculture, forestry, fishing, and construction, who have higher values. Measurements of body fat deposition, alkaline phosphatase, GFR, and white blood cell count are very similar for most white-collar workers.

4.4.9 Alkaline phosphatase

Alkaline phosphatase is an enzyme found in blood that help break down proteins with important role in liver function and bone development [256]. Biomarker levels below 20-44 U/L may indicate hypophosphatasia, a rare genetic disease that affects bones and teeth, while levels above 116-147 U/L may indicate liver damage. Biomarker levels of alkaline phosphatase are significantly higher in several blue collars compared to the reference group and the NHANES population (**Figure 4.10**). Most of the workers in each sector-collar combination are within the normal range.

However, there is a high percentage of workers with biomarker levels outside the normal range for the blue collars, especially those working in agriculture, forestry, and fishing. On the other hand, the lowest average concentrations of alkaline phosphatase are found predominantly in white-collar workers, who averages are 3-7 U/L lower than that of the NHANES population.

4.4.10 Indicators of Inflammatory Response

White blood cells are cells of the immune system responsible for protecting the body against infectious disease and foreign invaders [257]. C-reactive protein (CRP) is a protein made by the liver where increased levels are indicative of a condition causing inflammation in the body [258]. After correcting for multiple comparisons, no results were significant for white blood cell count. CRP levels are significantly lower in white-collar workers from education services, professional, technical services, and arts, entertainment, recreation and in most blue collars as shown with negative percent differences. This is not the case for blue collars working in mining, however, who have the highest difference observed at 57.1% compared to the reference group of white collars from public administration. In **Figure 4.11**, while the unadjusted average mean concentration of CRP is not the highest for blue-collar workers from mining, 60.3% of these workers have CRP levels higher than the clinical threshold established at 0.1 mg/dL.

4.4.11 Indicators of Insulin Resistance

Insulin resistance occurs when the body's cells do not respond normally to insulin to control blood glucose levels [259]. On average, higher values of HOMA-IR are found in white collars from agriculture, forestry, fishing, and construction along with blue collars from information and private household, which is indicative of higher to insulin resistance. However, these results are not significant after adjusting for multiple comparisons. The only significant result was for blue collars working in construction, whose values of HOMA-IR are 18.8% lower compared to those

for white-collars from public administration. Interestingly, in **Figure 4.12**, 51% and 43.4% of the white collars from agriculture, forestry, fishing, and construction, respectively, along with 41.9% and 46.2% blue collars from information and private household, respectively, are above the clinical threshold for HOMA-IR. This suggests that approximately more than 40% of the workers in these occupational groupings are susceptible to insulin resistance.

4.4.12 Glomerular Filtration Rate (GFR)

GFR is a measure of kidney function with values below 90 mL/min/1.73 m^2 indicative of kidney damage and higher values indictive of hyperfiltration [260]. The effect sizes for several blue collars and only one white-collar group from transportation and warehousing remain significant after correcting for multiple comparisons. Interesting all these effect sizes are positive, indicating that average adjusted GFRs for these occupational groups are higher compared to the reference group. Moreover, less than 25% of the workers for these groups have GFRs indicative of kidney disease (**Figure 4.13**). The most susceptible occupational groups are the white collars from professional and technical services, agriculture, forestry, fishing, and construction, with at least 2.9% of the population with GFR below 60 mL/min/1.73 m^2.

4.4.13 Indicators of Body Fat Deposition

Indicators of body fat deposition include BMI, relative fat mass index, waist circumference, weight, subscapular skinfold, and triceps skinfold. Triceps skinfold is the thickness measured at the back side in the middle of the upper arm [261], while subscapular skinfold is measured under the lowest point of the shoulder blade [262]. The thickness of the skinfold is a measure for subcutaneous fat. Several blue-collar workers, on average, have lower measurements of subscapular and triceps skinfold compared to the reference group, whereas most white collars show similar profiles with the reference group except for those working in agriculture, forestry, fishing, and manufacture of

non-durable goods. Almost all sector-collar combinations, on average, have lower measurements of subscapular skinfold compared the reference group of white collars from public administration. Several of these differences are significant within the blue-collar class. The effect sizes are concurrent with the unadjusted mean measurements for subscapular skinfold as the average measurement for the reference group was the highest (**Figure 4.14**). Blue collars working in construction have the lowest average measurements of subscapular skinfold.

4.4.14 Influence of Occupational Exposure on Physiological Function

Comparing the chemical exposure profiles (**Figure 4.15**) with the physiological response profiles (**Figure 4.16**) within the same sector-collar combination or clustered group may provide insights into the influence of occupational exposures on eliciting a physiological response. Interestingly, both clustering procedures first split the sector-collar combinations by collar class, implying that chemical exposures profiles and the physiological response profiles are similar for workers within the same collar class. More specifically, some of the sector-collar combinations that show the highest averages of alkaline phosphatase levels are blue-collars from agriculture, forestry, fishing, retail trade, education services, and professional, technical services. These same sector-collar combinations are in the same cluster for having similar chemical exposure profiles, which is characterized by higher biomarker levels of PFASs, heavy metals, VOCs, PAHs, and phthalates. As another example, blue-collar in mining have higher biomarker levels of some of the most toxic chemicals such as heavy metals including cadmium and lead, PAHs, and VOCs including benzene and toluene. They also have the highest levels of CRP, which may suggest that such occupational exposures may elicit an inflammatory response.

As another method to understand the influence of chemical exposures on physiological function, we defined clustered occupational groupings based on similar chemical exposures

profiles (**Figure 4.17**) and compared the physiological response across these chemical exposure-based clustered groups while adjusting for age, sex, and race (**Figure 4.18**). Like the results from comparing the clustered groups between the chemical exposures and physiological response profiles, blue-collar workers have similar chemical exposure profiles as well as physiological response profiles. In addition, the physiological responses profiles for the clustered occupational groups also suggest that blue collar workers, compared to the reference group of white collars, have higher levels of alkaline phosphatase, white blood cell count, and GFR while lower measurements of triceps and subscapular skinfold, weight, creatinine levels, and indicators of glucose control such as HOMA-IR and ratio of insulin to glucose. On the other hand, indicators of body fat deposition such as BMI, waist circumference, weight, subscapular skinfold, and triceps skinfold are comparable between the reference group of white collars and two clustered groups: 1) an occupational group comprised of white collars from construction, management, administration, waste services, transportation, and warehousing along with blue collars from private household and 2) blue collars from public administration and white collars from wholesale trade. In contrast, HOMA-IR, the ratio of insulin to glucose, and cardiovascular indicators such as the ratio of total to HDL cholesterol, ratio of LDL to HDL cholesterol, triglycerides, and LDL cholesterols are higher in the previous clustered groups compared to the reference group.

4.5 Discussions

In this study, we present an unbiased approach to characterize difference in chemical biomarker levels and physiological measurements across a diverse suite of chemical contaminants and physiological indicators. To our knowledge, this is the first application of hierarchical clustering on differences by chemical exposures to identify groups of workers with similar chemical exposure profiles. We applied the same procedure with the physiological responses. We

compared the chemical exposure profiles with the physiological response profiles within the same sector-collar combination or clustered group to gain insights into the influence of occupational exposures on eliciting a physiological response. These results provide insight on differences of occupational exposures and responses across a wide set of industrial sectors. They are also informative for identifying which workers are susceptible to higher occupational exposures and potential adverse health outcomes.

We identified substantial differences in chemical biomarker levels across the different occupational groups. Blue-collar workers show higher biomarker levels of heavy metals such as lead and cadmium, PAHs, PFASs, and VOCs such as toluene and benzene compared to their white-collar counterparts. Higher lead levels found in blue collars from construction, manufacture, wholesale trade, retail trade, professional, and technical services may be due to how lead can be found in old and commercial paint, car parts, batteries, glass, and consumer products made of plastics [263]. In addition, this same group of workers may be exposed to cadmium via industrial uses of cadmium in making batteries, plating, pigments, and plastics [264]. Sources of occupational PAH exposures to this group may be due to engaging in tasks that involved combustion emission, such as welding, firefight, metal refining, vehicle repairs, etc. [265–267]. Higher biomarker levels of PFASs may be due to contact with products containing PFASs, which include food packaging, stain-resistant furniture, non-sick cookware, water repellant clothing, and medical and automotive application [268]. Similarly, higher VOCs levels may also be due to working with product containing VOCs such as building material, person care products, air fresheners, cleaning productions, fuel oil, and gasoline [269]. Higher VOCs [270,271] and PAHs [272] levels may also be due to smoking habits. Overall, higher biomarker levels of heavy metals, PAHs, PFASs, and VOCs in predominantly

blue-collar workers may be due to contact with products containing these chemicals and/or higher smoking prevalence.

On the other hand, most white-collar workers have higher levels of metabolites of arsenic and mercury along with a biomarker of sunscreen use, BP-3. While arsenic is used in many industries in the manufacture of paints, wood preservatives, agriculture chemicals, and in glass [273], it is less likely that higher biomarker levels of arsenic metabolites in white collars are due to occupational exposures. A similar argument can be made for mercury as industrial uses of mercury involve producing thermometers, barometers, batteries, and electrical switches [274], though health care workers may be exposed to mercury via medical or dental equipment [275]. Higher mercury and arsenic biomarker levels among these white-collar workers may indicate higher fish consumption [276]. This could therefore be a surrogate for behaviors associated with higher socioeconomic status [277] instead of an indicator of occupational exposures, since fish is expensive and is more accessible to those with higher socioeconomic status. It is also doubtful that white-collars are exposed to BP-3 due to their occupation as this chemical is used to manufacture agricultural chemicals, pharmaceuticals, plastic packaging, and household cleaning products [278,279]. Instead, as BP3 is used to prevent UV light from damaging scents and colors in personal care products [280], it more likely that higher levels of this chemical may suggest that cosmetics usage has a major role in strategic self-presentation, which may be integral in succeeding at the jobs of white-collar workers. Interestingly, unlike how the chemical biomarker levels of blue collars may be due to occupational exposures, the chemical exposure profiles of white collars may be indicative of behaviors or habits associated with their occupation and socioeconomic status.

The physiological response profiles also differ between white- and blue-collar workers. We observed that alkaline phosphatase was significantly higher in blue-collar workers compared

to their white-collar counterparts in a representative US population. These findings are supported by a study characterizing alkaline phosphatase levels in a representative Korean population where most blue-collar workers have serum alkaline phosphatase levels greater than or equal to 264 U/L for men and 252 U/L for women compared to unemployed and white-collar workers [281]. We observed an inflammatory response in blue collars via higher levels of white blood cell counts, and the signal was the strongest in mining blue-collars via CRP. This finding could be associated with occupational chemical exposures or with stress, as increased psychosocial job stress is associated with increased inflammatory markers [282,283]. The physiological response profiles of blue-collar workers are also characterized by elevated GFR, indicative of hyperfiltration or improved kidney function. This contrasts with other studies that finding the lower GFR is associated with blue-collar jobs [284,285]. Across markers of body fat deposition, these measurements were on average, significantly lower in blue collars from accommodation, food services, arts, entertainment, recreation, retail trade, construction, agriculture, forestry, fishing, and professional and technical services. Obesity prevalence has also been found to be lower in these occupational groups [286]. These findings suggest that markers of body fat deposition may serve as surrogates of physical activity [287,288]. We observed that the ratio of insulin to glucose was significantly lower in blue collars working in accommodation, food services, arts, entertainment, recreation, and construction compared to the reference group of white collars from public administration. While the higher ratios indicate insulin resistance in white collars from agriculture, forestry, fishing, and construction, the result was not significant. Interestingly, we observed that the workers with significantly lower ratio of insulin to glucose also have lower measurements on marker of fat deposition. This may suggests that insulin resistance is prevented by the physical activity required by the job for these blue-collar works [289]. We observed a healthy worker effect in the blue-collar

workers for indicators of body fat deposition and kidney function but not for alkaline phosphatase. These results emphasize that using a broad set of physiological enable a more holistic characterization of human health.

While we did not directly assess the impact of multiple chemical exposures on physiological response, we noticed potential links between the chemical exposures and physiological response profiles within the same occupational group. In our study, blue-collar workers in mining are exposed to some well characterized highly toxic chemicals including heavy metals such as cadmium and lead, the PAHs, and VOCs such as benzene and toluene. Furthermore, these workers also have a strong inflammatory response. Even after adjusting for age, sex, and race and correcting for multiple comparison, biomarker levels of C-Reactive Protein are 57.1% higher and significant compared to reference group of white collars from public administration. In fact, several studies have observed that cadmium exposures induce inflammation in mouse [290–293] even at non-toxic levels [294]. Similarly, lead exposures are also associated with inflammation [295–297]. One study assessed the association between CRP and blood lead levels in NHANES and found men are at increased risk of lead-induced inflammation than women [298]. Interestingly, majority of mining blue collars in NHANES are men, thus occupational exposure to heavy metals like lead may elicit an inflammatory response in mining blue collars. In addition, PAHs [299–302] and VOCs such as toluene [303,304] and benzene [305,306] are also known to elicit an inflammatory response individually and in combination [97]. Overall, our findings provide additional evidence to suggest that occupational co-exposures to lead and cadmium, PAHs, and VOCs may induce an inflammatory response in blue collar workers in the mining sector.

As another example to link chemical exposures to physiological response, blue collars from agriculture, forestry, fishing, retail trade, education services, and professional, technical services

have the highest levels of alkaline phosphatase and have a chemical exposure profile characterized by higher levels of heavy metals such as lead and cadmium, PFASs, VOCs, PAHs, and phthalates. Numerous studies show that heavy metals do induce liver damage [307,308], which results in higher levels of alkaline phosphatase. PFASs have also been associated with altered liver function [309] and liver damage and cancer [310]. Several studies have discovered significant associations between exhaled VOCs and liver disease [311,312], and thus exhaled VOCs can be used as diagnostic indicators of liver disease [313–315] Exposures to PAHs induce precancerous liver lesions and cancer in mice [316] and chronic exposure to PAHs has been associated with liver damage in humans [317]. Phthalates also modulate liver function [318,319]. These findings emphasize that co-exposures of heavy metals, PFASs, VOCs, PAHs, and phthalates may induce liver damage and may explain the elevated levels of alkaline phosphatase observed in blue collars working in agriculture, forestry, fishing, retail trade, education services, professional services, and technical services. Since excess alcohol consumption can causes liver damage, the higher levels of alkaline phosphatase found in blue collars may be a signal for alcohol intake. Thus, future studies can adjust for alcohol consumption to help disentangle the association between occupational exposures and liver disease.

The physiological profiles of these workers are also characterized by elevated GFR, indicative of hyperfiltration. Hyperfiltration has been linked to high exposures to cadmium [320] and lead [321,322]. Moreover, exposures to lead or cadmium have been associated with an initial increase in GFR and then a decline [320,321,323]. This suggests a physiological response to eliminate the toxicants and then the subsequent loss of kidney function. Reduced [324] and increased GFRs have also been associated with PFASs [325,326]. While reduced GFR is concerning as it indicates kidney damage, increased GFRs have also been associated with higher risk of cardiovascular morbidity and mortality [327,328]. Exposures to VOCs have also been associated with kidney dysfunction [329]

with higher exhaled levels of VOCs observed in critically ill patients with acute kidney injury compared to their counterparts with normal renal function [330]. But VOCs exposures have not been linked to hyperfiltration. PAHs may contribute to kidney dysfunction [331] by disrupting the balance between production of harmful free radicals and the ability of the body to counteract their harmful effects with antioxidants [332], leading to oxidative stress [333]. Cigarette smoking is associated with elevated GFRs [334], and as PAHs are present in tobacco products [272], exposures to this chemical class may also be linked to hyperfiltration. Increase in phthalates, particularly DEHP metabolites, was associated with kidney disease [335]. On the other hand, these phthalates are associated with elevated GFRs [336]. Overall, evidence suggests that exposures to heavy metals, PFASs, VOCs, PAHs, and phthalates are associated with both reduced and elevated GFRs and may vary based on the dose and duration of exposure. But it is unclear whether chemical-mediated hyperfiltration is a precursor of hypofiltration and subsequently a primary factor in kidney dysfunction. Thus, future studies can observe how chemical exposures influence the temporal trends of GFR.

Many studies using NHANES have been limited to studying co-exposures and its impact on human health in one chemical family [93–95]. Our framework enabled the identification of co-exposure across a wide range of chemicals not only limited to one chemical family. The primary limitation of this study is the non-random sparsity of the chemical biomarker data in the worker population. This sparsity prevented the application of hierarchical clustering on individual workers as we could not build a complete dataset with participants, who have measurements for all studied chemicals. While there are no participants with measurements for all 108 studied chemical indicators, there are participants with measurements for a subset of these substances as shown in **Table 2**. For example, there are 811 participants with occupational data and measurements available in at least 90 chemicals. Thus, future studies can determine the set of overlapping

chemicals among these participants and conduct clustering analysis on the raw chemical biomarker dataset. By observing the similarity of the clustering configurations between using the regression statistics versus using the raw data, these future analyses will help substantiate the findings of this paper. As another way to address this sparsity challenge, we conducted clustering analysis on exposure differences among the different sector-collar combinations, i.e. we applied clustering analysis on statistics of the biomarker data instead of on the raw data. A dataset of statistics derived from the biomarker data is usually complete. By applying hierarchical clustering on a dataset of statistics that represents exposure or physiological differences, we identified group of workers with similar exposure or physiological response profiles. This framework can be applied in other settings to help cluster observations based on similar profiles especially in a non-randomly sparse dataset. This can be done without having to form a complete dataset of impute the missing values.

However, our framework does have disadvantages. First, there is a substantial reduction in sample size, which will affect statistical power to detect significant differences. Second, there is a high chance that the dataset of statistics is singular, which makes applying machine learning techniques almost impossible. Our dataset of percent differences on chemical exposures was singular, due to biomarker levels of several VOCs being the same across all sector-collar combinations. A solution to this problem of having a singular matrix would involve removing these chemicals from the analysis. Finally, the main disadvantage of our framework is that we did not model the impact of co-exposures on physiological response. Thus, future studies can use statistics of a given physiological indicator as the outcome variable in a regression model with statistics of the chemical indicators as the main predictors to identify the most important set of chemicals responsible for explaining the variation of the physiological indicator.

The present study has several other limitations. First, the sample size for participants with occupational data is only 25% of that of the NHANES population. This 75% reduction in sample size is due to 50% of the participants agreeing to provide information on their occupation and then 50% of those participants responding that they are not working at a job or business. The analysis was limited by the small sample size, which prevented us from establishing a direct link between multiple chemical exposures and physiologic dysfunction. As NHANES includes more occupational data with each study year, future analysis will have increased statistical power to model the associations between chemical exposures and associated effects across different occupational groups. Second, we assume a linear association between age and chemical biomarker levels, which may not reflect how the chemical biomarker levels changes across the life-stage. Hence, future studies can determine which higher order polynomial would best describe the association between age and chemical biomarker levels to best account for the confounding effect of age on chemical biomarker levels. Future work can also use Gaussian kernel regressions to model the non-linear associations of chemical biomarker levels for age or other continuous variable and chemical biomarker levels. A Gaussian kernel regression model involves taking a weighted average of the surrounding data points to predict a given data point [337] Third, we did not correct for physical activity, which may explain why measurements of body fat deposition in lower in blue collar workers compared to their white-collar counterparts. Thus, further studies can account for physical activity in their models by using metabolic equivalents. Fourth, the strong signal of VOCs and PAHs observed in blue-collar workers may be due to smoking habits, which we did not adjusting for smoking in our models. Future analysis can conduct a sensitivity analysis to observe how the results changes when smoking is accounted. Fifth, a limitation of using chemical biomarker data is that a delay between the time of exposure and time of data collection

may prevent the detection of higher occupational exposures. This limitation is especially salient for VOCs, which have short half-life ranging from 2 to 128 hours [134,135]. Thus, differences observed between the occupational groups could be substantially higher if data collection occurred at the time of exposure, e.g. at the workplace. Hence, future directions can involve comparing biomarker levels of chemicals between workers who had their measurements taken on a workday versus those who were measured on an off day. Sixth, long-term stress and social factors such as social economic status, education, and health access may impact health as much or more than environmental exposures. In addition, these factors can modulate the effects of environmental exposures [338]. Thus, future work can involve accounting for these confounders as well as studying the interaction effects to better understand the contribution of chemical exposures to adverse health outcomes. In addition, to identify which environmental factors are most important in explaining a health endpoint, future directions can incorporate feature selection techniques such as LASSO (least absolute shrinkage and selection operator), adaptive elastic net, random forest and feature extraction techniques such as partial least squares and principal component regression. Finally, we assessed physiological response by using individual physiological indicators instead of pooling them together to better evaluate overall physiological dysfunction. Therefore, future studies can use the studied physiological indicators to define an allostatic load score to characterize the overall physiologic dysfunction [339–341] associated with multiple chemical exposures across the occupational groups.

4.6 Conclusions

Evaluating the influence of multiple occupational exposures on the physiological functioning is essential to begin understanding how chemical exposures elicit physiological response associated with adverse health outcomes. We applied an unbiased approach to screen

across 108 chemical indicators and 27 physiological indicators to characterize the chemical exposure and physiological response profiles across white- and blue-collars from 20 different industrial sectors. We developed a framework using hierarchical clustering on differences of chemical biomarker levels and physiological measurements to identify clusters of workers with similar chemical exposure and physiological response profiles. This framework enabled us to identify similar clustering between chemical exposure and physiological response profiles to characterize the indirect impact of multiple chemical exposures on physiological response in the workers population. Our framework enabled 1) comprehensive characterization of chemical exposures and physiological responses across a wide variety of occupations, 2) evaluation of occupational co-exposures patterns, and 3) identification of occupations susceptible to high exposure and physiological dysfunction. These findings can also guide efforts to design targeted interventions to reduce health disparities in susceptible occupations.

4.7 Figures

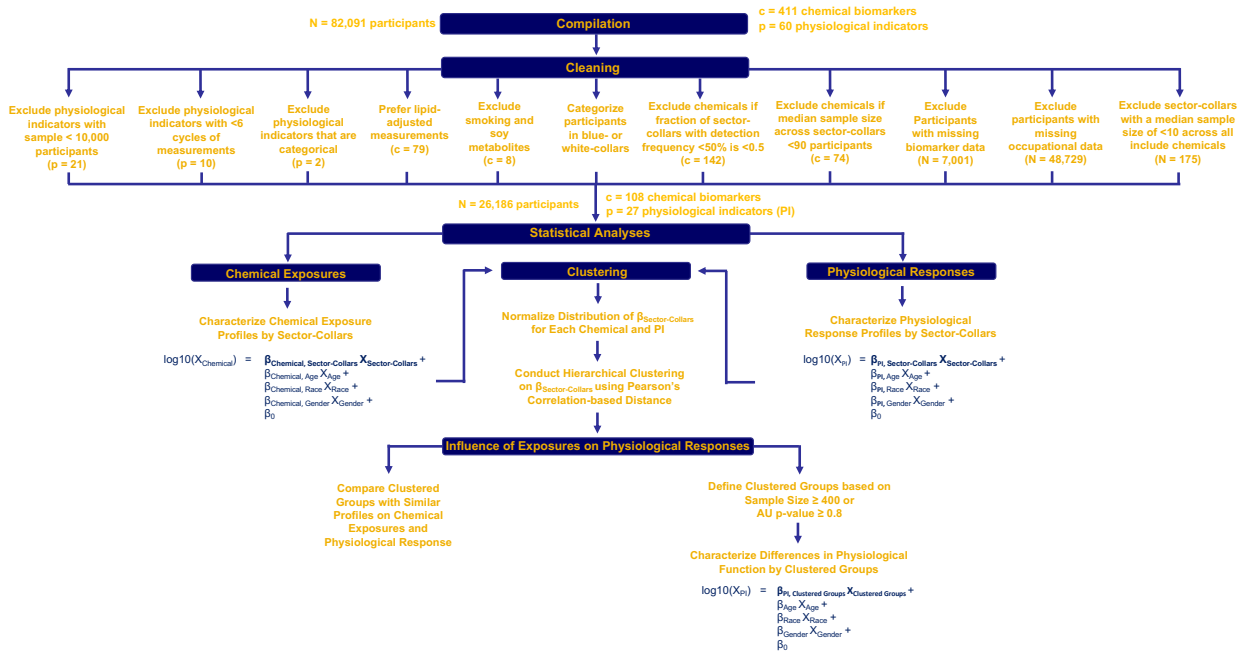

Figure 4.1. Schematic description on curation of chemical biomarker and physiological measurements and of the analytical methods used to characterize occupational variations in chemical exposures and physiological responses. Reference group for the analysis on the sector-collar combinations is White Collars from Public Administration.

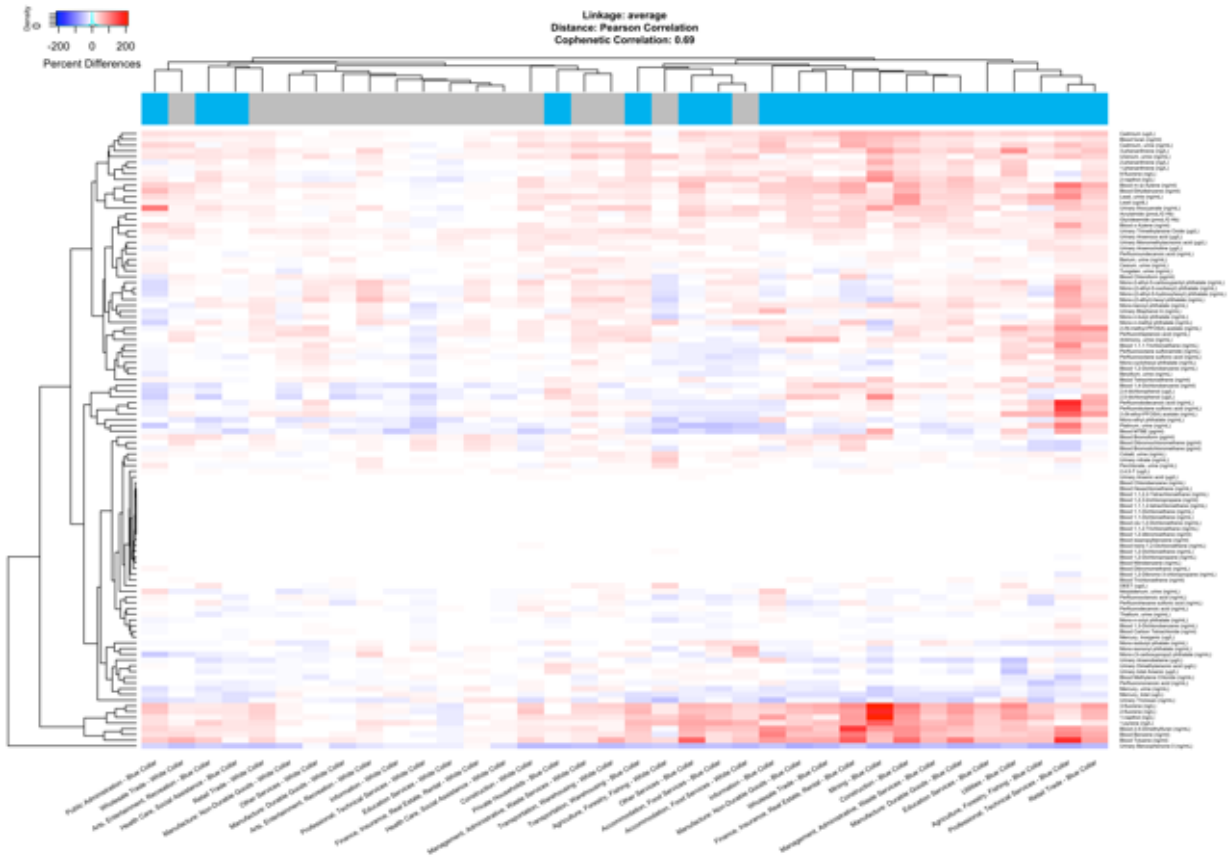

Figure 4.2. Heatmap of percent differences in chemical biomarker concentrations by sector-collar combinations, relative to Public Administration - White Collars. Chemical biomarkers in white color indicates that the concentrations are the same between the given sector-collar combination and the reference group. The color bar represents the collar categorization. Blue presents the blue-collar workers, while gray represents the white-collar workers. Results are adjusted for age, sex, and race.

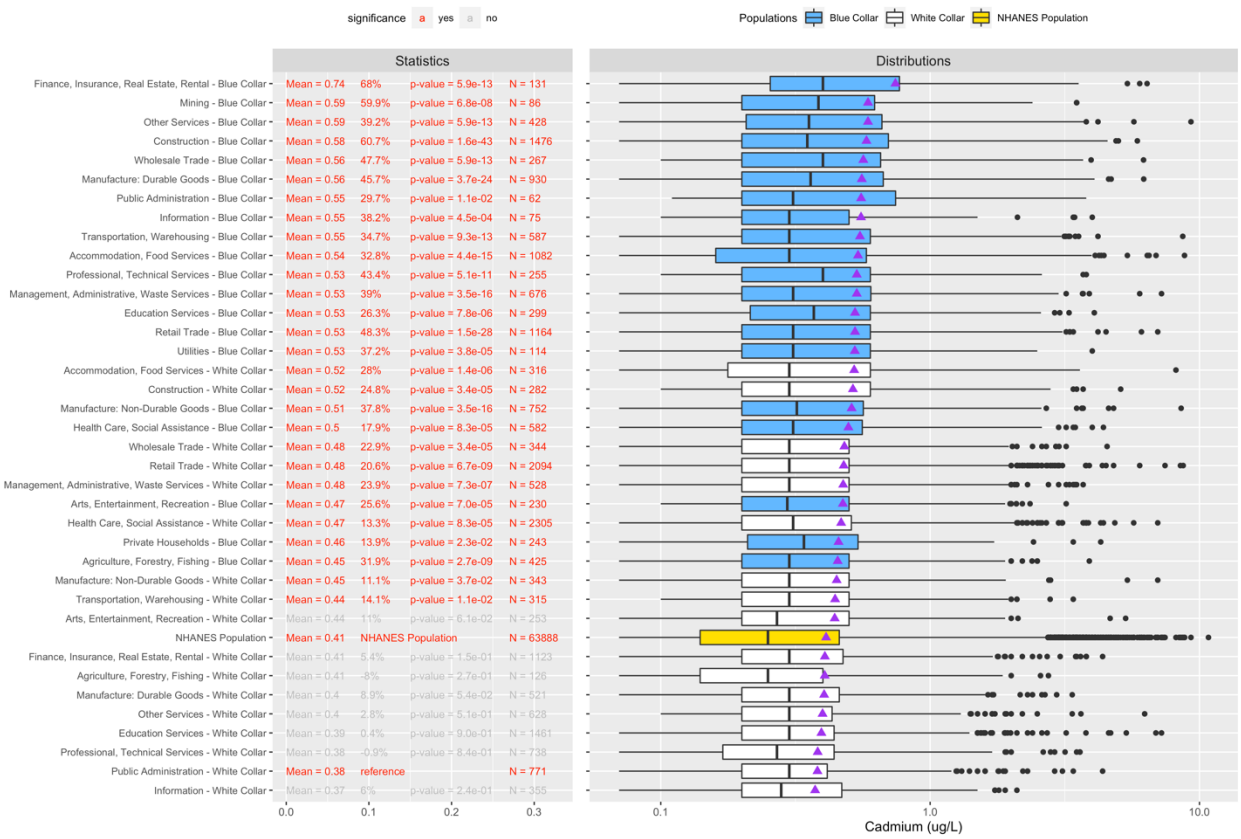

Figure 4.3. Box plot of distribution of blood cadmium. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for cadmium, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of blood cadmium for a given sector-collar combination. Results are adjusted for age, sex, and race.

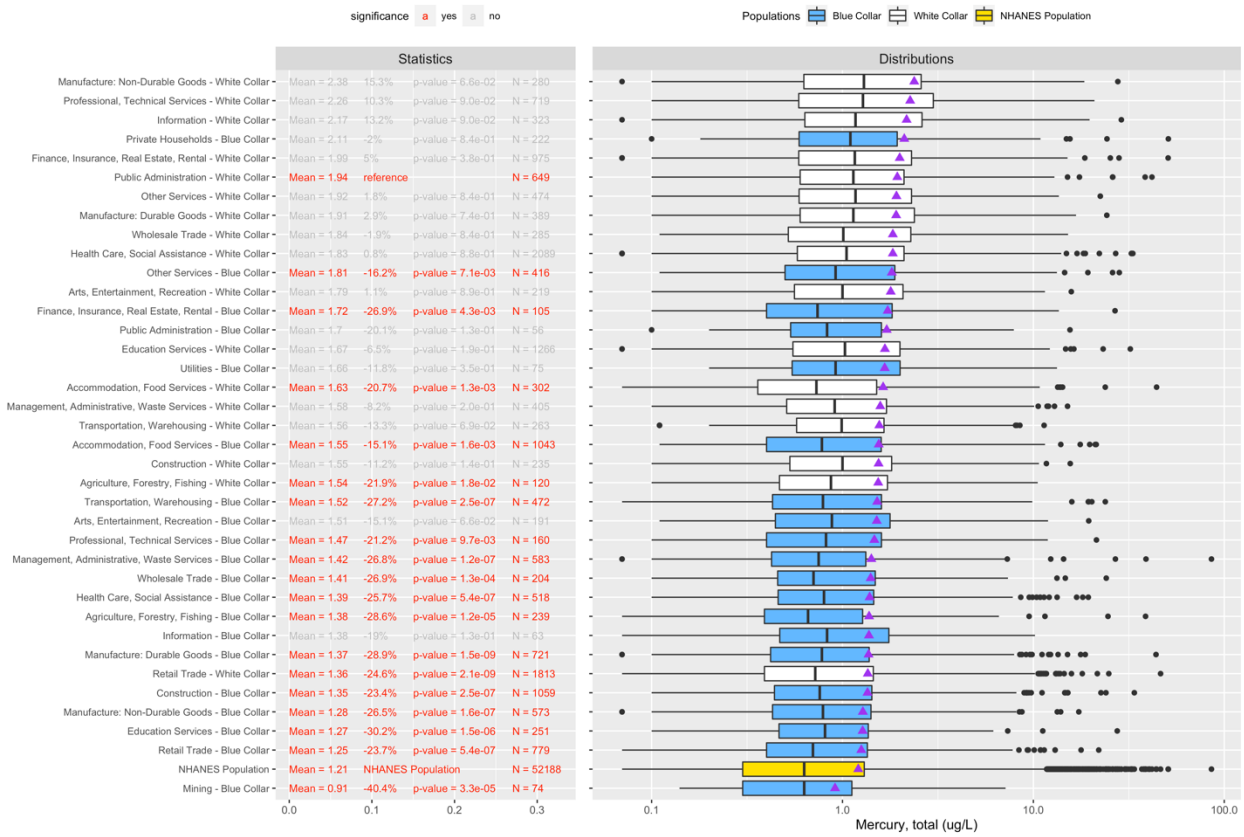

Figure 4.4. Box plot of distribution of total mercury in blood. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for cadmium, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of total mercury for a given sector-collar combination. Results are adjusted for age, sex, and race.

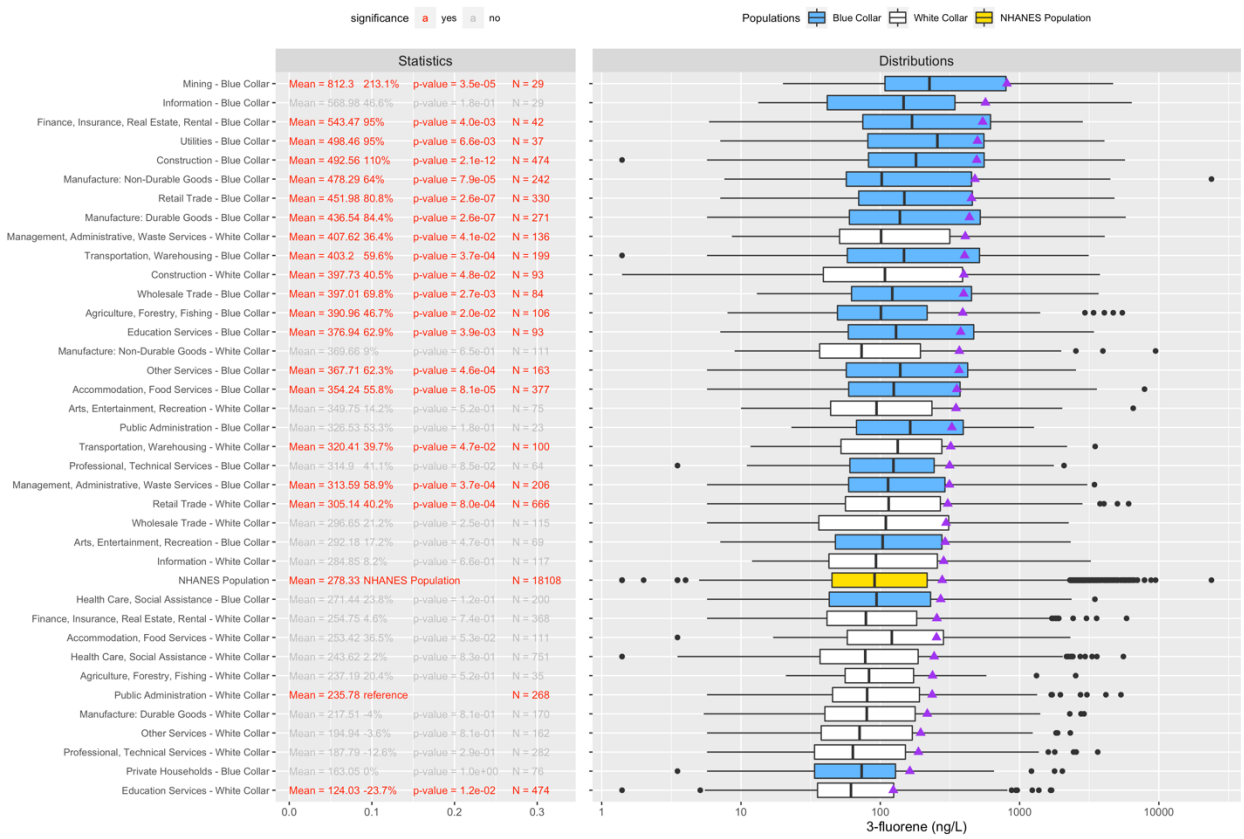

Figure 4.5. Box plot of distribution of urinary 3-fluorene. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for cadmium, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of urinary 3-fluorene for a given sector-collar combination. Results are adjusted for age, sex, and race.

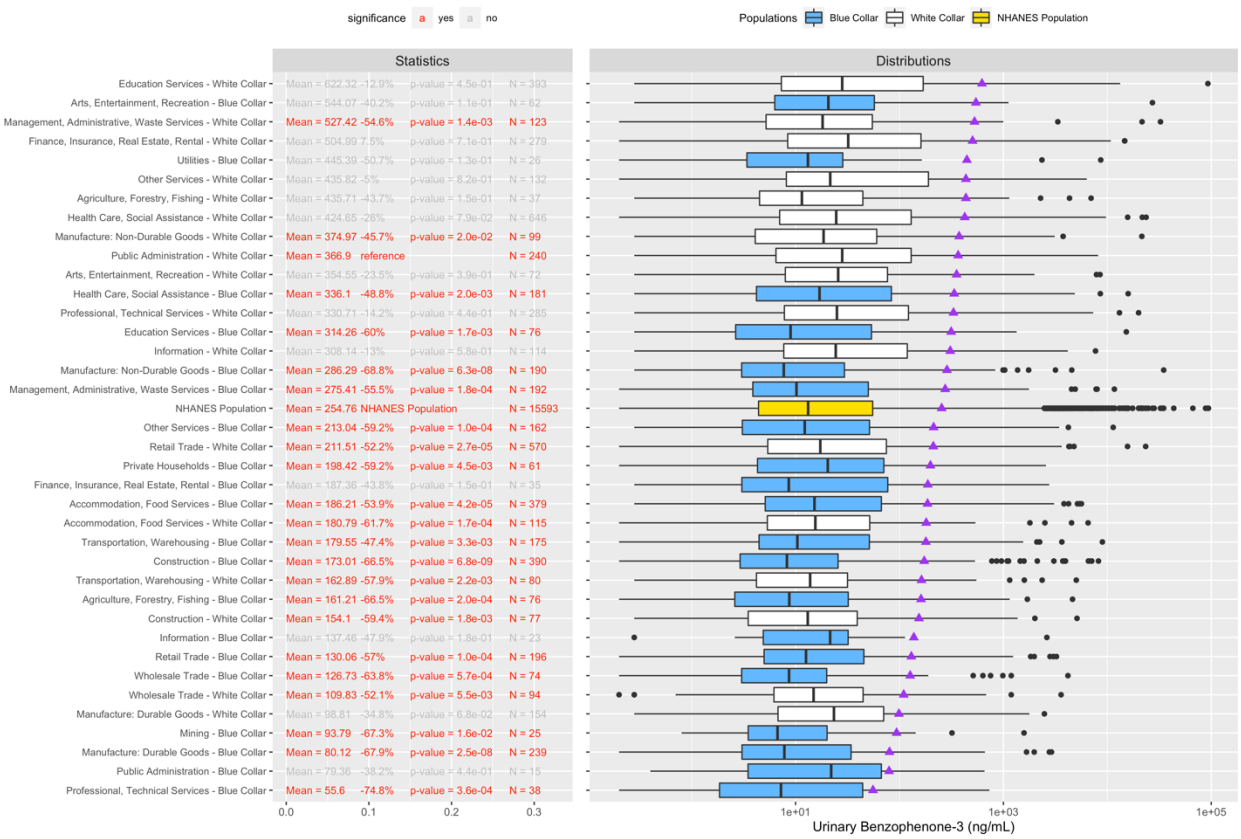

Figure 4.6. Box plot of distribution of urinary benzophenone-3 (BP-3), a biomarker of sunscreen use. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for cadmium, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of urinary BP-3 for a given sector-collar combination. Results are adjusted for age, sex, and race.

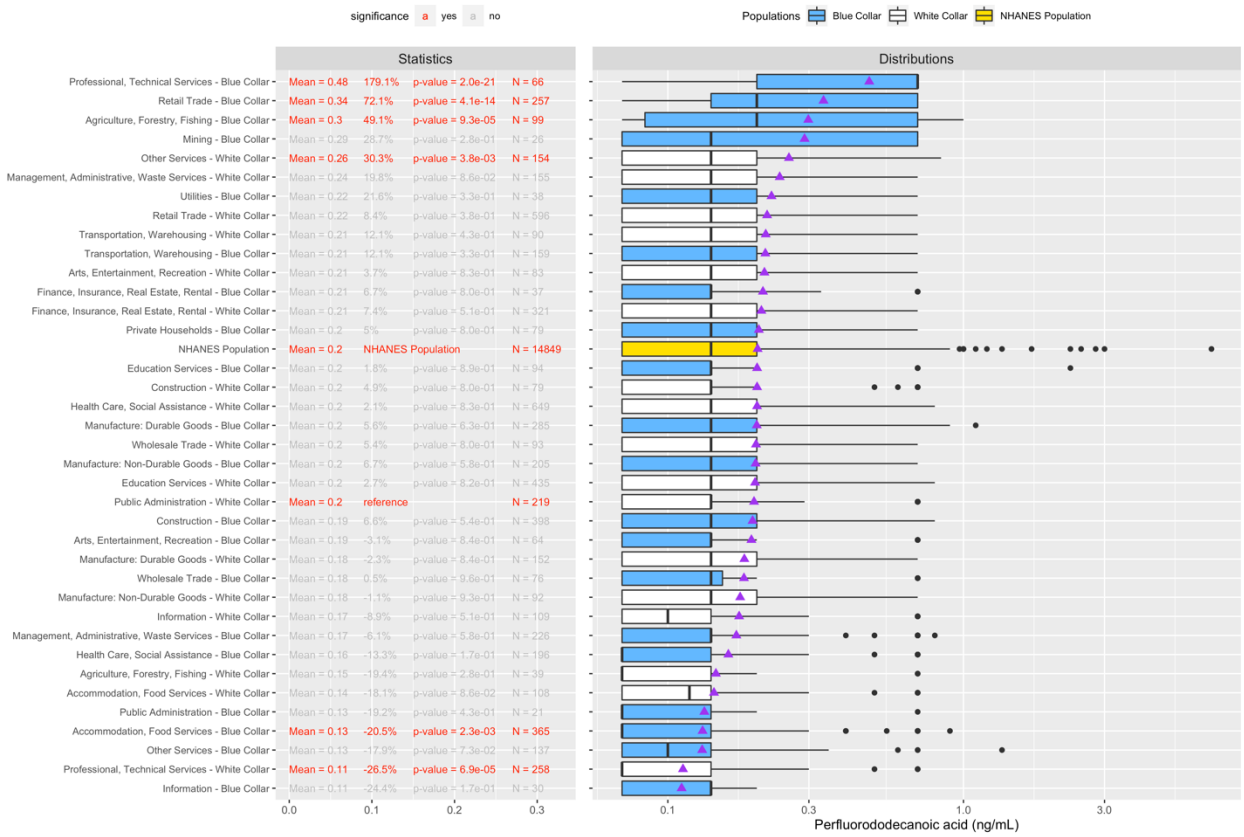

Figure 4.7. Box plot of distribution of perfluorododecanoic acid (PFDA). Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for cadmium, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of PFDA for a given sector-collar combination. Results are adjusted for age, sex, and race.

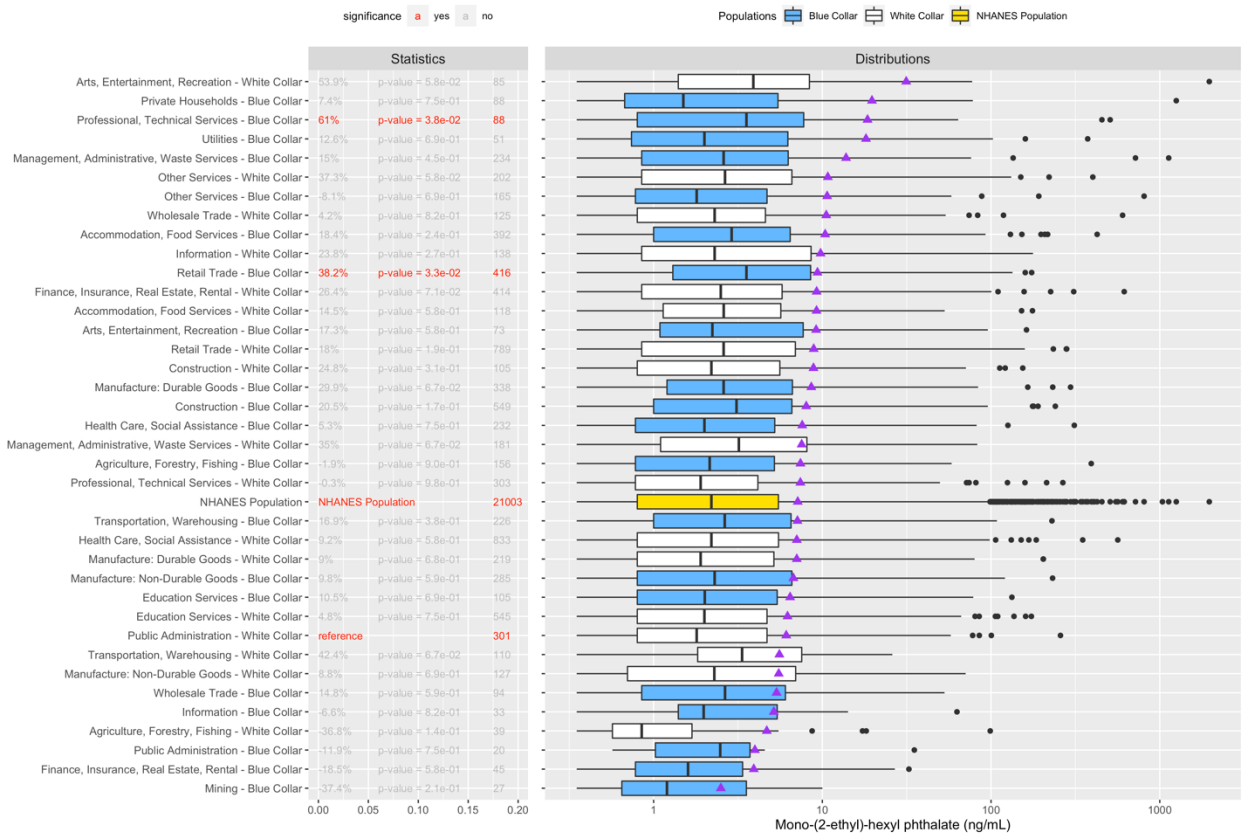

Figure 4.8. Box plot of distribution of urinary mono-(2-ethyl)-hexyl phthalate, a metabolite of the plasticizer Di-2-ethylhexyl phthalate, DEHP. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for cadmium, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of urinary mono-(2-ethyl)-hexyl phthalate for a given sector-collar combination. Results are adjusted for age, sex, and race.

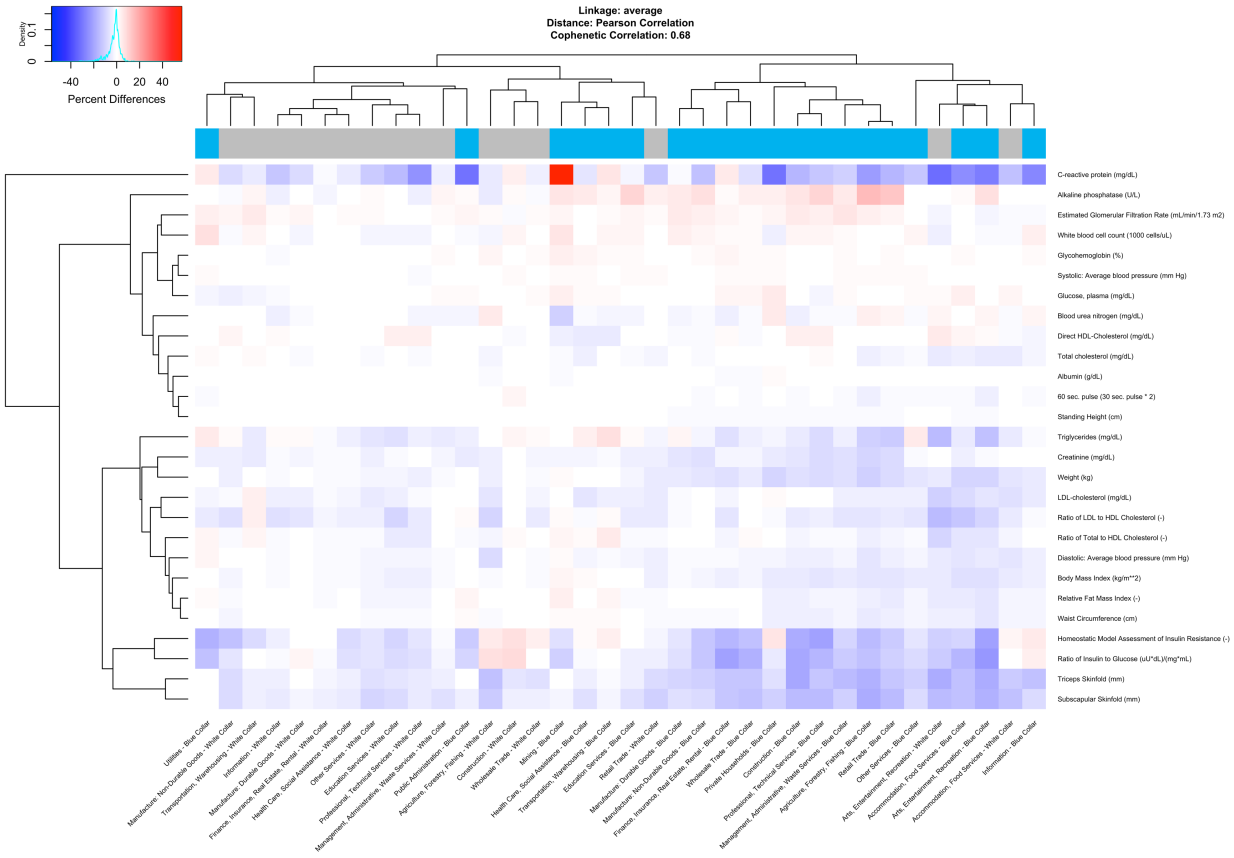

Figure 4.9. Heatmap of percent differences in physiological indicator measurements by sector-collar combinations, relative to Public Administration - White Collars. Physiological indicators in white color indicate that the measurements are the same between the given sector-collar combination and the reference group. Results are adjusted for age, sex, and race.

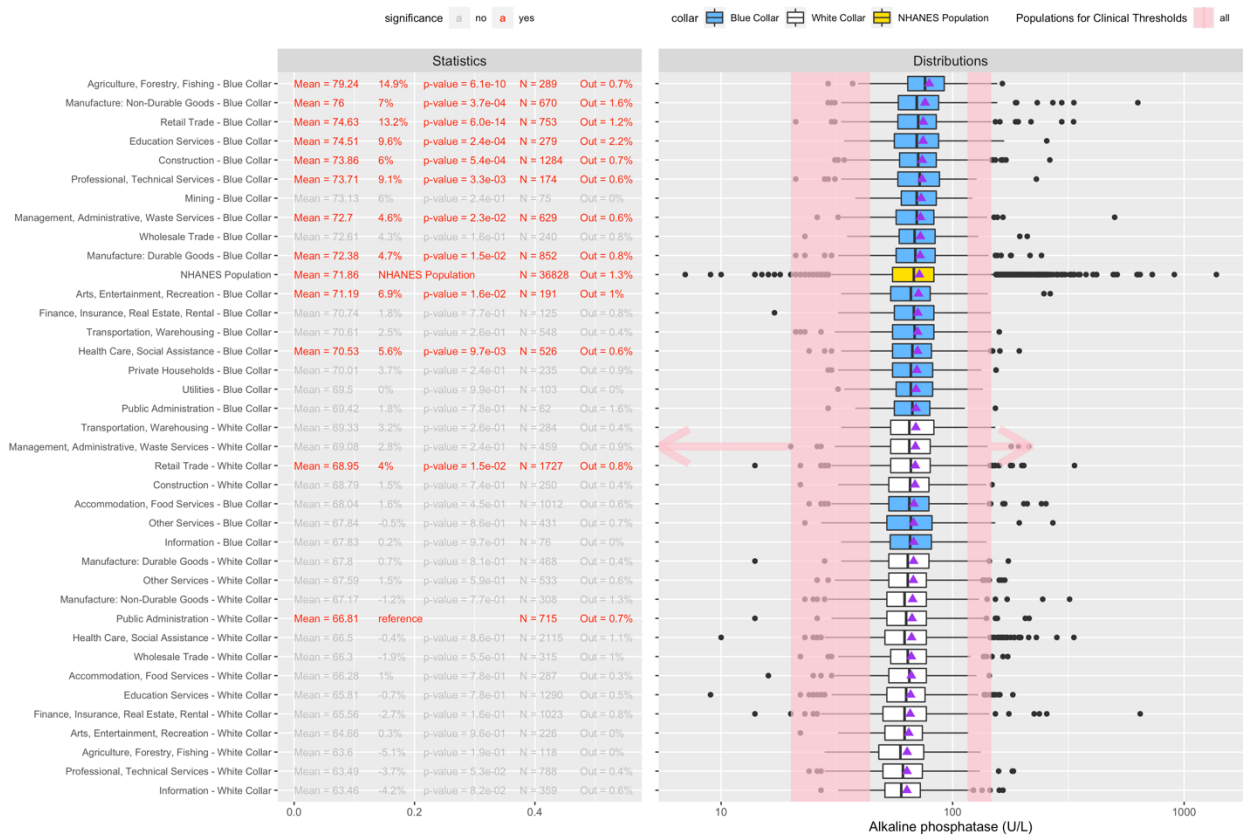

Figure 4.10. Box plot of distribution of alkaline phosphatase. Far-left statistics are the mean concentrations. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for alkaline phosphatase, including the sector-collar combinations. The middle statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. The middle-right statistics are the sample size of each sector-collar combinations. The far-right statistics are the percentage of participants outside of the range of normality, i.e. percentage of participants whose measurements are beyond the pink boxes. Purple triangle represents the mean concentration of alkaline phosphatase for a given sector-collar combination. Results are adjusted for age, sex, and race.

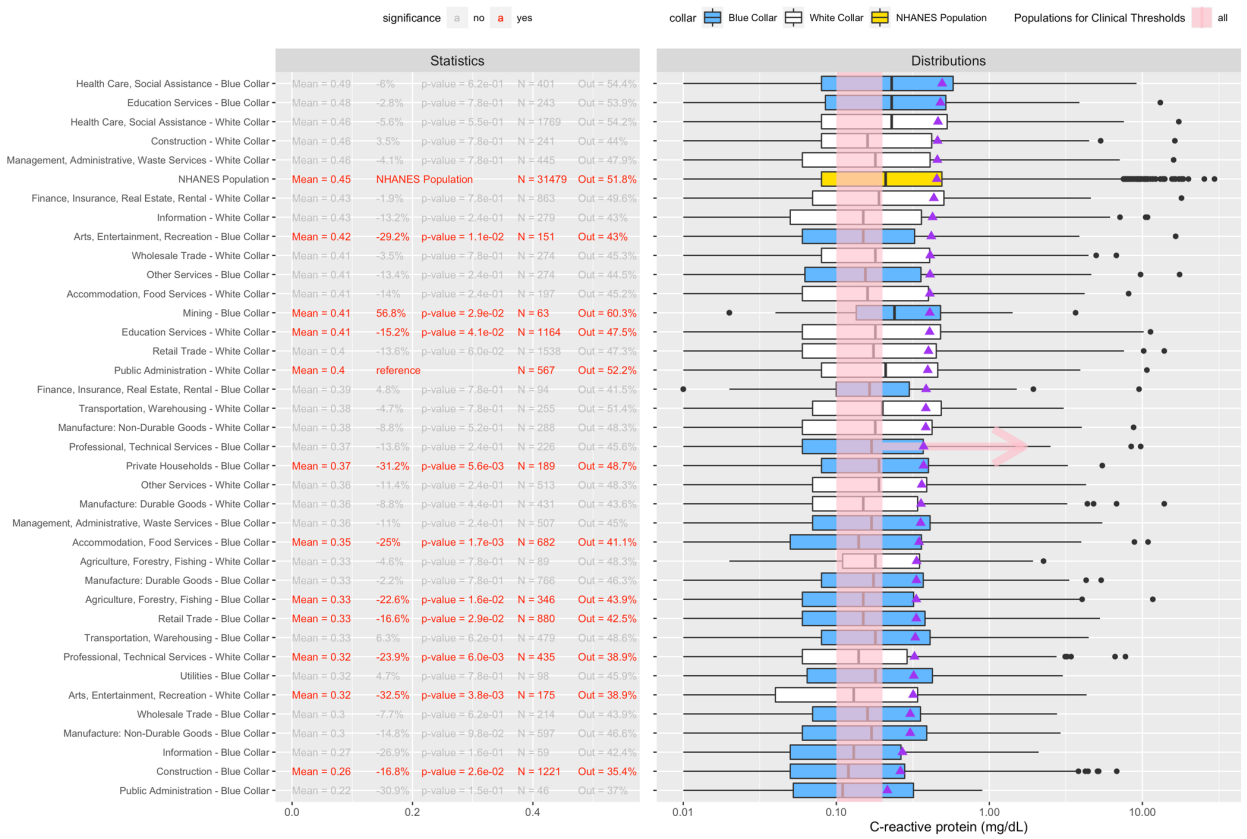

Figure 4.11. Box plot of distribution of C-reactive proteins (CRP). Far-left statistics are the mean concentrations. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for alkaline phosphatase, including the sector-collar combinations. The middle statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. The middle-right statistics are the sample size of each sector-collar combinations. The far-right statistics are the percentage of participants outside of the range of normality, i.e. percentage of participants whose measurements are beyond the pink boxes. Purple triangle represents the mean concentration of CRP for a given sector-collar combination. Results are adjusted for age, sex, and race.

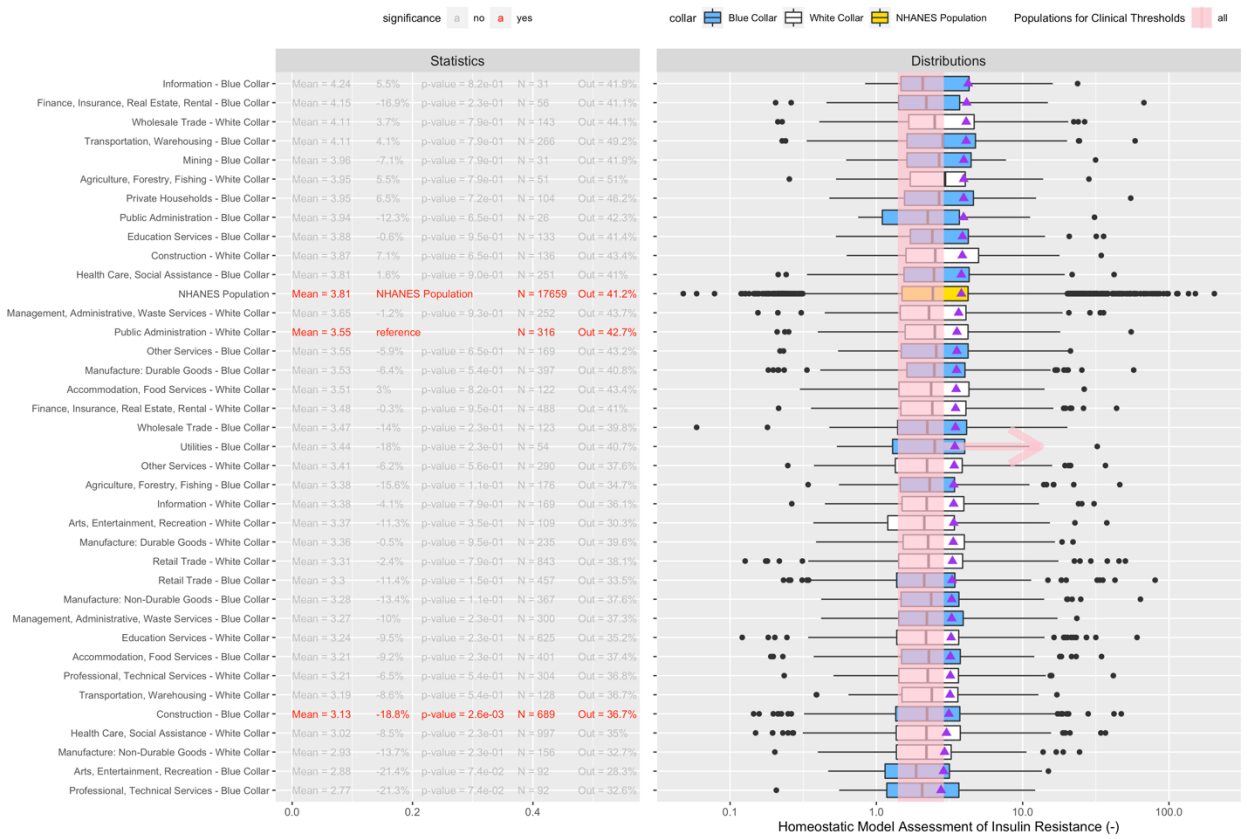

Figure 4.12. Box plot of distribution of Homeostatic Model Assessment of Insulin Resistance (HOMA-IR). Far-left statistics are the mean concentrations. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for alkaline phosphatase, including the sector-collar combinations. The middle statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. The middle-right statistics are the sample size of each sector-collar combinations. The far-right statistics are the percentage of participants outside of the range of normality, i.e. percentage of participants whose measurements are beyond the pink boxes. Purple triangle represents the mean concentration of HOMA-IR for a given sector-collar combination. Results are adjusted for age, sex, and race.

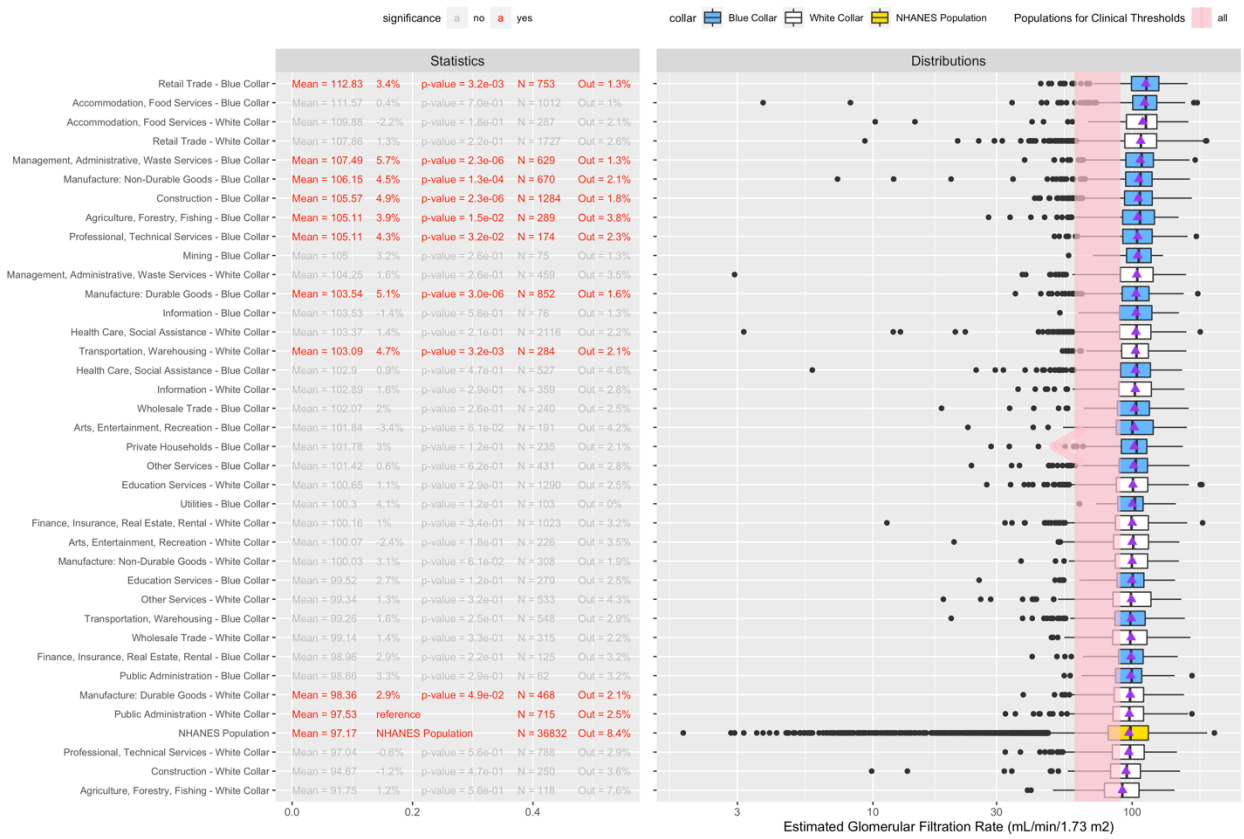

Figure 4.13. Box plot of distribution of glomerular filtration rate (GFR). Far-left statistics are the mean concentrations. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for alkaline phosphatase, including the sector-collar combinations. The middle statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. The middle-right statistics are the sample size of each sector-collar combinations. The far-right statistics are the percentage of participants outside of the range of normality, i.e. percentage of participants whose measurements are beyond the pink boxes. Purple triangle represents the mean concentration of GFR for a given sector-collar combination. Results are adjusted for age, sex, and race.

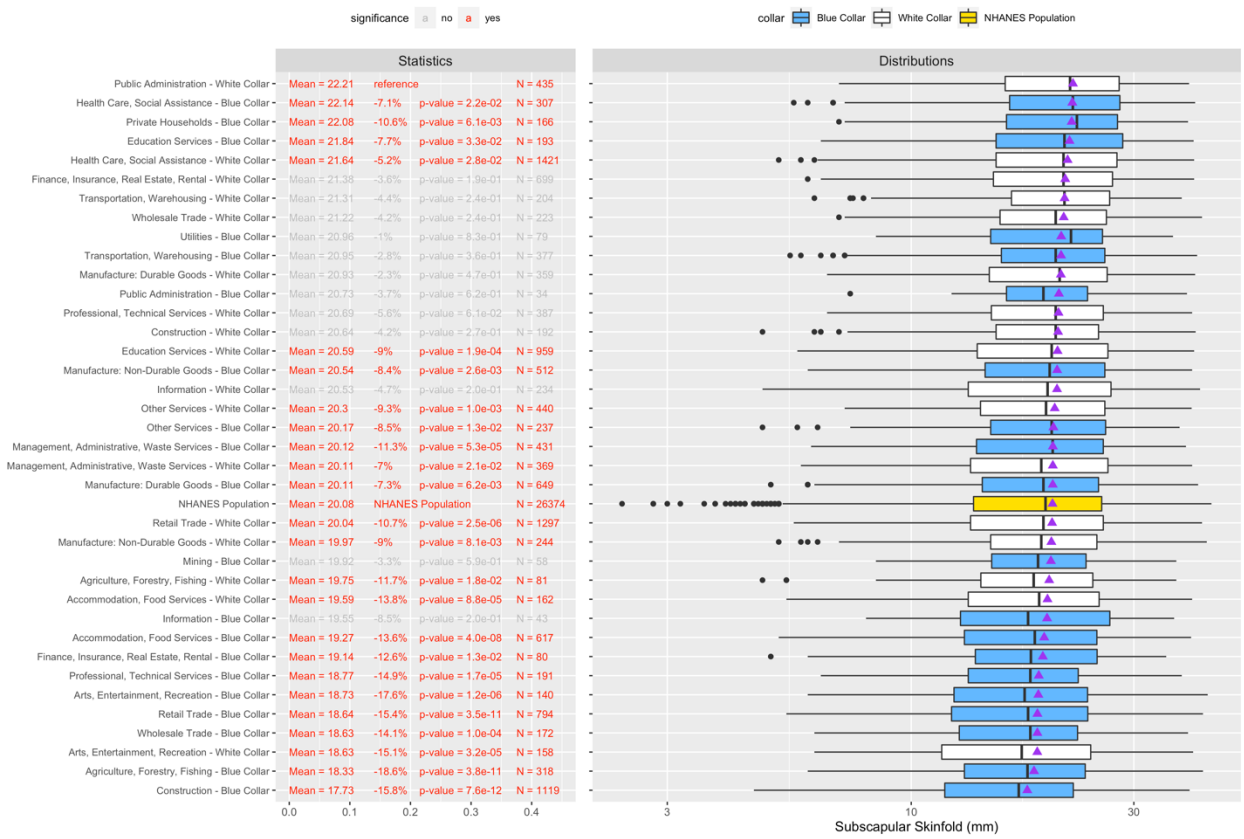

Figure 4.14. Box plot of distribution of subscapular skinfold. Far-left statistics are the mean concentrations. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for alkaline phosphatase, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. The far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of subscapular skinfold for a given sector-collar combination. Results are adjusted for age, sex, and race.

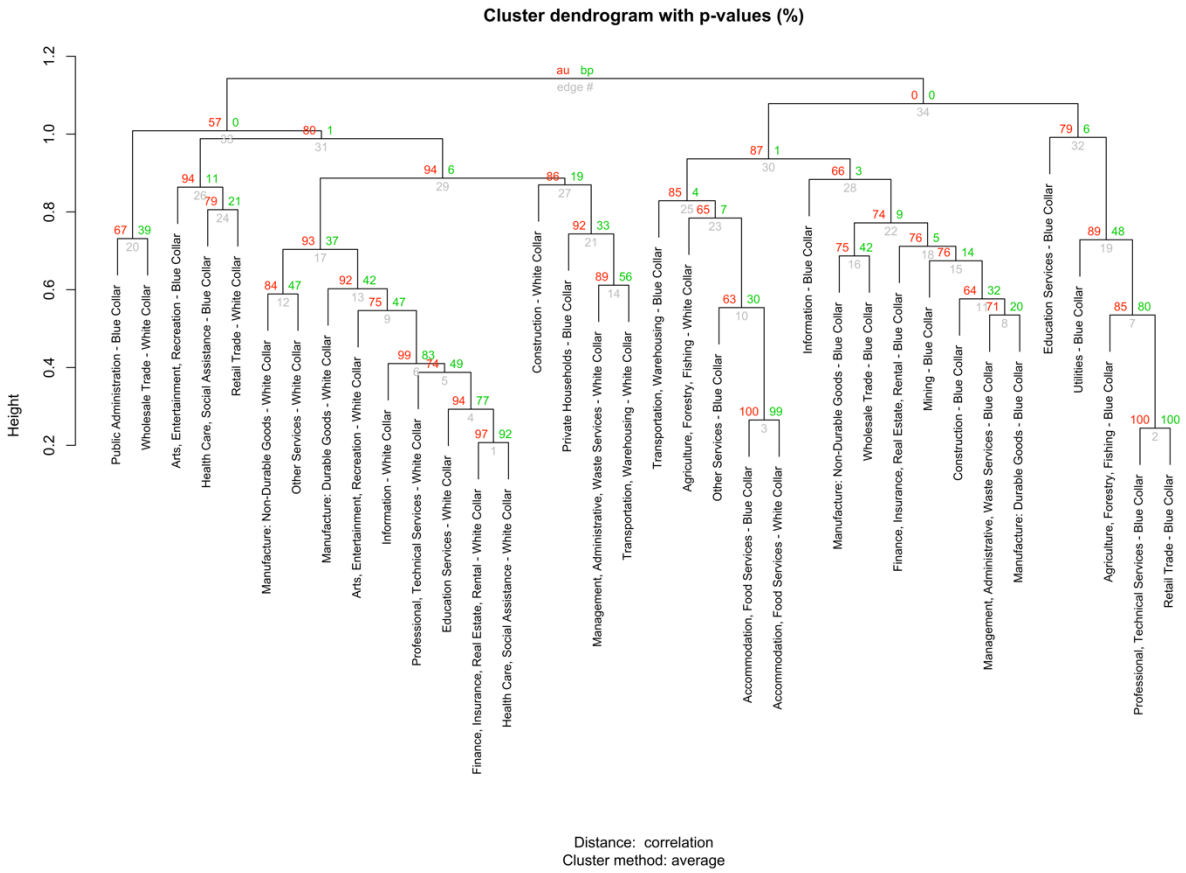

Figure 4.15. Dendrogram of sector-collar combinations based on exposure profiles for 108 chemical biomarkers. Pearson's correlation-based distance is the dissimilarity metric. Average linkage method was used due to having the highest cophenetic correlation coefficient, i.e. using this linkage method generated the dendrogram that best preserves the dissimilarity between the sector-collar combinations.

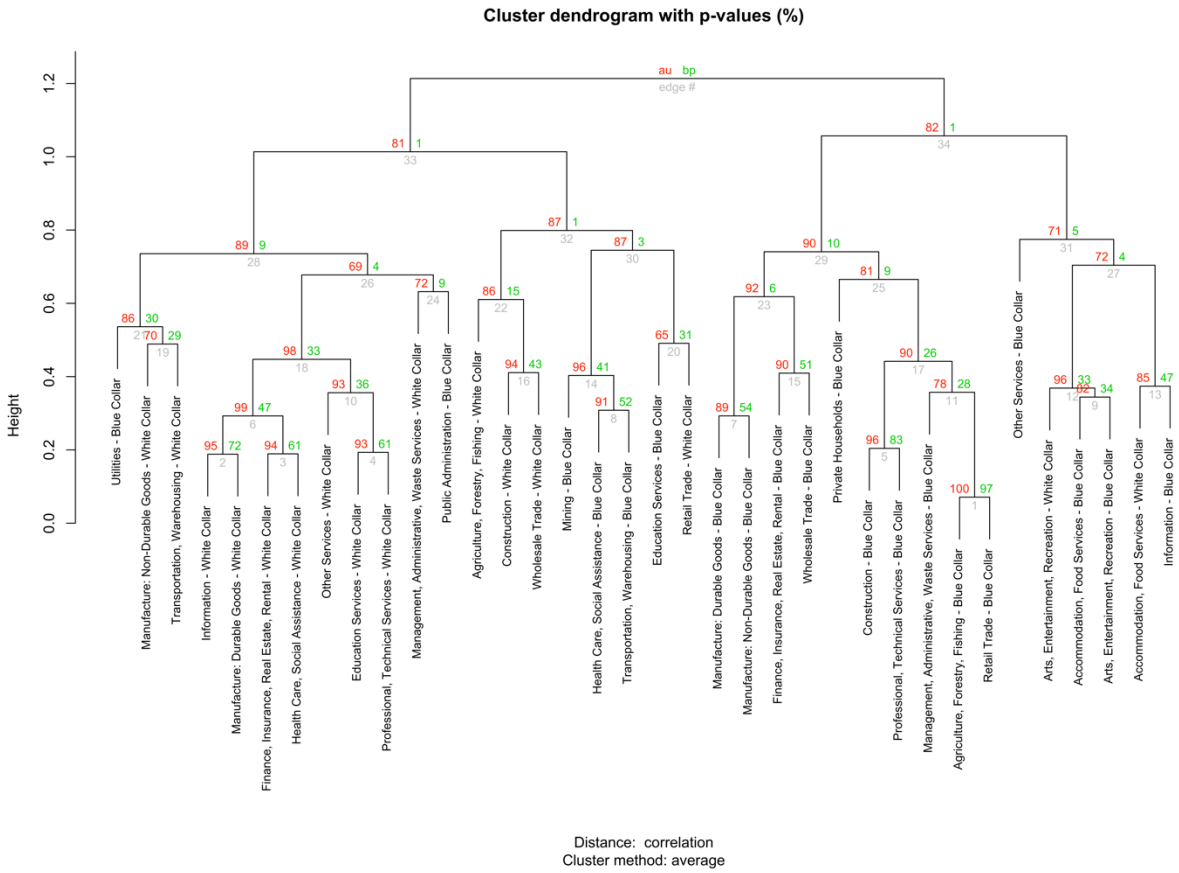

Figure 4.16. Dendrogram of sector-collar combinations based on physiological response profiles for 27 physiological indicators. Pearson's correlation-based distance is the dissimilarity metric. Average linkage method was used due to having the highest cophenetic coefficient, i.e. using this linkage method generated the dendrogram that best preserves the dissimilarity between the sector-collar combinations.

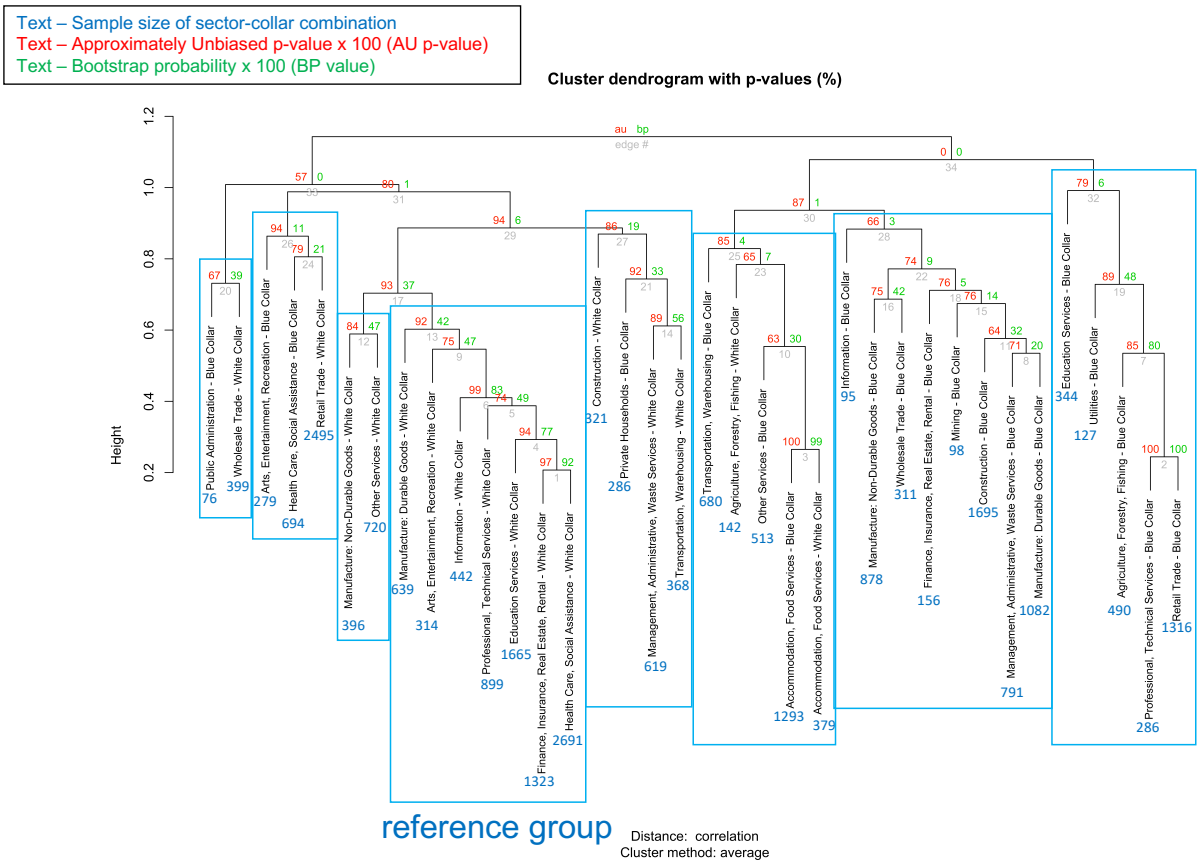

Figure 4.17. Dendrogram of sector-collar combinations based on exposure profiles for 108 chemical indicators. Red text represents the sample size of the sector-collar combination. Red boxes indicate the clustered occupational groups defined based on of sample size ≥ 400 or AU p-value ≥ 0.8. Pearson's correlation-based distance is the dissimilarity metric. Average linkage method was used due to having the highest cophenetic correlation coefficient, i.e. using this linkage method generated the dendrogram that best preserves the dissimilarity between the sector-collar combinations.

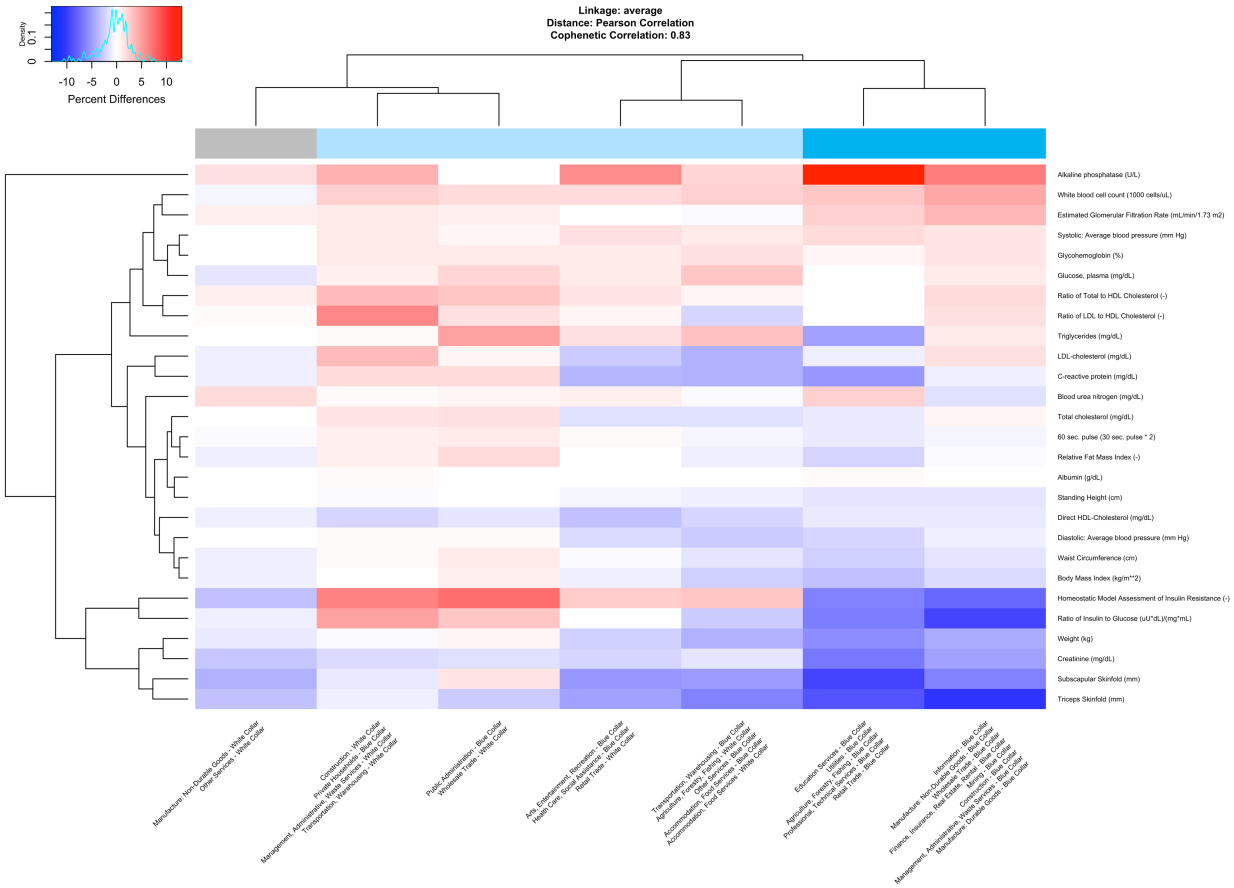

Figure 4.18. Heatmap of percent differences in physiological indicator measurements by clustered groups, relative to White Collars from Manufacture: Durable Goods; Arts, Entertainment, Recreation; Information; Professional, Technical Services; Education Services; Finance, Insurance, Real Estate, Rental; Health Care, Social Assistance; and Public Administration. Physiological indicators in white color indicate that the measurements are the same between the given sector-collar combination and the reference group. The color bar represents the collar gradient of the collar categorization. Dark blue indicates that the occupational cluster is comprised of all blue-collar workers. Lighter blue implies that the occupational cluster is comprised of a mixture of blue and white-collar workers. Gray implies that the occupational cluster includes only white-collar workers. Results are adjusted for age, sex, and race.

4.8 Tables

Table 4.1. Population statistics of 26,186 NHANES participants with occupational data.

CATEGORICAL			
Cycle	N (%)	Sex	N (%)
1999-2000 (Cycle 1)	3020 (11.5)	Male	14030 (53.6)
2001-2002 (Cycle 2)	3469 (13.2)	Female	12156 (46.4)
2003-2004 (Cycle 3)	3036 (11.6)	Race/Ethnicity	
2005-2006 (Cycle 4)	3340 (12.8)	Mexican Americans	5347 (20.4)
2007-2008 (Cycle 5)	3435 (13.1)	Other Hispanics	2001 (7.6)
2009-2010 (Cycle 6)	3472 (13.3)	Non-Hispanic Whites	11370 (43.4)
2011-2012 (Cycle 7)	3024 (11.5)	Non-Hispanic Blacks	5437 (20.8)
2013-2014 (Cycle 8)	3390 (12.9)	Other Races	2031 (7.8)

CONTINUOUS									
	N (%)	Minimum	5th	10th	Median	Mean (SD)	90th	99th	Maximum
Age (years)	26186 (100)	16	18	19	39	39.6 (15.3)	61	75	85

Chapter 5 Characterization of Linear and Non-linear Associations between Physiological Indicators and All-Cause Mortality

5.1 Abstract

Importance: Clinicians have used thresholds to prioritize high-risk patients usually without considering the linear or non-linear associations between a physiological indicator and a health endpoint.

Objective: To determine whether the clinical thresholds align with values indicative of increased mortality risk by characterizing the linear and non-linear relationships with all-cause mortality for 27 physiological indicators.

Design: We used the cross-sectional, biomarker data from the 1999-2014 National Health and Nutrition Examination Survey linked with National Death Index mortality data collected through December 31, 2015.

Setting: Data were analyzed from January 2019 to April 2020.

Participants: A nationally representative sample of 45,032 adults aged 18 years and older.

Exposures: 9 cardiovascular, 7 body composition, 5 metabolic, 4 nephrological, and 2 immune biomarkers

Main Outcomes and Measures: All-cause mortality. We used Cox proportional hazards regression models adjusted for age, sex, and race to associate physiological indicators with all-cause mortality. We used 10-fold cross validation to help select the most appropriate model using the Concordance Index, Nagelkerke R^2, and Akaike Information Criterion. We compare the

clinical thresholds to association-based thresholds, defined as values where the hazard ratio is 10% higher than the minimum hazard ratio.

Results: 24 out of 27 (88.9%) indicators show non-linear associations, while height, triglycerides, and 60-second pulse show linear associations. 13 out of 23 indicators (57%) showed some agreement between the association-based and clinical thresholds. Creatinine shows concordance for both bounds of the male thresholds. 12 indicators were concordant for one threshold with confirmation of the threshold for C-Reactive Protein (CRP). 14 indicators show no concordance. For 10 indicators which includes the cholesterol-related indicators, glucose control indicators, markers of body fat deposition, and Glomerular Filtration Rate, the parabolic associations suggest the need for two thresholds.

Conclusions and Relevance: The systematic characterization of the relationships with all-mortality confirmed that most of the studied physiological indicators show non-linear associations. It also confirms the relevance of the clinical threshold for half of the studied indicators; however, there is a need to reconsider the thresholds for indicators that show parabolic associations but only have one bound.

5.2 Introduction

Clinicians commonly use reference ranges[342–344] or clinical thresholds to determine whether a patient's test results are normal. A reference range is defined as a set of values between the 2.5th and 97.5th percentiles for a physiological indicator in a healthy population[345], which is the case for albumin, blood urea nitrogen (BUN), creatinine, alkaline phosphatase, and C-reactive proteins (CRP). Using percentiles without consideration of a health outcome can be presumptuous about disease pathophysiology[346,347]. Some clinical thresholds, however, were established via association with an outcome[348–350]. For example, the classification of body mass index (BMI) was

calibrated to all-cause mortality in 1995[348]. While the associations between BMI and mortality has changed, the classifications still remained unchanged. For an indicator of kidney function, the lower threshold for Glomerular Filtration Rate (GFR) was established based on associations with all-cause mortality and cardiovascular disease[350]. But the restricted distribution of GFR in the studied population prevented an upper bound from being defined. Thus, there is a need to evaluate whether the current reference ranges and clinical thresholds are relevant in distinguishing high-risk from low-risk patients.

Improved risk stratification requires understanding whether associations between physiological indicators and health outcomes are described by linearity or non-linearity[351–353]. Most association studies consider linearity[107–109] or non-linearity[111,112] separately without quantitatively assessing which of the models better describes the relationship between a given physiological indicator and a health outcome. Many studies do compare the prediction performance of various models such as linear, quadratic, cubic, and logarithmic[116,117] but have not validated the results on new data. This led to finding that the more flexible model is deemed better at characterizing the associations, which may reflect the actual relationship or show an overfitted model[118]. Since these gaps may lead to significant limitations in risk surveillance and management, there is a need to develop a statistical framework to compare models and identify the most appropriate linear or non-linear model while evaluating for overfitting.

To address these limitations, the goal of this study is to characterize the relationships between all-cause mortality and 27 physiological indicators in the US population. We focus on all-cause mortality as it is the ultimate health outcome. For each physiological indicator, we specifically 1) compare the prediction performance of linear and different nonlinear models by applying a machine learning approach, 2) assess the robustness of the models by observing changes

in prediction performance when extreme measurements are excluded, 3) describe the associations between the physiological indicator and mortality as characterized by the most appropriate model(s), and 4) determine the relevance of the current clinical thresholds by evaluating whether these values are indicative of increased mortality risk.

5.3 Methods

5.3.1 Study Population

The National Health and Nutrition Examination Survey (NHANES) is a cross-sectional study conducted by the Centers for Disease Control and Prevention (CDC) to characterize the health of a non-institutionalized, civilian US population[19]. For this analysis, we used the continuous 1999-2014 NHANES data and started with sample of 82,091 participants. We also used data linked with the National Death Index to ascertain mortality information collected through December 31, 2015[354]. We excluded participants for not having data on mortality status (N = 37,046) and those who were not followed up (N = 13), leading to a sample of 45,032 participants. **Figure A4.1** describes these exclusion criteria.

5.3.2 Measurements of Physiological Indicators

Figure A4.1 also describes the curation process on which physiological indicators to include in our analysis. We identified 60 biomarkers and anthropomorphic measures that characterize physiologic function. To exclude physiological indicators with low overlap with the mortality data, we excluded those with measurements in fewer than six NHANES cycles (n = 10) and with a sample size of less than 10,000 participants (n = 21). As we preferred continuous variables for studying linear and non-linear associations, we also excluded physiological indicators that are categorical (n = 2). The final dataset for analysis consisted of 27 physiological indicators.

Tables A4.1-A4.3 compare the observed characteristics between participants with complete and incomplete data.

5.3.3 Database of Clinical Thresholds

To characterize the relevance of the current clinical thresholds, we compiled a database of thresholds for 23 physiological indicators from literature. Thresholds were not available for height, subscapular skinfold, triceps skinfold, and weight. Since the same indicator may have several thresholds, we used the range of the thresholds. We used sex-specific thresholds when available.

5.3.4 Statistical Analyses

We performed all analyses using R version 3.6.0. To produce estimates that are representative of the non-institutionalized, civilian US population, we accounted for NHANES sampling designs by applying the survey weights to our statistical models[18].

We used Cox proportional-hazards regression models to characterize the associations between all-cause mortality and each physiological indicator[355]. For each physiological indicator, we assessed linear associations with all-cause mortality and tested two non-linear associations by discretizing each indicator into 9 quantiles ("novemtiles") and by using a weighted sum of cubic polynomials ("spline")[284,356] with the latter considered the more flexible, non-linear model. The reference group for the novemtiles models is the novemtile with the minimum mean hazard ratio. We conducted 10-fold cross validation[357] while adjusting for linear age (continuous), sex (categorical), and race/ethnicity (categorical) to compare the predictive capability of the linear versus non-linear models. We adjusted for linear age as it showed better prediction performances compared to non-linear age (**Figure A4.2**). We observed whether the association between mortality and BMI remains robust after adjusting for smoking with blood cotinine. We selected the model that best describes the association between mortality and each physiological indicator

by using the Akaike Information Criterion (AIC)[358], Concordance Index[359], and Nagelkerke R2[360]. We provide our justification for using these fits measures in **Table A4.4**. We defined confidence intervals around each measure by bootstrapping for 1000 replicates[361]. To account for multiple comparisons across the models, we used a False Detection Rate (FDR) method of 5% on the p-values of the regression coefficients pertaining to the physiological indicators[103].

To assess the influence of extreme measurements on the prediction performance, we conducted sensitivity analyses on the distributions of each physiological indicator. We applied a series of Cox Proportional Hazard models on sample subsets restricted to the study participants in the 1st to 99th, 5th to 95th, and 10th to 90th percentiles of each indicator.

To define clinical thresholds based on the associations with mortality, we used the spline model to identify values of the physiological indicator that show an increased mortality risk of 10% from the minimum risk. To define sex-specific clinical thresholds, we applied the same procedure to each sex.

5.3. Results

Table 5.1 presents population characteristics for the 45,032 participants. **Figure A4.3** displays the distributions of each physiological indicator.

Figure 5.1 displays the Concordance Index and Nagelkerke R2s across all the models and physiological indicators for two populations: one with all participants and another with participants who have measurements within the 1st and 99th percentiles. **Figure A4.4-A4.5** display the AICs and Concordance Index for the same populations, respectively. Across the fit measures, the non-linear models show better prediction performances compared to the linear models. The prediction performance of the novemtiles models were consistently high across all physiological indicators regardless of the extent of measurements exclusion and were also stable as reflected by

the narrower confidence intervals. In contrast, linearity shows low R^2s for inflammatory (CRP, alkaline phosphatase, white blood cell counts), metabolic (HOMA-IR), and nephrological (BUN, Creatinine) biomarkers when all participants were included. The spline model R^2s are also low for alkaline phosphatase and cardiovascular biomarkers involving LDL cholesterol. For these physiological indicators, exclusion of measurements outside the 1st and 99th percentiles resulted in improved prediction performance, particularly for the linear and spline models. The overfitting is due to the linear and spline models attempting to fit to the outliers. **Figure A4.6** shows the correlations between decreased sample size and improved prediction performance.

Figure A4.7-A4.9 compare the prediction performance on including all participants with those from the sensitivity analysis on restricting the distribution of the physiological indicators to the 5^{th} to 95^{th} and the 10^{th} to 90^{th} percentiles. **Figure A4.10** displays the statistical significance of all models with respect to the prediction performances. While prediction performance and statistical significance improved when the distribution was restricted to the 1^{st} and 99^{th} percentiles, further exclusion did not lead to further improvement for the linear nor spline models. In contrast, the novemtiles models show consistent prediction performance and significance regardless of the restrictions. This implies that studying non-linearity with quantile-based models will result in stable predictions, as outliers do not heavily influence these models given enough participants in each quantile. In our case, we ensured at least 1100 participants in each novemtile.

Figure 5.2 displays the relative risk for death through the hazard ratios across the distribution of BMI, Systolic Average Blood Pressure (SBP), Ratio of Total to HDL Cholesterol, CRP, HOMA-IR, and GFR for the different models. We select these examples to highlight expected and unexpected findings, challenges in model interpretation, and relevance of the clinical thresholds. To aid visualization, we show the associations when measurements outside the 1^{st} and

99th percentiles were excluded. **Figure A4.11** displays the associations for these physiological indicators in the entire population, while **Figure A4.12-A4.16** show the relative mortality risk for the remaining physiological indicators. **Figure A4.17-A4.22** present the sex-specific associations. These figures are available online [name of shiny app]. For BMI, while the non-linear models show a parabolic association with death in **Figure 5.2A**, the linear model suggests that mortality risk is the same across the entire distribution. The association-based thresholds for BMI are shifted higher compared to the current clinical thresholds. These findings were robust even after adjusting for smoking (**Figure A4.23**). Low and high BMI were strongly associated with all-cause mortality in males (**Figure A4.18A**) than in females (**Figure A4.18B**). For average SBP, the linear model identified increased risk with elevated blood pressure (**Figure 5.2B**). However, the non-linear models suggest a parabolic association between mortality and SBP. In **Figure 5.2C**, the non-linear models imply that lower and higher ratios of Total to HDL Cholesterol are associated with higher mortality risk. In contrast, the linear model shows a positive association between the ratios and mortality. Increased mortality risk aligns with the threshold for males (**Figure A4.17I**) but is higher than the threshold for females (**Figure A4.17J**). In **Figure 5.2D**, discretization of CRP levels into novemtiles shows that a sigmoidal function best characterizes this association compared to the linear model. Mortality risk becomes apparent for CRP at the current clinical threshold of 0.1 mg/dL. In **Figure 5.2E**, all models agree that elevated HOMA-IR is associated with increased mortality risk but disagree with the interpretation of the mortality risks at lower HOMA-IR with the spline, novemtiles, and linear models suggesting increased risk, no effect, and lower risk, respectively. In **Figure 5.2F** for GFR, the linear model suggests that participants below the clinical threshold of 90 mL/min/1.73m^2 [362,363] are at higher risk to death. But the non-linear models suggest that participants with GFR below 65 or above 90 mL/min/1.73m^2 are at risk.

Table 5.2 compares the current clinical thresholds with the association-based thresholds. Creatinine shows concordance for both bounds of the male thresholds. Twelve indicators were concordant for one threshold with Relative Fat Mass Index, Creatinine, HDL-Cholesterol, and Ratio of Total to HDL Cholesterol agreeing with one of the sex-specific thresholds. Fourteen indicators show no concordance.

5.4. Discussions

We present a data-driven approach to identify the models that most appropriately describe the association between physiological indicators and all-cause mortality in the US population. We applied a machine-learning based approach to test the prediction capability and robustness of linear and non-linear models while penalizing models that are more prone to overfitting. We observed that with exception of height, triglycerides, and 60-second pulse, the associations between the other indicators and mortality are non-linear. We used the "winning" model(s) to determine expected and unexpected directions associated with higher mortality risk for each physiological indicator. Results from the spline models can also be informative of whether the current clinical thresholds align with values identified at which mortality risk increases over baseline mortality risk.

There are advantages and disadvantages of using the spline models as an alternative approach to model the non-linear association with the novemtiles model. The spline models provide relative mortality risk for each measurement along the distribution of a given physiological indicator, while the novemtiles models provide an average risk for measurements within the same novemtile. Thus, using the spline model enable us to observe a gradient of mortality risk that is not possible in the novemtiles models. Due this attribute, spline models have been commonly used to define the clinical thresholds for physiological indicators such as GFR[350] and BMI[348]. A

disadvantage of the spline model was the overfitting to outliers, which resulted in lower and unstable prediction performance. The interpretations were also unstable as the confidence intervals for mortality risk was wide for participants with extreme measurements. However, an unexpected advantage of the spline model overfitting is the ability to find individuals with extraordinary risk at the tails of a biomarker distribution. This highlights the need for focused study of individuals with these extreme biomarker levels to build better predictive models. As another disadvantage, when all measurements were included, fitting a spline model created difficulties in interpretation. The relative risks for some extreme measurements were substantially higher, therefore making the relative risks associated with other measurements appear negligible. However, when measurements outside the 1st and 99th percentiles of a given physiological indicator were excluded, fitting a spline model on measurements improved the interpretation along with the prediction performance and its stability, making it at times on par with that of the novemtiles models. But excluding participants, especially those with extreme measurements, is not conducive to helping those who would be at most need for targeted interventions. Overall, while the spline models were prone to overfitting due to outliers, these models provide insights on the gradient of the mortality risk as well as showcase the need to sample for participants with extreme measurements in order to enhance risk stratification.

We found non-linear associations for cholesterol-related biomarkers. The non-linear models support the growing literature on the associations between higher HDL levels and increased mortality risk[364,365], possibly contradicting the impression of being the "good" cholesterol. The current notion of LDL being the "bad" cholesterol implies that higher LDL levels are associated with increased risk[366]. Hence, guidelines on cholesterol management recommend lipid-lowering therapy for patients who show high cardiovascular risk and have LDL levels of 70

mg/dL or higher[367]. Following the guidelines might lead to more patients being treated to abnormally low levels of LDL. Our non-linear results show a stronger association between mortality and lower LDL levels compared to higher levels, which are consistent with other studies[368–372]. It is unclear whether these associations are due to preexisting disease leading to low cholesterol levels[370,372], adverse side effects from lipid-lowering medications[369,371], or low LDL levels being a causal factor of mortality[370,373]. Thus, we advocate using longitudinal data to understand the risk associated with changing versus stable cholesterol levels to gain insights on the effects at low cholesterol levels.

Obesity in the US is a major health problem as 71% of Americans are overweight or obese[374]. Eating disorders and being underweight are also important health problems, since 10-15% of Americans are affected[375,376]. Participants with lower BMI are associated with an increased mortality risk compared to those with higher BMI, which may be attributed to efficacy of public health interventions or improvements in healthcare for obesity-related conditions[377]. In addition, the associations with mortality have shifted with the minimum risk found within 24-30 kg/m^2 instead of at 24 kg/m^2 [348]. Our results are consistent with other studies[378–380], suggesting additional investigation on other drivers leading to decreased risk associated with higher BMI. The associations with all-cause mortality in lower and higher BMI were attenuated in women than in men[381], which may suggest residual confounding due to sex differences[382] and/or the survival advantage of adipose distribution in females[383]. These findings emphasize the importance to understand the influence of caloric intake and adipose tissue mass on mortality risk.

Our results support low GFR (hypofiltration[384]) and surprisingly elevated GFR (hyperfiltration[385]) as potentially reflecting renal injury. Interestingly, mortality risk for participants with hyperfiltration is twice as high as those with hypofiltration in the novemtiles

model. This result is especially pertinent to Native Americans[386] and Non-Hispanic Black Americans[387], who on average have elevated GFR, prompting the need to establish an upper threshold for GFR. Overall, these findings prompt the need to also study the associations of these biomarkers with other health endpoints to help reevaluate these pathophysiological mechanisms for better informed medical decision making.

Precision prevention and precision medicine involve leveraging biologic, demographic, and epidemiologic data to develop personalized risk prediction models[388,389]. These models can be used to define a risk score to quantify a patient's likelihood to a disease event. Some clinical risk scores are based on assuming a linear relationship between a given biomarker and a health outcome[390–392]. We observed, however, that when including all participants, linearity has the worst prediction performance when describing the associations between inflammatory, metabolic, and nephrological biomarkers and all-cause mortality. Other risk scores dichotomized the physiological indicators based on a cutoff[393–396], such as a clinical threshold or a percentile. For physiological indicators that have parabolic associations with risk, using only one threshold to dichotomize the indicator will result in misclassifying the patients. This scoring methodology also assumes that risk is the same regardless of any deviation from the cutoff. However, our findings prompt the need to reexamine this assumption and suggest that estimating risk in this manner does not adequately represents patient's risk, especially when the measurements are extreme.

While some risk scores account for non-linear associations, some only penalize if measurements are in one unfavorable direction instead of in either direction[397–399], which is particularly problematic for indicators showing a parabolic association with an outcome. For example, a score estimating cardiovascular risk penalizes for high Total Cholesterol and high SBP[397] but not for low SBP. This suggests the need to better incorporate the non-linear associations

in the penalization. Our results may not be generalizable to the hospitalized population. Thus, future studies can use our framework to incorporate these associations in defining a more representative risk score to help clinicians provide tailored lifestyle and medical recommendations to high-risk patients.

This study has several limitations. First, the cross-sectional nature of NHANES does not consider changes in physiological measurements over the follow-up period. Thus, more temporal measurements may better dissect the unexpected associations with mortality. Second, we did not evaluate how other demographic or lifestyle factors may influence the relationship between a physiological indicator and mortality. Future analyses can build upon our baseline model to quantify the type of influence e.g. mediating, moderating, or confounding that such factors have on these associations, but the effect of bias becomes increasing difficult to untangle with more complex models. Third, we observed that physiological indicators with smaller sample size showed better prediction performance, which was especially true for the AIC as it showed a strong positive correlation with sample size. Thus, we caution against using our results to infer the relative importance of the indicators on mortality. Hence, future studies can apply feature selection techniques on a complete dataset comprised of participants with measurements available for all 27 physiological indicators to infer relative importance. Fourth, the spline model may not be the most optimized model, since we manually defined the interior knots to be at the 25^{th}, 50^{th}, and 75^{th} percentile of the distribution of a given physiological indicator. Hence, future studies can compare the cross-validated prediction performance of spline models defined with different number and locations of interior knots to identify the optimum spline model. In addition, future work can use Gaussian kernel regression models, which involves taking a weighted average of the surrounding data points to predict a given data point [337]. Such models may be a more automated alternative to

the spline models as the Gaussian kernel regression models do not require knots to be defined. Fifth, the discretization for the novemtiles model may not have led to the optimized quantile model, thus future directions can involve identifying the optimal number of discretization. Finally, we may not confidently characterize the morality risk for participants with extreme values due to lower sample size. Thus, there is a need to sample more participants with extreme measurements.

5.5. Conclusions

Accurate identification of high-risk patients requires systematic methodologies to characterize the linear and non-linear associations between a physiological indicator and any health outcome. In this study, we developed a machine-learning based framework to establish the appropriate model to best characterize the relationship with all-cause mortality for a wide variety of physiological indicators. Our framework led to 1) determining that non-linearity is better at charactering the relationships with death for most of the studied indicators, 2) identifying unexpected directionality in the relationships between GFR and cholesterol-related biomarkers and increased risk of death, and 3) observing the validity of the clinical thresholds for a few biomarkers and also the need for reevaluation for others. Our findings can guide future efforts in risk prioritization, so that preventative strategies can be personalized to promote better informed decision making to reduce the incidence of any disease event.

5.6 Figures

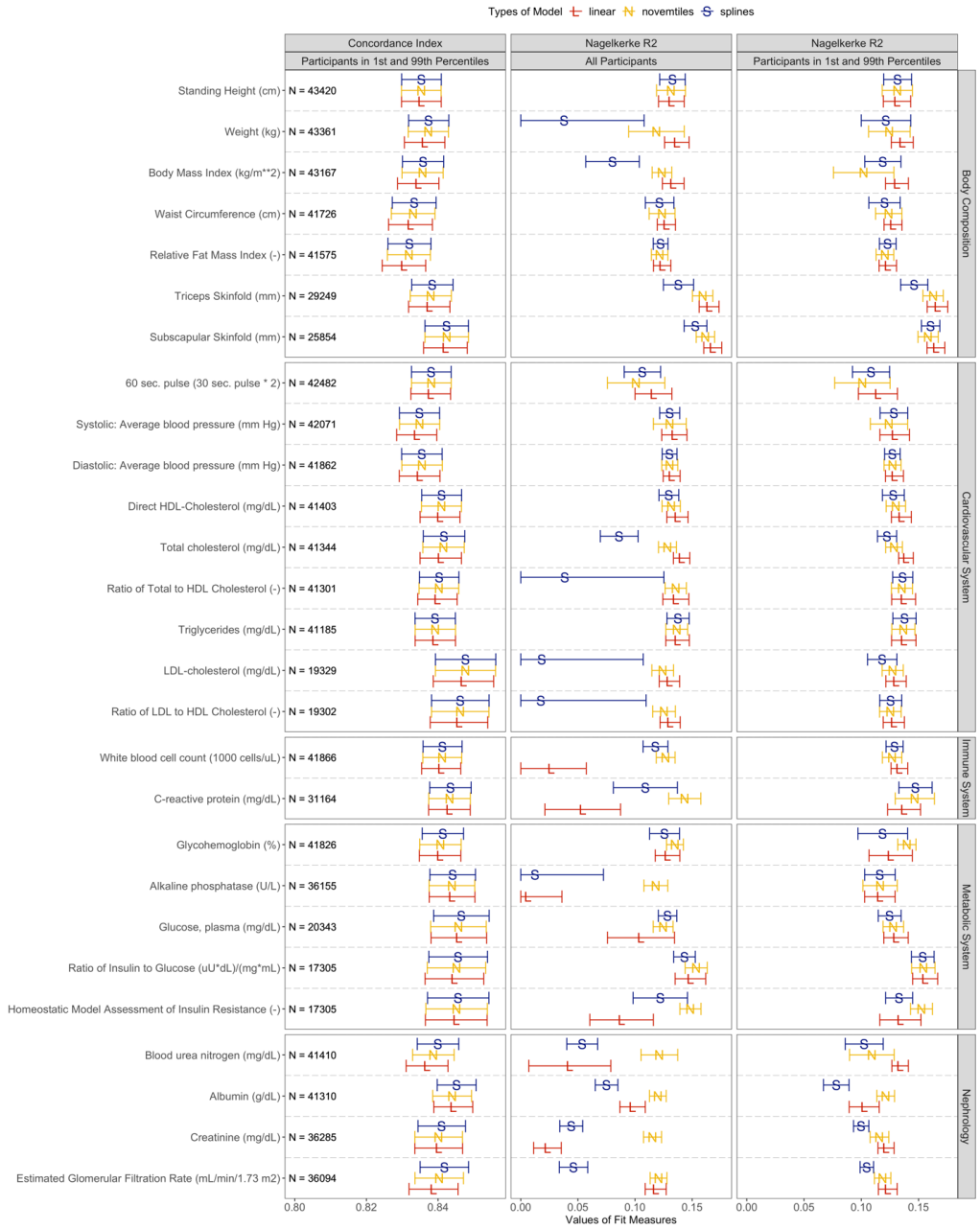

Figure 5.1. Alphabet soup plot displaying the AIC and Nagelkerke R2 for the associations with all-cause mortality for all physiological indicators, grouped by body system. The prediction performances are displayed for two populations: one with all participants and another with participants who have measurements within 1st and 99th percentiles for a given physiological indicator. Sample size for each physiological indicator is provided to indicate the number of participants who have data for mortality, age, sex, race, and the given indicator. Results were adjusted for age, sex, and race/ethnicity.

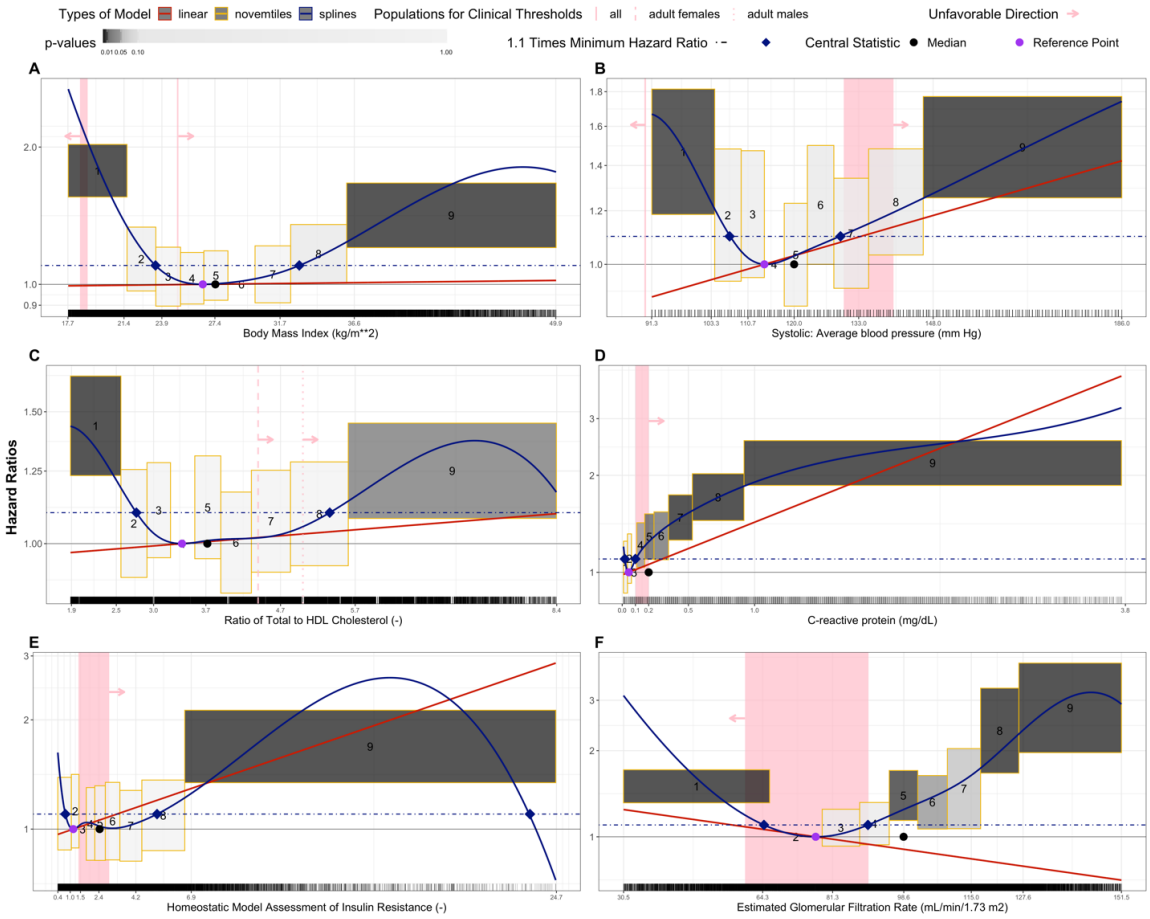

Figure 5.2. Stairway plots of hazard ratios relative to physiological indicators across all models to describe the relative mortality risk for A) Body Mass Index, B) Average Systolic Blood Pressure, C) Ratio of Total to HDL Cholesterol, D) C-Reactive Proteins, E) Homeostatic Model Assessment of Insulin Resistance, and F) Glomerular Filtration Rate. For visualization aid, participants with measurements between the 1st and 99th percentiles of a physiological indicator are included. Relative risks for mortality from the novemtiles model are represented by the boxes with the width representing the range of a novemtile and the height representing the 95% Confidence Interval of the hazard ratio. The mean hazard ratio for each novemtile is presented by a digit. The hazard compares participants in a novemtile to those in the reference group at the novemtile shown without a box. The purple dot represents the reference point and the measurement of a physiological indicator shown to have the lowest hazard ratio for the linear and spline models. The red and blue lines represent the relative mortality risk with respect to reference point for the linear and spline models, respectively. The black dot represents the median of a physiological indicator. The dashed navy line represents when the hazard ratio is 10% higher than the minimum hazard ratio, i.e. when the hazard ratio is 1.1. The navy diamonds indicate the concentration at which the hazard ratio shows a 10% increase from the minimum hazard ratio. The pink lines and rectangles represent the values of the clinical thresholds with the width of the rectangles representing the ranges of the threshold. The set of tick marks along the base of the plot represent the distribution of a physiological indicator with increased opacity implying increased number of participants. Results were adjusted for age, sex, and race/ethnicity.

5.7 Tables

Table 5.1. Characteristics of 45,032 participants and distributions of the 27 physiological indicators in the NHANES population.

Categorical Variables											
Mortality Status	N (%)		Sex	N (%)		Race	N (%)				
Deceased	5588 (12.4)		Males	21737 (48.3)		Mexican	8562 (19.0)				
Alive	39444 (87.6)		Females	23295 (51.7)		Other Hispanics	3286 (7.3)				
						Non-Hispanic Whites	20312 (45.1)				
						Non-Hispanic Blacks	9665 (21.5)				
						Other Race/Multi-Racial	3207 (7.2)				
Continuous Variables											
	N	Min	1st	5th	10th	Median	Mean	90th	95th	99th	Max
Time of follow up (months)	45032	1	12	19	27	92	97.30	177	189	199	201
Age (years)	45032	18	18	19	21	46	47.09	75	80	85	85
Body Composition											
Body Mass Index (kg/m**2)	44047	12.04	17.72	19.81	21.2	27.43	28.52	37.1	40.937	49.92	130.21
Standing Height (cm)	44288	123.3	145.7	151.3	154.3	166.9	167.32	180.8	184.4	190.3	204.5
Subscapular Skinfold (mm)	26374	2.4	6.5	8.4	9.9	19.4	20.08	31.3	34.5	38.5	44
Triceps Skinfold (mm)	29824	2.4	4.9	7	8.5	17.8	18.92	31.2	34.2	38.4	45
Waist Circumference (cm)	42558	55.5	67.5	73.4	77.4	96.2	97.34	118.2	126	142.2	179
Weight (kg)	44233	25.6	45.1	52.2	56.7	77.1	80.04	106.8	118	145.4	371
Relative Fat Mass Index (-)	42425	2.82	15.26	20.17	23.49	34.40	34.91	46.63	48.67	51.89	58.41
Cardiovascular System											
60 sec. pulse (30 sec. pulse * 2):	43147	32	48	54	58	72	72.84	90	96	106	224

Diastolic: Average blood pressure (mm Hg)	42647	3.33	39.33	50	54.67	70	69.88	84.67	89.33	100	132
Systolic: Average blood pressure (mm Hg)	42881	64.67	91.33	98.67	102.67	120	123.74	149.33	161.33	186	270
Direct HDL-Cholesterol (mg/dL)	42145	7	26	32	35	50	52.77	74	82	101	188
LDL-cholesterol (mg/dL)	19696	9	45	62	72	112	114.81	161	177	213	629
Ratio of LDL to HDL Cholesterol (-)	19696	0.15	0.70	1.01	1.21	2.15	2.31	3.6	4.08	5.15	33.11
Ratio of Total to HDL Cholesterol (-)	42144	1.31	1.89	2.26	2.49	3.72	3.99	5.82	6.60	8.40	37.05
Total cholesterol (mg/dL)	42147	59	112	133	144	192	195.39	250	270	313	813
Triglycerides (mg/dL)	42018	9	34	46	56	116	148.28	269	347	589	6057
Immune System											
C-reactive protein (mg/dL)	31478	0.01	0.01	0.02	0.04	0.21	0.45	1.06	1.63	3.77	29.6
White blood cell count (1000 cells/uL)	42703	1.5	3.5	4.31	4.8	6.9	7.27	10	11.2	13.798	99.99
Metabolic System											
Alkaline phosphatase (U/L)	36827	7	32	41	46	68	71.86	101	115	153	1378
Glucose, plasma (mg/dL)	20757	36	72.556	81	84.7	98	105.92	128	159	281	686.2
Glycohemoglobin:(%)	42657	2	4.5	4.8	4.9	5.4	5.64	6.4	7.4	10.7	18.8
Homeostatic Model Assessment of Insulin Resistance (-)	17659	0.05	0.40	0.74	0.98	2.43	3.81	7.24	10.31	24.72	204.47
Ratio of Insulin to Glucose	17659	0.00	0.02	0.03	0.04	0.10	0.13	0.24	0.31	0.56	5.06
Nephrology											
Albumin (g/dL)	42041	1.2	3.1	3.6	3.8	4.3	4.25	4.7	4.8	5	5.7
Blood urea nitrogen (mg/dL)	42038	1	4	6	7	12	13.19	20	23	35	122

Creatinine (mg/dL)	36831	0.16	0.4	0.5	0.6	0.82	0.89	1.17	1.3	1.957	17.8
Estimated Glomerular Filtration Rate (mL/min/1.73 m^2)	36831	1.85	30.46	51.91	62.70	98.55	97.17	128.48	136.81	151.50	207.58

Table 5.2. Comparison of current clinical thresholds or reference ranges with association-based thresholds of the 27 physiological indicators.

		Clinical		Association-based[a]		Concordance[b]	
Physiological Indicators	population[c]	lower threshold	upper threshold	lower threshold	upper threshold	lower threshold	upper threshold
Body Composition							
Body Mass Index (kg/m**2)	all	18.5-19	24.9-25	23.48	32.97	No	No
Standing Height (cm)	all			168.6			
Subscapular Skinfold (mm)	all			29.9	37.1		
Triceps Skinfold (mm)	all			22.6			
Waist Circumference (cm)	females		80-88	73.6-86.8	105	No	No
Waist Circumference (cm)	males		90-102	87.6	115.5	No	No
Waist Circumference (cm)	all			86.6	113.3		
Weight (kg)	all			78.6	104.8		
Relative Fat Mass Index (-)	females	21-24	33-36	27.01	34.01-39.53	No	Yes
Relative Fat Mass Index (-)	males	8-13	19-25	25.14	31.36	No	No
Relative Fat Mass Index (-)	all			25.57	30.67		
Cardiovascular System							
60 sec. pulse (30 sec. pulse * 2)	all	50-60	80-100		65	No	No
Direct HDL-Cholesterol (mg/dL)	females	50		57	91	No	No
Direct HDL-Cholesterol (mg/dL)	males	40		42	68	Yes	No
Direct HDL-Cholesterol (mg/dL)	all			49	83		
LDL-cholesterol (mg/dL)	all		100-130	107	168	No	No
Triglycerides (mg/dL)	all		150	43	207	No	No
Total cholesterol (mg/dL)	all		200	186	284	No	No
Diastolic: Average blood pressure (mm Hg)	all	60	80-90	69.33	89.33	No	Yes
Systolic: Average blood pressure (mm Hg)	all	90	130-140	107	129.33	No	Yes

Ratio of LDL to HDL Cholesterol (-)	all		1.4	1.51	3.62-5.15	No	No
Ratio of Total to HDL Cholesterol (-)	females		4.4	2.42	5.02-8.38	No	No
Ratio of Total to HDL Cholesterol (-)	males		5	3.07	5.74	No	No
Ratio of Total to HDL Cholesterol (-)	all			2.77	5.36		
Immune System							
C-reactive protein (mg/dL)	all		0.1-0.2	0.02	0.1	No	Yes
White blood cell count (1000 cells/uL)	all	4-15	10-11	4.2	5.9	Yes	No
Metabolic System							
Glycohemoglobin (%)	all		5.7-7	4.9	5.9	No	Yes
Glucose, plasma (mg/dL)	all	60-70	99-126	95.4	107.9	No	Yes
Alkaline phosphatase (U/L)	all	20-44	116-147		64	No	No
Homeostatic Model Assessment of Insulin Resistance (-)	all		1.4-2.9	0.77	5.23	No	No
Ratio of Insulin to Glucose (uU*dL)/(mg*mL)	all		0.2-0.3	0.07	0.15	No	No
Nephrology							
Albumin (g/dL)	all	3.4-3.5	5.4-5.5	4.65	4.95	No	Yes
Blood urea nitrogen (mg/dL)	all	6-10	20-21	12	19	No	Yes
Creatinine (mg/dL)	females	0.5-0.6	1-1.1	0.6	0.81	Yes	No
Creatinine (mg/dL)	males	0.6-0.9	1.2-1.3	0.92	1.12	Yes	Yes
Creatinine (mg/dL)	all			0.63	0.93-1.15		
Estimated Glomerular Filtration Rate (mL/min/1.73 m2)	all	60-90		64.61	89.89	No	No

[a] The association-based thresholds are values of the physiological indicator that that show an increased mortality risk of 10% from the minimum risk of the spline model. All models are adjusted for age, sex, and race, while the sex-specific models are adjusted for age and race. Participants with measurements between the 1st and 99th percentiles of the physiological indicator were used.

[b] Concordance is achieved when the association-based threshold is within 10% of the clinical threshold. When there is a range for the clinical threshold, concordance is achieved when the association-based threshold is within the range or within 10% of the maximum clinical threshold.

[c] The thresholds are applicable to the entire US population ("all") or sex-specific ("females" or "males").

Chapter 6 Conclusion

My dissertation research interfaces data science, epidemiology, and environmental health as I applied computational approaches to understand how the environment can influence human health. This research required addressing two challenges: 1) characterizing chemical exposure ad disparities for a wide range of chemicals to identify populations that are at risk for high chemical exposures and 2) characterizing associations along the spectrum of chemical exposures to physiological indicators to adverse health outcomes. Addressing these challenges is necessary for understanding how certain chemical exposures perturb physiologic function and thereby result in increased risk of adverse health effects. Motivated by this goal, I developed and applied unbiased approaches to systematically screen for chemical exposure disparities by age, sex, race, and occupation for a wide range of toxicants (Chapter 2-4), and characterize the types of associations between all-cause mortality for 27 indicators used to describe physiologic function (Chapter 5).

Sections 6.1 to 6.4 present the key achievements, limitations, future directions, and contribution of the frameworks developed in the four studies (Chapter 2-5). Section 6.5 is an integrated discussion of the four studies. Finally, Section 6.6 presents my closing remarks about integrating these frameworks to ground exposome research in data science.

6.1 Characterization of Age-Based Trends of Chemical Biomarker Levels

This study focused on evaluating age-based exposure trends by applying a screening approach to a diverse suite of 141 chemical biomarkers in a representative sample of 74,942 participants in the US. Specifically, I defined characteristics to distinguish chemicals indicative of

legacy exposures from those reflecting current, ongoing exposures. I conducted a targeted analysis of PFASs to further substantiate the criteria for legacy exposures. In addition, I developed a framework using a series of quadratic regression models to explore non-linear age-based trends in identifying chemicals of higher concentrations in children versus those in the older population.

Summary and Insights: Childhood development is a crucial window of susceptibility for the effects of toxic exposures, so prioritizing chemicals that are more common in children can drive future work on studying the effects of these exposures. It can also lead to developing interventions to reduce exposures and prevent subsequent adverse outcomes. To systematically characterize age-based trends of chemical biomarkers, I defined a metric to rank chemical indicators from most to least concerning for children by using the first derivative and second derivative of the chemical biomarker levels with respect to age. The first derivative represents the association between age and the chemical biomarker levels, while the second derivative describes the curvature of the association. Using this metric, I identified several chemical contaminants of elevated and ongoing exposures in children: phthalates, brominated flame retardants, pesticides, and metals, such as lead and tungsten. On the other hand, there were several legacy chemicals such as PCBs and Dioxins of higher concentrations in adults. These legacy chemicals were characterized by having a minimum half-life of one year in the human body and being restricted in North American commerce due to legislation or a voluntary commercial phaseout. Moreover, a targeted analysis of highly persistent PFASs such as PFOS and PFOA with known phaseout dates further confirmed these characteristics of legacy chemicals, since the differences in their biomarker levels between the young and older populations continued to increase since the phaseout period. This implies that persistent chemicals, that have been banned or phased out for longer than their half-lives, have the potential to be legacy chemicals as time passes. This pattern was not observed

for all PFASs, especially those still in use, so we will still be exposed to particular PFASs such as PFNA and PFDA, which will require continued biomonitoring.

Limitations and Future Directions: While I characterized the influence of several important factors in explaining age-based trends in chemical exposures, I did not study which factors among half-lives, time-trends, legislation, and consumer product usages, are most important in explaining these age-based trends. Thus, future work could use machine learning techniques such as feature selection to help identify the most important factors in explaining age-based trends for legacy versus emerging chemicals. The second limitation is that I excluded 182 chemicals that did not satisfy the inclusion threshold of at least 50% detection in the study population. As such, I may have excluded chemicals that exhibit low dose effects or that have exposures only in population subsets. The final limitation is that I did not incorporate toxicological data to prioritize chemicals based on both exposure and toxicity. Now, with the US Environmental Protection Agency Toxicity Forecaster (ToxCast) data on *in vitro* toxicity for thousands of chemicals [400], future work can integrate both biomonitoring and bioactivity data to prioritize chemicals that elicit low-dose and/or high-dose effects.

Contributions: Applying an unbiased, screening approach to biomonitoring data enabled a characterization of age-based trends for a diverse suite of toxicants spanning 16 chemical classes. It also facilitated the ease of identifying chemicals of higher exposure in children versus the older population. My current framework can be coupled with toxicity data to help facilitate chemical prioritization and risk stratification.

6.2 Racial Disparities in Chemical Biomarker Concentrations in United States Women

This study was motivated by a need to understand whether environmental exposures can help explain why the incidence of triple negative breast cancer (TNBC) is approximately three

times higher in Non-Hispanic Black women than in Non-Hispanic White women [79,80]. While the mechanisms driving this difference likely involve interactions between genetic and environmental factors, the influence of environmental risk factors such as chemical exposures may be substantial albeit poorly understood. Thus, I evaluated chemical exposure disparities by race in 38,030 female participants for 143 chemical contaminants using a series of generalized linear models to prioritize chemicals for further experimental work.

Summary and Insights: Non-Hispanic Black women had significantly higher concentrations of many chemicals in their bodies relative to non-Hispanic White women. This included chemicals used in cosmetics such as methyl paraben (2.39-fold), propyl paraben (2.09-fold), and mono-ethyl phthalate (1.74-fold). It also includes 1,4-Dichlorobenzene (3.26-fold), which is used as a disinfectant, pesticide, and deodorant in mothballs, and its metabolite 2,5-dichlorophenol (4.56-fold). Overall, the results showed substantial differences in chemical body burden by race in a representative sample of US women.

Limitations and Future Directions: A major limitation is that we excluded chemicals with detection frequencies below 50%, thereby potentially excluding chemicals that exhibit low dose effects that should have been prioritized. Future studies could therefore benefit from evaluating differences in detection frequency by race to identify additional chemicals that show substantial racial disparities. A second limitation is that this study did not link exposure disparities with toxicological data specifically to breast cancer. However, two ongoing studies are integrating these exposure disparities with ToxCast database on chemical bioactivity [400] and with the Comparative Toxicogenomics Database on chemical-gene interactions [401] to understand why TNBC [79,80] and preterm birth [402,403], respectively, occurs disproportionally in non-Hispanic Black women compared to Non-Hispanic White women.

Contributions: My framework is currently being integrated with ToxCast bioactivity data to enable prioritization of three chemicals, thiram, propyl paraben, and p,p'-DDE, to understand how chemical exposure at environmentally relevant doses in humans can change cell morphology in ways that can progress toward breast cancer in the wet lab. While my framework is being used in the context of breast cancer disparities, it can be coupled with other data characterizing any health effect. Thus, this coupling can help prioritize chemicals to facilitate data-driven experimental work to better understand how the environment drives health disparities.

6.3 Biomarker-Based Occupational Exposome

The main objective of this study was to understand the distribution of occupational exposures to 108 chemicals in a US sample of 26,361 workers across 20 industrial sectors and 19 occupational job titles and the resultant impact on physiological function. I used these job titles to classify participants as blue- or white-collar workers. I then performed clustering analysis on the sector-collar combinations to define occupational groups with similar chemical exposure profiles. I used these exposure-based groups to characterize differences in 27 physiological indicators to identify which occupational groups are most susceptible to chemical-mediated effects on physiological function.

Summary and Insights: I identified higher levels of heavy metals, VOCs, and PAHs in most blue-collar workers across the different sectors compared their white-collar counterparts. In contrast, the white-collar workers had higher exposures to certain heavy metals such as mercury, arsenic, and their metabolites, along with a biomarker of sunscreen use, benzophenone-3. Higher mercury and arsenic biomarker levels among these white-collar workers may indicate higher fish consumption [276]. This could therefore be a surrogate for behaviors associated with higher socioeconomic status [277] instead of an indicator of occupational exposures, since fish is expensive

and is more accessible to those with higher socioeconomic status. Based on these differences in chemical exposures between white- and blue-collar workers, the defined clustered groups were almost mutually exclusive with few clusters including both blue- and white-collar workers. Interestingly, in the clustered groups of blue-collar workers, alkaline phosphatase (biomarker of liver disease or bone disorder), white blood cell count (biomarker of inflammation), and glomerular filtration rate (biomarker of kidney function) showed higher values compared to the reference group of white-collar workers. These results suggest that certain occupational exposures can diminish important physiological functions in blue-collar workers.

Limitations and Future Directions: A primary limitation is the non-random sparsity of the chemical biomarker data in the worker population. As no participant has data available for all studied toxicants, I could not conduct unsupervised clustering of individual workers. Instead, I had to work with groups of workers via the sector-collar combinations. This challenge resulted in conducting clustering analysis on exposure differences among the different sector-collar combinations. These differences then were quantified with the regression coefficients to compare differences in chemical biomarker levels between a given sector-collar combination and the reference group of white-collar workers in public administration. Ideally, if the chemical biomonitoring data were not sparse, then I could compare the clustering method conducted on the regression statistics versus that performed on the raw data to help substantiate the results. Second, due to the sparsity of the occupational data, the analysis was limited by the small sample size, since a quarter of NHANES participants have data on industrial sectors and occupational titles. This prevented me from establishing a direct link between multiple chemical exposures and physiologic dysfunction. As NHANES includes more occupational data with each study year, future analysis could have increased statistical power to model the associations between chemical exposures and

associated effects across different occupational groups. Third, I could not directly compare the chemical biomarker levels to the permissible exposures limits (PELs) defined by the Occupational Safety and Health Administration (OSHA). This prevented me from calculating the percentage of workers above acceptable or safe exposure levels. Thus, future studies can incorporate exposure modeling to predict exposures levels from biomarker levels to enable the comparison to PELs to help identify workers susceptible to high exposures. Fourth, I did not account for confounders such as socioeconomic indicators, alcohol consumption, dietary habits, and smoking habits that may explain differences in occupational exposures across the sector-collars. Hence, future directions can include quantifying differences in these factors by occupation and then account for such differences by including these variables as covariates in the regression models or by conducting stratified analysis in order to better understand if health disparities are due to differences in occupational exposures. Finally, another limitation involved using individual physiological indicators instead of pooling them together to better evaluate overall physiologic dysfunction. Therefore, future studies can define an allostatic load score [339,340] to characterize the overall physiologic dysfunction associated with multiple chemical exposures across the occupational groups.

Contributions: I developed a framework using clustering analysis to characterize the indirect impact of multiple chemical exposures on physiological indicators in the workers population. This method enables me to apply unsupervised learning techniques to non-randomly sparse data without having to define a complete dataset nor impute.

6.4 Characterization of Linear and Non-linear Associations between Physiological Indicators and All-Cause Mortality

This study focused on characterizing linear and non-linear associations between all-cause mortality and each of the 27 physiological indicators, to identify locations in the distribution associated with increased mortality risk. This study is foundational to better understanding the influence of environmental exposures on physiological functions that can impact mortality risk. **Summary and Insights**: I observed that 24 out of 27 (88.9%) indicators show non-linear associations with all-cause mortality, while height, triglycerides, and 60-second pulse show linear associations. Unexpectedly, I observed parabolic associations between cholesterol levels and all-cause mortality. Also, unexpectedly, the results showed a parabolic association between glomerular filtration rate (GFR) and mortality where participants with higher GFR had a higher mortality risk compared to those with lower GFR. The current clinical threshold for GFR, however, only suggests that a value less than 90 mL/min/1.73m^2 implies kidney damage [362,363]. As another unexpected finding, mortality risk for obese participants was lower compared to that for underweight participants.

Limitations and Future Directions: The unexpected lower mortality risk observed in obese patients may indicate the efficacy of public health interventions or improvements in healthcare for obesity-related conditions [377], which was not accounted in our models. Thus, future analysis can study trends of medical and public health interventions for obesity to account for residual confounding, in order to better understand why mortality risk is lower for obese or overweight participants compared to underweight participants. Moreover, the higher mortality risk observed in participants with lower LDL cholesterol may be overestimated due to the inclusion of participants on lipid-lower medications [370,372]. However, it is not well understood whether

mortality risk is higher in participants on lipid-lower medication versus those who are not, even though their LDL cholesterol levels are the same. Thus, future studies could conduct a sensitivity analysis to study whether the associations would change if participants on lipid-lower medications were excluded. Finally, some of the associations may be overestimated due to inclusion of unhealthy participants, who tend to have higher mortality risk. Therefore, future work can include a sensitivity analysis to observe whether these associations would change if unhealthy participants were removed.

Contributions: I developed a machine-learning framework to establish the best model to characterize the relationship with all-cause mortality for a wide variety of physiological indicators. This framework enabled identifying both expected and unexpected directionalities in associations for all-cause mortality. Thus, the framework could be applied to study the associations of any physiological indicators and health endpoints for more accurate risk identification. Furthermore, this framework could be applied to study any associations along the continuum of chemicals indicators to physiological indicators to any health endpoints.

6.5 Integrated Discussion

Drawing inspirations from GWAS for screening across millions of genetic factors to identify those significantly associated with a particular disease, I screened across a wide range of chemicals to characterize biomarker levels across the life stage, race/ethnicity, sex, and occupation in order to identify population susceptible to high exposures, and I screened across a broad suite of physiological indicators to identify populations vulnerable to physiological dysfunction along with directionalities associated with increased mortality risk. I applied a range statistical techniques such as generalized linear regression models, splines models, quantiles model, hierarchical clustering, bootstrapping, and cross validation. Using these techniques, I confirmed

expected findings and identified unexpected findings that would not have been detected if I did not apply these techniques coupled with an untargeted approach.

The chemicals that showed the highest disparities were those that are found in consumer products. In children, several phthalates, brominated flame retardants, pesticides, and metals such as lead and tungsten are found at higher biomarker levels [185]. Phthalates are used as plasticizers to make vinyl flexible and pliant [404]. Due to flexibility of this chemical family, phthalates are detected in flooring, walls, ceiling, food containers, and children's toys [405]. Brominated flame retardants reduce the flammability of product containing them, which includes but not limited to furniture, carpets, pillows, paints, upholstery, and kitchen applicants [406,407]. Lead can be found in household dust, toy jewelry, old toys, plastic toys, cosmetics, and paints [62,408,409]. Pesticides can be found in food, insect repellents, rodent control product, garden care products, and pet products [410,411]. Tungsten is found in jewelry rings and in light bulb filaments [412,413]. Children may be in contact with such products and thus exposed to these chemicals due to normal development behaviors such as crawling [59], mouthing [60,61], and playing [62].

Studying racial disparities in women reiterates the theme of observing higher biomarker levels from chemicals used in consumer products. In non-Hispanic Black, Mexican American, Other Hispanic, and Other Race/Multi-Racial women, pesticides and their metabolites, including 2,5-dichlorophenol, o,p'-DDE, beta-hexachlorocyclohexane, and 2,4-dichlorophenol and personal care and consumer product compounds, including parabens and mono-ethyl phthalate [120]. 2,4-dichlorophenol is a photo-degradation product of triclosan, which is a common antibacterial and antifungal agent used in soap and hand sanitizer [414,415]. 2,5-dichlorophenol is a metabolite of 1,4-dichlorobenzene, which has been used in moth balls, room and toilet deodorizers, and as an insecticidal fumigant [416,417]. o,p'-DDE is a metabolite of DDT, which was developed as an

insecticide but has been banned in 1972 [418,419]. Beta-hexachlorocyclohexane is a byproduct of the production of the insecticide lindane and was deemed as a Persistent Organic Pollutant by the Stockholm Convention [420,421]. Parabens are used as preservatives in personal care products [422,423]. Mono-ethyl phthalate is a metabolite of diethyl phthalate, which is a plasticizer used in plastic packaging, toiletries, fragrance, and cosmetics. Diethyl phthalate is used to extend the aromatic strength of the fragrance, prevent brittleness in nail polishes, and enable hairspray to form a flexible film on the hair [162,424]. Cotinine levels, which is a metabolite of nicotine [425], is the highest in non-Hispanic White women. Interestingly, similar to the findings in children, chemicals showing the highest biomarker levels are also the ones being used in common consumer products, highlighting how the products that we use on a daily basis drives an individual's "chemosome".

Then, this theme is further emphasized when I studied the occupational exposome. Blue-collar workers from retail trade, professional, and technical services show some of the highest biomarker levels of heavy metals such as lead and cadmium, PAHs, phthalates, PFASs, and VOCs such as toluene and benzene compared to any of the other occupations. Furthermore, these chemicals are some of the most toxic substances measured in NHANES. In addition, these blue-collar workers from retail trade, professional, and technical services have a physiological profile characterized by liver damage and elevated kidney filtration, which are both indicative of increased risk for death. If these chemical biomarkers are indicative of occupational exposures and responsible for eliciting these physiological responses, then these findings would imply that handling or producing the consumer products are linked to elevated toxicant levels and subsequently physiological dysfunction conducive to increased risk of mortality.

Every day, we come into contact with many out of the 80,000 chemicals on the market, but only about 1% of them have been evaluated for safety [426]. Though these statistics are daunting, we

do not necessarily need to measure biomarker concentrations for all 80,000 toxicants. The findings in this dissertation suggest that the chemicals found at the highest biomarker levels are those that are used in consumer and personal care products. Thus, biomonitoring studies can measure for toxicants that are commonly used in consumer and personal care products to better prioritize which chemicals to use for further toxicological evaluation.

While I did a comprehensive analysis of characterizing chemical exposure disparities across the life-stage, sex, race/ethnicity, and occupation, I did not ascertain the contribution of chemical exposures to adverse health outcomes nor did I compare these contributions to the contributions of other factors such as genetics, long-term stress, socioeconomic status, education, and health access. These factors are known to impact health outcomes and could even predispose individuals to adverse responses to environmental chemical exposures. This illustrates the need to measure genetic as well as a broad set of environmental factors in as many people as possible. Furthermore, how these genetic and environmental risk factors relate to health is further complicated by the intertwining associations among the risk factors. Thus, untangling this intertwining web will enable the identification of pertinent drivers of health disparities. This insight will be essential in identifying susceptible population and designing targeted intervention to mediate and prevent disease.

6.6 Closing Remarks

Throughout these studies, a quote from Dr. Eberhard Voit resonated with me: "Data generated from the exposome defy human intuition" [427]. This quote was confirmed by the unexpected patterns I observed from applying unbiased approaches to large-scale biomonitoring data in the context of exposome research in data science. This systematic, screening approach is particularly salient to exposome research, as there are more than 80,000 chemicals estimated to be

used in commerce in the US [426]. In addition, there continues to be an increasing availability of multiple layers of data on exposure, toxicity, and human health. Thus, integrating omics techniques with such data will be integral to deriving insights essential for more informed decision-making and prioritization with the goal of protecting human health.

APPENDICES

Appendix 1. Characterization of Age-Based Trends of Chemical Biomarker Levels

Table A1.1.Indicator of excluded measurements by NHANES Cycles for chemical biomarkers.

		Percentage of Participants below LOD by NHANES cycles							
Codename	Comment Name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8
LBXD03	1,2,3,6,7,8-hxcdd (fg/g)	X							
LBXD03LA	1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	X							
LBXD05	1,2,3,4,6,7,8-hpcdd (fg/g)	X							
LBXD05LA	1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	X							
LBXD07	1,2,3,4,6,7,8,9-ocdd (fg/g)	X							
LBXD07LA	1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	X							
LBXF03	2,3,4,7,8-pncdf (fg/g)	X							
LBXF03LA	2,3,4,7,8-pncdf Lipid Adj (pg/g)	X							
LBXF04	1,2,3,4,7,8-hxcdf (fg/g)	X							
LBXF04LA	1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	X							
LBXF08	1,2,3,4,6,7,8-hpcdf (fg/g)	X							
LBXF08LA	1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	X							
URXPAR	Paranitrophenol (ug/L)	X							
LBXHXC	3,3',4,4',5,5'-hxcb (fg/g)	X							
LBXHXCLA	3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	X							
LBXPCB	3,3',4,4',5-pcnb (pg/g)	X							
LBXPCBLA	3,3',4,4',5-pcnb Lipid Adj (pg/g)	X							
LBX028	PCB28 (ng/g)	X							
LBX028LA	PCB28 Lipid Adj (ng/g)	X							
LBD199	PCB199 (ng/g)	X	X						
LBD199LA	PCB199 Lipid Adj (ng/g)	X	X						
LBX180	PCB180 (ng/g)	X	X						

LBX180LA	PCB180 Lipid Adj (ng/g)	X	X						
LBX146	PCB146 (ng/g)	X	X						
LBX146LA	PCB146 Lipid Adj (ng/g)	X	X						
LBX170	PCB170 (ng/g)	X	X						
LBX170LA	PCB170 Lipid Adj (ng/g)	X	X						
LBX194	PCB194 (ng/g)	X	X						
LBX194LA	PCB194 Lipid Adj (ng/g)	X	X						
LBX187	PCB187 (ng/g)	X	X						
LBX187LA	PCB187 Lipid Adj (ng/g)	X	X						
LBX153	PCB153 (ng/g)	X	X						
LBX153LA	PCB153 Lipid Adj (ng/g)	X	X						
LBX196	PCB196 (ng/g)	X	X						
LBX196LA	PCB196 Lipid Adj (ng/g)	X	X						
LBX138	PCB138 (ng/g)	X	X						
LBX138LA	PCB138 Lipid Adj (ng/g)	X	X						
LBX118	PCB118 (ng/g)	X	X						
LBX118LA	PCB118 Lipid Adj (ng/g)	X	X						
LBX099	PCB99 (ng/g)	X	X						
LBX099LA	PCB99 Lipid Adj (ng/g)	X	X						
LBX074	PCB74 (ng/g)	X	X						
LBX074LA	PCB74 Lipid Adj (ng/g)	X	X						
LBXMPAH	2-(N-methyl-PFOSA) acetate (ng/mL)			X					
URX24D	2,4-D (ug/L)	X				X			
URXP10	1-pyrene (ng/L)								X
URXPAR	Paranitrophenol (ug/L)	X							

Table A1.2. Corresponding NHANES codename, CAS NO., and chemical classification for each chemical biomarker.

Codename	Chemical Name	CAS NO.	Chemical Class	Chemical Class Shortened
LBXGLY	Glycideamide (pmoL/G Hb)	5694-00-8	Acrylamide	Acrylamide
LBXACR	Acrylamide (pmoL/G Hb)	79-06-1	Acrylamide	Acrylamide
LBXBB1LA	2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)	59080-40-9	Brominated Flame Retardants (BFR)	BFRs
LBXBR2LA	2,4,4'-tribromodiphenyl ether lipid adj (ng/g)	41318-75-6	Brominated Flame Retardants (BFR)	BFRs
LBXBR8LA	2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)	207122-15-4	Brominated Flame Retardants (BFR)	BFRs
LBXBR6LA	2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)	189084-64-8	Brominated Flame Retardants (BFR)	BFRs
LBXBR3LA	2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)	5436-43-1	Brominated Flame Retardants (BFR)	BFRs
LBXBR5LA	2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)	60348-60-9	Brominated Flame Retardants (BFR)	BFRs
LBXBR7LA	2,2',4,4',5,5'-hexabromodiphenyl lipid adj (ng/g)	68631-49-2	Brominated Flame Retardants (BFR)	BFRs
LBXD03LA	1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	57653-85-7	Dioxins	Dioxins
LBXD07LA	1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	3268-87-9	Dioxins	Dioxins
LBXD05LA	1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	35822-46-9	Dioxins	Dioxins
LBXF04LA	1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	70648-26-9	Furans	Furans
LBXF03LA	2,3,4,7,8-pncdf Lipid Adj (pg/g)	57117-31-4	Furans	Furans
LBXF08LA	1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	67562-39-4	Furans	Furans
SSMEL	Melamine (ng/mL)	108-78-1	Melamine	Melamine
SSCYA	Cyanuric acid (ng/mL)	108-80-5	Melamine	Melamine
LBXBPB	Lead (ug/dL)	7439-92-1	Metals	Metals
LBXBCD	Cadmium (ug/L)	7440-43-9	Metals	Metals
LBXTHG	Mercury, total (ug/L)	7439-97-6	Metals	Metals
URXUHG	Mercury, urine (ng/mL)	7439-97-6	Metals	Metals
URXUBA	Barium, urine (ng/mL)	7440-39-3	Metals	Metals
URXUCO	Cobalt, urine (ng/mL)	7440-48-4	Metals	Metals
URXUCS	Cesium, urine (ng/mL)	7440-46-2	Metals	Metals
URXUMO	Molybdenum, urine (ng/mL)	7439-98-7	Metals	Metals
URXUPB	Lead, urine (ng/mL)	7439-92-1	Metals	Metals
URXUSB	Antimony, urine (ng/mL)	7440-36-0	Metals	Metals
URXUTL	Thallium, urine (ng/mL)	7440-28-0	Metals	Metals
URXUTU	Tungsten, urine (ng/mL)	7440-33-7	Metals	Metals
URXUUR	Uranium, urine (ng/mL)	7440-61-1	Metals	Metals

URXUAS	Urinary total Arsenic (µg/L)	7440-38-2	Metals	Metals
URXUAB	Urinary Arsenobetaine (µg/L)	64436-13-1	Metals	Metals
URXUDMA	Urinary Dimethylarsonic acid (µg/L)	75-60-5	Metals	Metals
URXUCD	Cadmium, urine (ng/mL)	7440-43-9	Metals	Metals
LBXBMN	Blood manganese (ug/L)	7439-96-5	Metals	Metals
LBXSCU	Serum Copper (ug/dL)	7440-50-8	Metals	Metals
LBXSZN	Serum Zinc (ug/dL)	7440-66-6	Metals	Metals
LBXBGM	Mercury, methyl (ug/L)	22967-92-6	Metals	Metals
URXSCN	Urinary thiocyanate (ng/mL)	302-04-5	Other	Other
URXUIO	Iodine, urine (ng/mL)	7553-56-2	Other	Other
URXNO3	Urinary nitrate (ng/mL)	14797-55-8	Other	Other
URXUP8	Perchlorate, urine (ng/mL)	14797-73-0	Other	Other
URXTRS	Urinary Triclosan (ng/mL)	3380-34-5	Personal Care & Consumer Product Compounds	PCCPCs
URXBPS	Urinary Bisphenol S (ug/L)	80-09-1	Personal Care & Consumer Product Compounds	PCCPCs
URXMPB	Methyl paraben (ng/ml)	99-76-3	Personal Care & Consumer Product Compounds	PCCPCs
URXPPB	Propyl paraben (ng/ml)	94-13-3	Personal Care & Consumer Product Compounds	PCCPCs
URXBPH	Urinary Bisphenol A (ng/mL)	80-05-7	Personal Care & Consumer Product Compounds	PCCPCs
URXBPF	Urinary Bisphenol F (ug/L)	620-92-8	Personal Care & Consumer Product Compounds	PCCPCs
URXBP3	Urinary Benzophenone-3 (ng/mL)	131-57-7	Personal Care & Consumer Product Compounds	PCCPCs
LBXTNALA	Trans-nonachlor Lipid Adj (ng/g)	39765-80-5	Pesticides	Pesticides
LBXOXYLA	Oxychlordane Lipid Adj (ng/g)	27304-13-8	Pesticides	Pesticides
LBXPDELA	p,p'-DDE Lipid Adj (ng/g)	72-55-9	Pesticides	Pesticides
LBXDIELA	Dieldrin Lipid Adj (ng/g)	60-57-1	Pesticides	Pesticides
LBXBHCLA	Beta-hexachlorocyclohexane Lipid Adj (ng/g)	319-85-7	Pesticides	Pesticides
URXCPM	3,5,6-trichloropyridinol (ug/L)	6515-38-4	Pesticides	Pesticides
URXOP3	Dimethylthiophosphate (ug/L)	1112-38-5	Pesticides	Pesticides
URXOPM	3-phenoxybenzoic acid (ug/L)	3739-38-6	Pesticides	Pesticides
URX14D	2,5-dichlorophenol (ug/L)	583-78-8	Pesticides	Pesticides
URXDCB	2,4-dichlorophenol (ug/L)	120-83-2	Pesticides	Pesticides
URXPAR	Paranitrophenol (ug/L)	100-02-7	Pesticides	Pesticides
URX24D	2,4-D (ug/L)	94-75-7	Pesticides	Pesticides
URXDEA	DEET acid (ug/L)	72236-23-8	Pesticides	Pesticides
URXCNP	Mono(carboxynonyl) phthalate (ng/mL)	26761-40-0	Phthalates	Phthalates

SSURHIBP	Mono-2-hydroxy-iso-butyl phthlte (ng/mL)	64339-39-5	Phthalates	Phthalates
URXCOP	Mono(carboxyoctyl) phthalate (ng/mL)	898544-09-7	Phthalates	Phthalates
URXMIB	Mono-isobutyl pthalate (ng/mL)	30833-53-5	Phthalates	Phthalates
URXMHP	Mono-(2-ethyl)-hexyl phthalate (ng/mL)	4376-20-9	Phthalates	Phthalates
URXMOH	Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)	40321-98-0	Phthalates	Phthalates
URXMZP	Mono-benzyl phthalate (ng/mL)	2528-16-7	Phthalates	Phthalates
URXMHH	Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)	40321-99-1	Phthalates	Phthalates
URXECP	Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)	40809-41-4	Phthalates	Phthalates
URXMBP	Mono-n-butyl phthalate (ng/mL)	131-70-4	Phthalates	Phthalates
URXMEP	Mono-ethyl phthalate (ng/mL)	2306-33-4	Phthalates	Phthalates
URXMNM	Mono-n-methyl phthalate (ng/mL)	4376-18-5	Phthalates	Phthalates
SSURMHBP	Mono-3-hydroxy-n-butyl phthalate (ng/mL)	57074-43-8	Phthalates	Phthalates
URXMC1	Mono-(3-carboxypropyl) phthalate (ng/mL)	66851-46-5	Phthalates	Phthalates
URXDMA	o-Desmethylangolensin (O-DMA) (ng/mL)	21255-69-6	Phytoestrogens	Phytoestrogens
URXETL	Enterolactone (ng/mL)	78473-71-9	Phytoestrogens	Phytoestrogens
URXETD	Enterodiol (ng/mL)	80226-00-2	Phytoestrogens	Phytoestrogens
URXP05	3-phenanthrene (ng/L)	605-87-8	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP06	1-phenanthrene (ng/L)	2433-56-9	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP07	2-phenanthrene (ng/L)	605-55-0	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP19	4-phenanthrene (ng/L)	7651-86-7	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP25	2 & 3-Hydroxyphenanthrene (ng/L)	605-55-0	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP10	1-pyrene (ng/L)	129-00-0	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP01	1-napthol (ng/L)	90-15-3	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP02	2-napthol (ng/L)	135-19-3	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP17	9-fluorene (ng/L)	484-17-3	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP03	3-fluorene (ng/L)	6344-67-8	Polyaromatic Hydrocarbons (PAH)	PAHs
URXP04	2-fluorene (ng/L)	2443-58-5	Polyaromatic Hydrocarbons (PAH)	PAHs
LBD199LA	PCB199 Lipid Adj (ng/g)	52663-75-9	Polychlorinated Biphenyls (PCB)	PCBs

LBX180LA	PCB180 Lipid Adj (ng/g)	35065-29-3	Polychlorinated Biphenyls (PCB)	PCBs
LBX209LA	PCB209 Lipid Adj (ng/g)	2051-24-3	Polychlorinated Biphenyls (PCB)	PCBs
LBX146LA	PCB146 Lipid Adj (ng/g)	51908-16-8	Polychlorinated Biphenyls (PCB)	PCBs
LBX170LA	PCB170 Lipid Adj (ng/g)	35065-30-6	Polychlorinated Biphenyls (PCB)	PCBs
LBX194LA	PCB194 Lipid Adj (ng/g)	35694-08-7	Polychlorinated Biphenyls (PCB)	PCBs
LBX187LA	PCB187 Lipid Adj (ng/g)	52663-68-0	Polychlorinated Biphenyls (PCB)	PCBs
LBX153LA	PCB153 Lipid Adj (ng/g)	35065-27-1	Polychlorinated Biphenyls (PCB)	PCBs
LBX196LA	PCB196 Lipid Adj (ng/g)	42740-50-1	Polychlorinated Biphenyls (PCB)	PCBs
LBX138LA	PCB138 Lipid Adj (ng/g)	35065-28-2	Polychlorinated Biphenyls (PCB)	PCBs
LBXHXCLA	3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	32774-16-6	Polychlorinated Biphenyls (PCB)	PCBs
LBX118LA	PCB118 Lipid Adj (ng/g)	31508-00-6	Polychlorinated Biphenyls (PCB)	PCBs
LBX099LA	PCB99 Lipid Adj (ng/g)	38380-01-7	Polychlorinated Biphenyls (PCB)	PCBs
LBXPCBLA	3,3',4,4',5-pcnb Lipid Adj (pg/g)	57465-28-8	Polychlorinated Biphenyls (PCB)	PCBs
LBX074LA	PCB74 Lipid Adj (ng/g)	32690-93-0	Polychlorinated Biphenyls (PCB)	PCBs
LBX028LA	PCB28 Lipid Adj (ng/g)	7012-37-5	Polychlorinated Biphenyls (PCB)	PCBs
LBX049LA	PCB49 Lipid Adj (ng/g)	41464-40-8	Polychlorinated Biphenyls (PCB)	PCBs
LBX044LA	PCB44 Lipid Adj (ng/g)	41464-39-5	Polychlorinated Biphenyls (PCB)	PCBs
LBXPFDE	Perfluorodecanoic acid (ng/mL)	335-76-2	Per- and Polyfluoroalkyl Substances (PFAS)	PFASs
LBXMPAH	2-(N-methyl-PFOSA) acetate (ng/mL)	2355-31-9	Per- and Polyfluoroalkyl Substances (PFAS)	PFASs
LBXPFNA	Perfluorononanoic acid (ng/mL)	375-95-1	Per- and Polyfluoroalkyl Substances (PFAS)	PFASs
LBXPFHS	Perfluorohexane sulfonic acid (ng/mL)	355-46-4	Per- and Polyfluoroalkyl Substances (PFAS)	PFASs
LBXPFOA	Perfluorooctanoic acid (ng/mL)	335-67-1	Per- and Polyfluoroalkyl Substances (PFAS)	PFASs
LBXPFOS	Perfluorooctane sulfonic acid (ng/mL)	1763-23-1	Per- and Polyfluoroalkyl Substances (PFAS)	PFASs
LBXCOT	Cotinine (ng/mL)	486-56-6	Smoking Related Compounds	SRCs
URXNAL	NNAL , urine (ng/mL)	76014-81-8	Smoking Related Compounds	SRCs
LBXVXY	Blood m-/p-Xylene (ng/ml)	108-38-3/106-42-3	Volatile Organic Compounds (VOC)	VOCs
LBXVDB	Blood 1,4-Dichlorobenzene (ng/ml)	106-46-7	Volatile Organic Compounds (VOC)	VOCs

LBXVBM	Blood Bromodichloromethane (pg/ml)	75-27-4	Volatile Organic Compounds (VOC)	VOCs
URXHEM	N-Ace-S-(2-Hydroxyethyl)-L-cys (ng/mL)	15060-26-1	Volatile Organic Compounds (VOC)	VOCs
URXPHG	Phenylglyoxylic acid (ng/mL)	611-73-4	Volatile Organic Compounds (VOC)	VOCs
URXHP2	N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)	75-56-9	Volatile Organic Compounds (VOC)	VOCs
URXMB3	N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)	106-99-0	Volatile Organic Compounds (VOC)	VOCs
URX34M	3-methipurc acd & 4-methipurc acd (ng/mL)	27115-49-7	Volatile Organic Compounds (VOC)	VOCs
URXBMA	N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)	19542-77-9	Volatile Organic Compounds (VOC)	VOCs
URXBPM	N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)	106-94-5	Volatile Organic Compounds (VOC)	VOCs
URXGAM	N-ac-S-(2-carbmo-2-hydxel)-L-cys (ng/mL)	79-06-1	Volatile Organic Compounds (VOC)	VOCs
URXMAD	Mandelic acid (ng/mL)	90-64-2	Volatile Organic Compounds (VOC)	VOCs
URXAAM	N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)	81690-92-8	Volatile Organic Compounds (VOC)	VOCs
URXAMC	N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)	103974-29-4	Volatile Organic Compounds (VOC)	VOCs
URXATC	2-amnothiazolne-4-carbxylic acid (ng/mL)	16899-18-6	Volatile Organic Compounds (VOC)	VOCs
URXCYM	N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)	74514-75-3	Volatile Organic Compounds (VOC)	VOCs
URXTTC	2-thoxothazlidne-4-carbxylic acid (ng/mL)	20933-67-9	Volatile Organic Compounds (VOC)	VOCs
URX2MH	2-Methylhippuric acid (ng/mL)	42013-20-7	Volatile Organic Compounds (VOC)	VOCs
URXPMM	N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)	33164-70-4	Volatile Organic Compounds (VOC)	VOCs
URXCEM	N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)	51868-61-2	Volatile Organic Compounds (VOC)	VOCs
URXHPM	N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)	23127-40-4	Volatile Organic Compounds (VOC)	VOCs
URXDHB	N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)	144889-50-9	Volatile Organic Compounds (VOC)	VOCs
LBXVTO	Blood Toluene (ng/ml)	108-88-3	Volatile Organic Compounds (VOC)	VOCs
LBXNM	Blood Nitromethane (pg/mL)	75-52-5	Volatile Organic Compounds (VOC)	VOCs
LBXVCF	Blood Chloroform (pg/ml)	67-66-3	Volatile Organic Compounds (VOC)	VOCs

Text A1.1. Overestimation of QSARs half-lives for PFASs and Other Estimation Methods

The estimated persistency of QSAR PFASs varies drastically from 3.5 years to 220 years. The half-lives of PFOS (4.80 years) and PFOA (5.42 years) from Arnot et al, 2014 Training Dataset are comparable with other estimated half-lives such as 5 years for PFOS and 2.3 to 4 years for PFOA. In addition, the QSAR-estimated half-life of PFHxS (7.29 years) is comparable to 8.5 years. A one-compartment model was used to estimate the half-lives for PFASs in a sample of Chinese volunteers in Shijiazhuang and Handan and estimated to be 0.38 to 20 years for the half-life of PFNA and 1.2 to 60 years for PFDA, which are very different from the QSAR-estimated half-lives of PFNA (45.03 years) and PFDA (205.82 years). There are no measured or estimated half-lives for 2-(N-methyl-PFOSA) acetate in human.

This model is a screening-level, fragment-based QSAR that for instance, focuses on "fragments" of the compound i.e. F-C bond. Since PFNA have several of these bonds and such bonds are associated with high persistency, the QSAR predicted a higher persistency for these cases. In addition, half-life of PFOS, PFOA, and PFHxS were the only ones used in the training set to train the QSAR model to predict half-lives for PFASs. So all of these stipulations can lead to overestimations on some of the predictions.

Thus, to address these overestimations, we looked at different literature sources for the half-lives and defined a hierarchy to determine more reasonable half-lives for these substances. When multiple human biomonitoring-based half-lives are available for a given chemical, the median of all available half-lives is used, or else we perform an extrapolation from animal data using the ratio of the half-life of PFOSA to the half-life of PFOS from rat (10.6 and 30 days, respectively) and rainbow trout (6 and 16.9 days, respectively). So, the extrapolation factors were 0.355 for rainbow trout and 0.353 for rats.

Table A1.3. Authorship for half-lives of PFASs with other estimation or extrapolation methods.

Studies	Type of HL	Types	LBXPFDE	LBXMPAH	LBXPFNA	LBXPFHS	LBXPFOA	LBXPFOS
Arnot et al., 2014	intrinsic	Training Dataset				7.3	5.4	4.8
		QSAR	205.8	45	45	2.2	9.9	45
Olsen et al., 2017	apparent					7.3	3.5	4.8
Bartell et al., 2010	intrinsic						2.3	
Zhang et al., 2013	renal and menstrual clearance	Young Females	4.2		1.5	2.3	1.8	6
	renal clearance	All Males and Older Females	9.2		3.5	29	1.7	18
Gomis et al., 2017	intrinsic	Female Australian					1.8	5
		Male Australian					2	4.9
		Female US					2.1	3.3
		Male US					2.4	3.8
Wong et al., 2018	intrinsic	Male						4.7
		Female (exclude loss from menstruation)						3.7
		Female (include loss from menstruation)						4
Russell et al., 2015	intrinsic						2.4	
Gomis et al., 2016	intrinsic						2.4	
Brede et al., 2010	intrinsic						3.26	
Spliethoff et al., 2008	apparent							4.4
D'Eon and Mabury, 2011	apparent							5.4
Olsen et al., 2012	apparent							4.3
Glynn et al., 2012	apparent							8.2
Yeung et al., 2013	apparent							4.55

Worley et al., 2017	intrinsic					15.5	3.9	3.3
Conglomerated Estimation or Animal-Human Extrapolation	years		6.7	1.664653	2.5	7.3	2.35	4.7
	hours		58692	14582.36	21900	63948	20586	41172
	log10 hours		4.768579	4.163828	4.340444	4.805827	4.313572	4.614602

Table A1.4. References for half-lives of inorganic substances for which the half-life could not be estimated by the QSAR models. Half-Lives (hours) were used in the analysis.

Chemical	References for Half-Lives	Half-Lives (units)	Half-Lives (hours)	Biological Compartments
Propineb	Watson M. IPCS INCHEM - PROPINEB. Available: http://www.inchem.org/documents/jmpr/jmpmono/v93pr16.htm.	20 days	480	Body
Arsenic	Bradberry SM, Harrison WN, Beer ST, Vale JA. 1997. IPCS INCHEM - arsenic acid. Available: http://www.inchem.org/documents/ukpids/ukpids/ukpid41.htm.	60 hours	60	Blood, Serum, Plasma
	ATSDR. 2007. Toxicological profile for arsenic. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp2.pdf.	4 hours 3 hours 60 hours 40 hours	4 3 60 40	Blood, Serum, Plasma Blood, Serum, Plasma Body Body
Arsenic Acid	National Center for Biotechnology Information. Pubchem compound database cid=234: Arsenic acid. Available: https://pubchem.ncbi.nlm.nih.gov/compound/234.	35 hours 2.40 days	35 57.6	Blood, Serum, Plasma Body
Arsenobetaine	Lehmann B, Ebeling E, Alsen-Hinrichs C. 2001. Kinetics of arsenic in human blood after a fish meal. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)) 63:42-48.	63 hours	63	Blood, Serum, Plasma
Methanearsonic acid	ATSDR. 2007. Toxicological profile for arsenic. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp2.pdf.	4.90 hours 4.20 hours	4.9 4.2	Blood, Serum, Plasma Blood, Serum, Plasma
Trimethylarsine Oxide	CDC. 2009. Fourth national report on human exposure to environmental chemicals. Available: https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf.	4 days 2 days	96 48	Body Body
Antimony	ATSDR. 2017. Toxicological profile for antimony and compounds. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp23.pdf.	62 days 47 days 1.99 hours	1488 1128 1.99	Lung Lung Blood, Serum, Plasma
	Gerhardsson L, Brune D, Nordberg GF, Wester PO. 1982. Antimony in lung, liver and kidney tissue from deceased smelter workers. Scandinavian journal of work, environment & health 8:201-208.	100 days 36 days	2400 864	Body Body

Barium	Bastarache E. Barium and compounds / toxicology. Available: https://digitalfire.com/4sight/hazards/ceramic_hazard_barium_and_compounds__toxicology_321.html.	20 hours 2 hours	20 2	Body Body
	Ayres DC, Hellier DG. 1997. Dictionary of environmentally important chemicals. Available: https://books.google.com/books?id=UTKWehimCkEC&pg=PA44&lpg=PA44&dq=barium+half+life+blood&source=bl&ots=r-sg5_z8vf&sig=3cyavKJmEbMfXCQV_heAGywb6Bw&hl=en&sa=X&ved=0ahUKEwjwp5PA59nVAhUB2mMKHaWzBiMQ6AEITjAI#v=onepage&q=barium half life blood&f=false.	50 days	1200	Bones
Beryllium	ATSDR. 2002. Toxicological profile for beryllium. Available: https://www.atsdr.cdc.gov/toxprofiles/tp4.pdf..	400 days 300 days 833 days 8 weeks 2 weeks	9600 7200 19992 1344 336	Body Body Lung Blood, Serum, Plasma Blood, Serum, Plasma
	CDC. 2009. Fourth national report on human exposure to environmental chemicals. Available: https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf.	450 days	10800	Bones
Cacodylic Acid	National Center for Biotechnology I. Pubchem compound database cid=2513: Cacodylic acid. Available: https://pubchem.ncbi.nlm.nih.gov/compound/2513.	92 days 90 days 76 days	2208 2160 1824	Blood, Serum, Plasma Blood, Serum, Plasma Blood, Serum, Plasma
Cadmium	Scoullos M, Vonkeman GH, Thornton I, Makuch Z. 2001. Mercury — cadmium — lead handbook for sustainable heavy metals policy and regulation. 1st ed:Springer Netherlands.	30 years	262800	Body
	Bernhoft RA. 2013. Cadmium toxicity and treatment. The Scientific World Journal 2013.	128 days 75 days 25 years	3072 1800 219000	Blood, Serum, Plasma Blood, Serum, Plasma Body
Cesium	ATSDR. 2004. Toxicological profile for cesium. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp157.pdf.	150 days 135 days 110 days 110 days 93 days 90 days 84 days 83 days 73 days 72 days	3600 3240 2640 2640 2232 2160 2016 1992 1752 1728	Body Body Body Body Body Body Body Body Body Body

		70 days 67 days 66 days 58 days 50 days 47 days 46 days	1680 1608 1584 1392 1200 1128 1104	Body Body Body Body Body Body Body
	CDC. 2009. Fourth national report on human exposure to environmental chemicals. Available: https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf.	70 days 109 days	1680 2616	Body Body
Cobalt	Tvermoes BE, Unice KM, Paustenbach DJ, Finley BL, Otani JM, Galbraith DA. 2014. Effects and blood concentrations of cobalt after ingestion of 1 mg/d by human volunteers for 90 d. The American journal of clinical nutrition 99:632-646.	36 days 22 days	864 528	Blood, Serum, Plasma Blood, Serum, Plasma
	ATSDR. 2004. Toxicological profile for cobalt. Available: https://www.atsdr.cdc.gov/toxprofiles/tp33.pdf.	99.50 days	2388	Body
	CDC. 2009. Fourth national report on human exposure to environmental chemicals. Available: https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf.	2 years 1 year	17520 8760	Lung Lung
Copper	Johnson PE, Milne DB, Lykken GI. 1992. Effects of age and sex on copper absorption, biological half-life, and status in humans. The American journal of clinical nutrition 56:917-925.	33 days 13 days	792 312	Body Body
	Lyon TDB, Fell GS, Gaffney D, McGaw BA, Russell RI, Park RHR, et al. 1995. Use of a stable copper isotope in the differential diagnosis of wilson's disease. Clinical Science 88:727 LP-732.	14 days 26 days	336 624	Blood, Serum, Plasma Blood, Serum, Plasma
Lead	ATSDR. 2007. Toxicological profile for lead. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp13.pdf.	27 years	236520	Bones
	ATSDR. Toxguide for lead. Available: https://www.atsdr.cdc.gov/toxguides/toxguide-13.pdf.	30 days	720	Blood, Serum, Plasma
	National Research Council Committee on Measuring Lead in Critical P. 1993. 4, biologic markers of lead toxicity. Available: https://www.ncbi.nlm.nih.gov/books/NBK236462/.	25 days	600	Blood, Serum, Plasma
	Barbosa F, Tanus-Santos JE, Gerlach RF, Parsons PJ. 2005. A critical review of biomarkers used for monitoring human exposure to lead: Advantages, limitations,	30 years 10 years	262800 87600	Bones Bones

	and future needs. Environmental Health Perspectives 113:1669-1674.			
	Rahde AF. 1994. Ipcs inchem - lead, inorganic. Available: http://www.inchem.org/documents/pims/chemical/inorglea.htm#SectionTitle:6.3 Biological half-life by route of exposure.	30 days 20 days 30 years 20 years	720 480 262800 175200	Blood, Serum, Plasma Blood, Serum, Plasma Bones Bones
	Oregon Department of Human Services. Health effects of lead exposure. Available: http://www.oregon.gov/oha/PH/HEALTHYENVIRONMENTS/HEALTHYNEIGHBORHOODS/LEADPOISONING/MEDICALPROVIDERSLABORATORIES/Documents/introhealtheffectsmedicalprovider.pdf.	30 years 25 years	262800 219000	Bones Bones
	CDC. 2016. Biomonitoring summary: Lead. Available: https://www.cdc.gov/biomonitoring/Lead_BiomonitoringSummary.html.	2 months 1 month	1440 720	Blood, Serum, Plasma Blood, Serum, Plasma
	NCHS. 2012. Laboratory procedure manual: Cadmium, Lead, Manganese, Mercury, and Selenium. Available: https://www.cdc.gov/nchs/data/nhanes/nhanes_11_12/PbCd_G_met_blood%20metals.pdf.	30 days	720	Blood, Serum, Plasma
	Swedish Chemicals Agency. 2012. Proposal for harmonised classification and labelling: Lead. Available: https://echa.europa.eu/documents/10162/13626/lead_clh_proposal_en.pdf.	40 days	960	Blood, Serum, Plasma
	Dobbs MR. 2009. Clinical neurotoxicology: Syndromes, substances, environments. Available: https://books.google.com/books/about/Clinical_Neurotoxicology.html?id=Pmcy24y2HyMC.	36 days	864	Blood, Serum, Plasma
Manganese	O'Neal SL, Zheng W. 2015. Manganese toxicity upon overexposure: A decade in review. Current environmental health reports 2:315-328.	8 years 9 years	70080 78840	Bones Bones
	Santamaria AB. 2008. Manganese exposure, essentiality & toxicity. The Indian journal of medical research 128:484-500.	42 days 10 days	1008.00 240.00	Blood, Serum, Plasma Blood, Serum, Plasma
	Mayo Foundation for Medical E, Research. Test ID: MNB - manganese, blood. Available: https://www.mayomedicallaboratories.com/test-catalog/Clinical+and+Interpretive/89120.	40 days	960	Body
	O'Neal SL, Hong L, Fu S, Jiang W, Jones A, Nie LH, et al. 2014. Manganese accumulation in bone following chronic exposure in rats: Steady-state	8.50 years	74460	Bones

	concentration and half-life in bone. Toxicology Letters 229:93-100.			
Mercury	ATSDR. 2000. Managing hazardous material incidents: Mercury (hg). Available: https://www.atsdr.cdc.gov/mhmi/mmg46.pdf.	90 days 60 days	2160 1440	Body Body
	US EPA. 1984. Mercury health effects update. Available: https://pubchem.ncbi.nlm.nih.gov/compound/mercury#section=Biological-Half-Life.	30 days 70 days 50 days	720 1680 1200	Blood, Serum, Plasma Body Body
	Hyman MH. 2004. The impact of mercury on human health and the environment. 10:70-75.	18 years 1 year	157680 8760	Bones Bones
	ATSDR. 1999. Toxicological profile for mercury. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp46.pdf.	45 days 40.50 days 36 days	1080 972 864	Blood, Serum, Plasma Blood, Serum, Plasma Body
Ethyl Mercury	WHO. 2006. Statement on thiomersal. Available: http://www.who.int/vaccine_safety/committee/topics/thiomersal/statement_jul2006/en/.	1 week	168	Blood, Serum, Plasma
	University of Minnesota E, Occupational H. Mercury: Dose, absorption, distribution, biotransformation, and excretion of mercury. Available: http://enhs.umn.edu/current/5103_spring2003/mercury/mercdose.html.	50 days	1200	Body
Methyl Mercury	Jo S, Woo HD, Kwon H-J, Oh S-Y, Park J-D, Hong Y-S, et al. 2015. Estimation of the biological half-life of methylmercury using a population toxicokinetic model. International Journal of Environmental Research and Public Health 12:9054-9067.	88.8 days 71.6 days 35.1 days 52.8 days	2131.20 1718.40 842.40 1267.20	Blood, Serum, Plasma Blood, Serum, Plasma Blood, Serum, Plasma Blood, Serum, Plasma
	WHO. 2006. Statement on thiomersal. Available: http://www.who.int/vaccine_safety/committee/topics/thiomersal/statement_jul2006/en/.	1.5 months	1080	Blood, Serum, Plasma
	ATSDR. 1999. Toxicological profile for mercury. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp46.pdf.	65 days 48 days	1560 1152	Blood, Serum, Plasma Blood, Serum, Plasma
	CDC. 2009. Fourth national report on human exposure to environmental chemicals. Available: https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf.	50 days	1200	Body
Inorganic Mercury	CDC. 2016. Biomonitoring summary: Mercury. Available:	3 weeks 1 week	504 168	Blood, Serum, Plasma

	https://www.cdc.gov/biomonitoring/Mercury_BiomonitoringSummary.html.			Blood, Serum, Plasma
	ATSDR. 2000. Managing hazardous material incidents: Mercury (hg). Available: https://www.atsdr.cdc.gov/mhmi/mmg46.pdf.	15 days 28 days 60 days 60 days	360 672 1440 1440	Blood, Serum, Plasma Blood, Serum, Plasma Body Body
Molybdenum	ATSDR. 2017. Toxicological profile for molybdenum. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp212.pdf.	6.60 hours	6.60	Blood, Serum, Plasma
	Vyskočil, Adolf, and Claude Viau. "Assessment of molybdenum toxicity in humans." Journal of Applied Toxicology 19.3 (1999): 185-192.	19 hours	19	Body
Perchlorate	Srinivasan A, Viraraghavan T. 2009. Perchlorate: Health effects and technologies for its removal from water resources. International Journal of Environmental Research and Public Health 6:1418-1442.	8 hours 6 hours	8 6	Blood, Serum, Plasma Blood, Serum, Plasma
	Crump KS, Gibbs JP. 2005. Benchmark calculations for perchlorate from three human cohorts. Environmental Health Perspectives 113:1001-1008.	7.50 hours	7.50	Blood, Serum, Plasma
	ATSDR. 2008. Toxicological profile for perchlorates. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp162.pdf.	8 hours	8	Blood, Serum, Plasma
	WHO. 2004. Perchlorate in Drinking-water Background document for development of WHO Guidelines for Drinking-water Quality. Available: http://www.who.int/water_sanitation_health/water-quality/guidelines/chemicals/perchlorate-background-jan17.pdf	6 hours 9.30 hours	6 9.30	Blood, Serum, Plasma Blood, Serum, Plasma
Platinum	Herr CEW, Jankofsky M, Angerer J, Kuster W, Stilianakis NI, Gieler U, et al. 2003. Influences on human internal exposure to environmental platinum. Journal of exposure analysis and environmental epidemiology 13:24-30.	720 days	17280	Body
Thallium	AcuteTox. Thallium sulfate. Available: http://www.acutetox.eu/pdf_human_short/66-Thallium sulphate revised.pdf.	8 days	192	Blood, Serum, Plasma
	Quest Diagnostics. Thallium, blood. Available: http://www.questdiagnostics.com/testcenter/testguide.action%3Fdc%3DTH_Thallium.	4 days 2 days	96 48	Blood, Serum, Plasma Blood, Serum, Plasma
Thiomersal	Burbacher, Thomas M. et al. “Comparison of Blood and Brain Mercury Levels in Infant Monkeys Exposed to	8.60 days	206.40	Blood, Serum, Plasma

	Methylmercury or Vaccines Containing Thimerosal." Environmental Health Perspectives 113.8 (2005): 1015–1021. PMC. Web. 6 Sept. 2018.			
Tungsten	Lemus R, Venezia CF. 2015. An update to the toxicological profile for water-soluble and sparingly soluble tungsten substances. Critical reviews in toxicology 45:388-411.	67 days 6 days	1608 144	Body Body
	ATSDR. 2005. Toxicological profile for tungsten. Available: https://www.atsdr.cdc.gov/toxprofiles/tp186.pdf.	23 years 4 years 100 days	201480 35040 2400	Bones Bones Bones
Uranium	ATSDR. 2013. Toxicological profile for uranium. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp150.pdf.	200 days 70 days 240 days 110 days 2 weeks 1 week	4800 1680 5760 2640 336 168	Bones Bones Lungs Lungs Body Body
Zinc	INCHEM I. IPCS INCHEM - ZINC. Available: http://www.inchem.org/documents/jecfa/jecmono/v17je33.htm.	500 days 300 days	12000 7200	Body Body
	Nriagu J. 2007. Zinc deficiency in human health. School of Public Health.	280 days	6720	Body

Table A1.5. Maximum composite half-life in hours, log-transformed half-life, and types of methods on how the half-lives were determined for each chemical biomarker.

NHANES Codename	Chemical Name (units)	Maximum Composite Half-Lives (hours)	Log 10(Maximum Composite Half-Lives)	Method Types to Find or Estimate Half-Lives
LBXGLY	Glycideamide (pmoL/G Hb)	4.716143	0.673587	Estimated by QSARs
LBXACR	Acrylamide (pmoL/G Hb)	2.238721	0.35	Estimated by QSARs
LBXBB1LA	2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)	56940	4.755417	Arnot et al., 2014 Training Set
LBXBR2LA	2,4,4'-tribromodiphenyl ether lipid adj (ng/g)	26400	4.421604	Arnot et al., 2014 Training Set
LBXBR8LA	2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)	18214.72	4.260422	Arnot et al., 2014 Training Set
LBXBR6LA	2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)	14831.67	4.17119	Arnot et al., 2014 Training Set
LBXBR3LA	2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)	13818.15	4.14045	Arnot et al., 2014 Training Set
LBXBR5LA	2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)	12699.6	4.10379	Arnot et al., 2014 Training Set
LBXBR7LA	2,2',4,4',5,5'-hexabromodiphenyl lipid adj (ng/g)	6194.411	3.792	Estimated by QSARs
LBXD03LA	1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	244555	5.388377	Arnot et al., 2014 Training Set
LBXD07LA	1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	66104.34	4.82023	Arnot et al., 2014 Training Set
LBXD05LA	1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	39400.41	4.595501	Arnot et al., 2014 Training Set
LBXF04LA	1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	60954.86	4.785008	Arnot et al., 2014 Training Set
LBXF03LA	2,3,4,7,8-pncdf Lipid Adj (pg/g)	51002.33	4.70759	Arnot et al., 2014 Training Set
LBXF08LA	1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	22792.4	4.35779	Arnot et al., 2014 Training Set
SSMEL	Melamine (ng/mL)	3.404082	0.532	Estimated by QSARs
SSCYA	Cyanuric acid (ng/mL)	2.238721	0.35	Estimated by QSARs

LBXBPB	Lead (ug/dL)	262800	5.419625	Literature
LBXBCD	Cadmium (ug/L)	262800	5.419625	Literature
LBXTHG	Mercury, total (ug/L)	157680	5.197777	Literature
URXUHG	Mercury, urine (ng/mL)	157680	5.197777	Literature
URXUBA	Barium, urine (ng/mL)	1200	3.079181	Literature
URXUCO	Cobalt, urine (ng/mL)	17520	4.243534	Literature
URXUCS	Cesium, urine (ng/mL)	3600	3.556303	Literature
URXUMO	Molybdenum, urine (ng/mL)	19	1.278754	Literature
URXUPB	Lead, urine (ng/mL)	262800	5.419625	Literature
URXUSB	Antimony, urine (ng/mL)	2400	3.380211	Literature
URXUTL	Thallium, urine (ng/mL)	192	2.283301	Literature
URXUTU	Tungsten, urine (ng/mL)	201480	5.304232	Literature
URXUUR	Uranium, urine (ng/mL)	5760	3.760422	Literature
URXUAS	Urinary total Arsenic (µg/L)	60	1.778151	Literature
URXUAB	Urinary Arsenobetaine (µg/L)	98	1.991226	Literature
URXUDMA	Urinary Dimethylarsonic acid (µg/L)	2243	3.350829	Literature
URXUCD	Cadmium, urine (ng/mL)	262800	5.419625	Literature
LBXBMN	Blood manganese (ug/L)	78840	4.896747	Literature
LBXSCU	Serum Copper (ug/dL)	792	2.898725	Literature
LBXSZN	Serum Zinc (ug/dL)	12000	4.079181	Literature
LBXBGM	Mercury, methyl (ug/L)	3360	3.526339	Literature
URXSCN	Urinary thiocyanate (ng/mL)	4.477442	0.65103	Estimated by QSARs
URXUIO	Iodine, urine (ng/mL)	744	2.871573	Literature
URXNO3	Urinary nitrate (ng/mL)	1.588547	0.201	Estimated by QSARs
URXUP8	Perchlorate, urine (ng/mL)	9.3	0.968483	Literature
URXTRS	Urinary Triclosan (ng/mL)	254.0973	2.405	Estimated by QSARs
URXBPS	Urinary Bisphenol S (ug/L)	7.144963	0.854	Estimated by QSARs
URXMPB	Methyl paraben (ng/ml)	2.09894	0.322	Estimated by QSARs
URXPPB	Propyl paraben (ng/ml)	2.09894	0.322	Estimated by QSARs
URXBPH	Urinary Bisphenol A (ng/mL)	2	0.30103	Arnot et al., 2014 Training Set
URXBPF	Urinary Bisphenol F (ug/L)	1.96336	0.293	Estimated by QSARs
URXBP3	Urinary Benzophenone-3 (ng/mL)	1.202264	0.08	Estimated by QSARs
LBXTNALA	Trans-nonachlor Lipid Adj (ng/g)	685510.4	5.836014	Estimated by QSARs
LBXOXYLA	Oxychlordane Lipid Adj (ng/g)	624539.5	5.79556	Estimated by QSARs

LBXPDELA	p,p'-DDE Lipid Adj (ng/g)	77236.13	4.887821	Arnot et al., 2014 Training Set
LBXDIELA	Dieldrin Lipid Adj (ng/g)	36525.47	4.562596	Arnot et al., 2014 Training Set
LBXBHCLA	Beta-hexachlorocyclohexane Lipid Adj (ng/g)	3365.348	3.52703	Estimated by QSARs
URXCPM	3,5,6-trichloropyridinol (ug/L)	197.6464	2.295889	Estimated by QSARs
URXOP3	Dimethylthiophosphate (ug/L)	117.1998	2.068927	Estimated by QSARs
URXOPM	3-phenoxybenzoic acid (ug/L)	67.02382	1.826229	Estimated by QSARs
URX14D	2,5-dichlorophenol (ug/L)	39.88466	1.600806	Estimated by QSARs
URXDCB	2,4-dichlorophenol (ug/L)	35.81212	1.55403	Estimated by QSARs
URXPAR	Paranitrophenol (ug/L)	13.95087	1.144601	Estimated by QSARs
URX24D	2,4-D (ug/L)	32.9997	1.51851	Arnot et al., 2014 Training Set
URXDEA	DEET acid (ug/L)	9.29979	0.968473	Estimated by QSARs
URXCNP	Mono(carboxynonyl) phthalate (ng/mL)	10.35214	1.01503	Estimated by QSARs
SSURHIBP	Mono-2-hydroxy-iso-butyl phthlte (ng/mL)	5.936395	0.773523	Estimated by QSARs
URXCOP	Mono(carboxyoctyl) phthalate (ng/mL)	5.176068	0.714	Estimated by QSARs
URXMIB	Mono-isobutyl pthalate (ng/mL)	1.75945	0.245377	Estimated by QSARs
URXMHP	Mono-(2-ethyl)-hexyl phthalate (ng/mL)	1.117285	0.048164	Estimated by QSARs
URXMOH	Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)	1.117285	0.048164	Estimated by QSARs
URXMZP	Mono-benzyl phthalate (ng/mL)	1.093676	0.038889	Estimated by QSARs
URXMHH	Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)	1.022749	0.009769	Estimated by QSARs
URXECP	Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)	0.990511	-0.00414	Estimated by QSARs
URXMBP	Mono-n-butyl phthalate (ng/mL)	0.951643	-0.02153	Estimated by QSARs
URXMEP	Mono-ethyl phthalate (ng/mL)	0.951643	-0.02153	Estimated by QSARs
URXMNM	Mono-n-methyl phthalate (ng/mL)	0.951643	-0.02153	Estimated by QSARs
SSURMHBP	Mono-3-hydroxy-n-butyl phthalate (ng/mL)	0.866928	-0.06202	Estimated by QSARs
URXMC1	Mono-(3-carboxypropyl) phthalate (ng/mL)	0.837084	-0.07723	Estimated by QSARs
URXDMA	o-Desmethylangolensin (O-DMA) (ng/mL)	3.116994	0.493736	Estimated by QSARs

URXETL	Enterolactone (ng/mL)	0.820352	-0.086	Estimated by QSARs
URXETD	Enterodiol (ng/mL)	0.412098	-0.385	Estimated by QSARs
URXP05	3-phenanthrene (ng/L)	10.13458	1.005806	Estimated by QSARs
URXP06	1-phenanthrene (ng/L)	10.13458	1.005806	Estimated by QSARs
URXP07	2-phenanthrene (ng/L)	10.13458	1.005806	Estimated by QSARs
URXP19	4-phenanthrene (ng/L)	10.13458	1.005806	Estimated by QSARs
URXP25	2 & 3-Hydroxyphenanthrene (ng/L)	10.13458	1.005806	Estimated by QSARs
URXP10	1-pyrene (ng/L)	8.830799	0.946	Estimated by QSARs
URXP01	1-napthol (ng/L)	6.424005	0.807806	Estimated by QSARs
URXP02	2-napthol (ng/L)	6.424005	0.807806	Estimated by QSARs
URXP17	9-fluorene (ng/L)	5.480988	0.738859	Estimated by QSARs
URXP03	3-fluorene (ng/L)	1.689685	0.227806	Estimated by QSARs
URXP04	2-fluorene (ng/L)	1.689685	0.227806	Estimated by QSARs
LBD199LA	PCB199 Lipid Adj (ng/g)	2023998	6.30621	Arnot et al., 2014 Training Set
LBX180LA	PCB180 Lipid Adj (ng/g)	451554.3	5.65471	Arnot et al., 2014 Training Set
LBX209LA	PCB209 Lipid Adj (ng/g)	303598.5	5.4823	Arnot et al., 2014 Training Set
LBX146LA	PCB146 Lipid Adj (ng/g)	233539.3	5.36836	Arnot et al., 2014 Training Set
LBX170LA	PCB170 Lipid Adj (ng/g)	226996.9	5.35602	Arnot et al., 2014 Training Set
LBX194LA	PCB194 Lipid Adj (ng/g)	170215.9	5.231	Estimated by QSARs
LBX187LA	PCB187 Lipid Adj (ng/g)	136442.9	5.134951	Arnot et al., 2014 Training Set
LBX153LA	PCB153 Lipid Adj (ng/g)	129038.7	5.11072	Arnot et al., 2014 Training Set
LBX196LA	PCB196 Lipid Adj (ng/g)	101199.9	5.00518	Arnot et al., 2014 Training Set
LBX138LA	PCB138 Lipid Adj (ng/g)	84739.08	4.928084	Arnot et al., 2014 Training Set

LBXHXCLA	3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	73872.02	4.86848	Arnot et al., 2014 Training Set
LBX118LA	PCB118 Lipid Adj (ng/g)	45459.01	4.65762	Arnot et al., 2014 Training Set
LBX099LA	PCB99 Lipid Adj (ng/g)	30359.85	4.4823	Arnot et al., 2014 Training Set
LBXPCBLA	3,3',4,4',5-pcnb Lipid Adj (pg/g)	14646.56	4.165736	Arnot et al., 2014 Training Set
LBX074LA	PCB74 Lipid Adj (ng/g)	8552.045	3.93207	Arnot et al., 2014 Training Set
LBX028LA	PCB28 Lipid Adj (ng/g)	5284.107	3.722972	Arnot et al., 2014 Training Set
LBX049LA	PCB49 Lipid Adj (ng/g)	1046.887	3.0199	Arnot et al., 2014 Training Set
LBX044LA	PCB44 Lipid Adj (ng/g)	948.7452	2.97715	Arnot et al., 2014 Training Set
LBXPFDE	Perfluorodecanoic acid (ng/mL)	58692	4.768578909	Conglomerated Estimation
LBXMPAH	2-(N-methyl-PFOSA) acetate (ng/mL)	14582.35905	4.163827787	Animal-Human Extrapolation
LBXPFNA	Perfluorononanoic acid (ng/mL)	21900	4.340444115	Conglomerated Estimation
LBXPFHS	Perfluorohexane sulfonic acid (ng/mL)	63948	4.805826966	Conglomerated Estimation
LBXPFOA	Perfluorooctanoic acid (ng/mL)	20586	4.313571968	Conglomerated Estimation
LBXPFOS	Perfluorooctane sulfonic acid (ng/mL)	41172	4.614601964	Conglomerated Estimation
LBXCOT	Cotinine (ng/mL)	17.30103	1.238072	Arnot et al., 2014 Training Set
URXNAL	NNAL , urine (ng/mL)	2.203265	0.343067	Estimated by QSARs
LBXVXY	Blood m-/p-Xylene (ng/ml)	31.62105	1.499976	Arnot et al., 2014 Training Set
LBXVDB	Blood 1,4-Dichlorobenzene (ng/ml)	21.9786	1.342	Estimated by QSARs
LBXVBM	Blood Bromodichloromethane (pg/ml)	18.40772	1.265	Estimated by QSARs
URXHEM	N-Ace-S-(2-Hydroxyethyl)-L-cys (ng/mL)	7.353935	0.86652	Estimated by QSARs
URXPHG	Phenylglyoxylic acid (ng/mL)	4.400217	0.643474	Estimated by QSARs
URXHP2	N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)	4.954844	0.69503	Estimated by QSARs

URXMB3	N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)	4.477442	0.65103	Estimated by QSARs
URX34M	3-methipurc acd & 4-methipurc acd (ng/mL)	4.824392	0.683443	Estimated by QSARs
URXBMA	N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)	4.594688	0.662256	Estimated by QSARs
URXBPM	N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)	4.477442	0.65103	Estimated by QSARs
URXGAM	N-ac-S-(2-carbmo-2-hydxel)-L-cys (ng/mL)	4.477442	0.65103	Estimated by QSARs
URXMAD	Mandelic acid (ng/mL)	4.005215	0.602626	Estimated by QSARs
URXAAM	N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)	3.992602	0.601256	Estimated by QSARs
URXAMC	N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)	3.992602	0.601256	Estimated by QSARs
URXATC	2-amnothiazolne-4-carbxylic acid (ng/mL)	3.992602	0.601256	Estimated by QSARs
URXCYM	N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)	3.992602	0.601256	Estimated by QSARs
URXTTC	2-thoxothazlidne-4-carbxylic acid (ng/mL)	3.992602	0.601256	Estimated by QSARs
URX2MH	2-Methylhippuric acid (ng/mL)	3.867926	0.587478	Estimated by QSARs
URXPMM	N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)	3.615931	0.55822	Estimated by QSARs
URXCEM	N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)	3.483236	0.541983	Estimated by QSARs
URXHPM	N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)	3.483236	0.541983	Estimated by QSARs
URXDHB	N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)	3.213711	0.507007	Estimated by QSARs
LBXVTO	Blood Toluene (ng/ml)	2.576321	0.411	Estimated by QSARs
LBXNM	Blood Nitromethane (pg/mL)	2.238721	0.35	Estimated by QSARs
LBXVCF	Blood Chloroform (pg/ml)	1.5	0.176091	Arnot et al., 2014 Training Set

Table A1.6. References on regulation, legislation, and restriction dates of substances.

Chemical	References for Ban/Phase-out Dates
Acephate	US EPA. 2017. Food and pesticides. Available: https://www.epa.gov/safepestcontrol/food-and-pesticides.
Aldrin/Endrin	ATSDR. 2011. Toxic substances portal - aldrin/dieldrin. Available: https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=56. US EPA. 2006. Procedures for the derivation of equilibrium partitioning sediment benchmarks (esbs) for the protection of benthic organisms: Endrin. Available: https://nepis.epa.gov/Exe/ZyNET.exe/P100G7G6.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2011+Thru+2015&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=.
Azinphos-Methyl	Gilbert S. 2014. Azinphos-methyl. Available: http://www.toxipedia.org/display/toxipedia/Azinphos-Methyl.
Bisphenol A (BPA)	Houlihan J, Lunder S, Jacob A. 2008. Timeline: Bpa from invention to phase-out. Available: https://www.ewg.org/research/timeline-bpa-invention-phase-out#.WjsfefmnEdU.
Cadmium	Spencer J. 2008. Toys 'R' us, mattel phase out cadmium batteries. The Wall Street Journal.
Carbofuran	Foley S. 2009. Carbofuran. Available: http://www.toxipedia.org/display/toxipedia/Carbofuran. US EPA. 2011. Carbofuran cancellation process. Available: https://archive.epa.gov/pesticides/reregistration/web/html/carbofuran_noic.html#revocation.
Carbon Tetrachloride	ATSDR. 2015. Toxic substances portal - carbon tetrachloride. Available: https://www.atsdr.cdc.gov/phs/phs.asp?id=194&tid=35.
Chlorobenzene	ATSDR. 2011. Toxic substances portal - chlorobenzene. Available: https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=87.
Chlordane	ATSDR. 2015. Toxic substances portal - chlordane. Available: https://www.atsdr.cdc.gov/phs/phs.asp?id=353&tid=62.
Chlorpyrifos	US EPA. 2017. Ingredients used in pesticide products: Chlorpyrifos. Available: https://www.epa.gov/ingredients-used-pesticide-products/chlorpyrifos.
Chlorpyrifos-Methyl	US EPA. 2000. Chlorpyrifos-methyl facts. Available: https://www3.epa.gov/pesticides/chem_search/reg_actions/reregistration/fs_PC-059102_1-Oct-00.pdf.
Cyanide	Jensen J. Ban on cyanide mining in montana with initiative 137. Available: http://meic.org/issues/mining-in-montana/hardrock-and-cyanide-mining-in-montana/ban-on-cyanide-mining-in-montana-with-initiative-137/. Representatives of Wisconsin. 2001. 2001 Assembly Bill 95. Available: http://docs.legis.wisconsin.gov/2001/related/proposals/ab95.
Diazinon	Gilbert S. 2014. Diazinon. Available: http://www.toxipedia.org/display/toxipedia/Diazinon.
Disulfoton and Methamidophos	US EPA. 2009. Disulfoton and methamidophos; product cancellation order. Available: https://www3.epa.gov/pesticides/chem_search/reg_actions/reregistration/frn_UG-2_23-Sep-2009.pdf.
DDT (dichloro-diphenyl-trichloroethane)	US EPA. 1975. DDT regulatory history: A brief survey (to 1975). Available: https://archive.epa.gov/epa/aboutepa/ddt-regulatory-history-brief-survey-1975.html.
Dioxins and dioxin-like compounds	US EPA. 1995. Municipal waste combustors.

1,2-Dibromo-3-Chloropropane	US EPA. 2000. 1,2-dibromo-3-chloropropane (DBCP). Available: https://www.epa.gov/sites/production/files/2016-09/documents/1-2-dibromo-3-chloropropane.pdf.
1,2-Dibromoethane	ATSDR. 2011. Toxic substances portal - 1,2-dibromoethane. Available: https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=131.
Dichlorvos (DDVP)	US EPA. 1995. Dichlorvos (ddvp); deletion of certain uses and directions. Available: https://www.federalregister.gov/documents/1995/04/19/95-9166/dichlorvos-ddvp-deletion-of-certain-uses-and-directions.
Ethyl Chloride	US EPA. Ethyl chloride (chloroethane). Available: https://www.epa.gov/sites/production/files/2016-09/documents/ethyl-chloride.pdf.
Ethylene Oxide	Gilbert S. 2009. Ethylene oxide (eto). Available: http://www.toxipedia.org/pages/viewpage.action?pageId=2822700.
Formaldehyde	Thomas K. 2014. The 'no more tears' shampoo, now with no formaldehyde. New York Times.
Heptachlor	ATSDR. 2015. Toxic substances portal - heptachlor/heptachlor epoxide. Available: https://www.atsdr.cdc.gov/phs/phs.asp?id=743&tid=135. Extension Toxicology Network. 1996. Pesticide information profile - heptachlor. Available: http://extoxnet.orst.edu/pips/heptachl.htm. Gilbert S. 2014. Heptachlor. Available: http://www.toxipedia.org/display/toxipedia/Heptachlor#Heptachlor-ATSDRPublicHealthStatement.
Hexachlorobenzene	ATSDR. 2015. Toxicological profile for hexachlorobenzene. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp90.pdf.
Hexachlorocyclohexanes	ASTDR. 2005. Toxicological profile for alpha-, beta-, gamma-, and delta-hexachlorocyclohexane. Available: https://www.atsdr.cdc.gov/toxprofiles/tp43.pdf. ATSDR. 2005. 5. Production, import/export, use, and disposal: Hexachlorocyclohexane. Available: https://www.atsdr.cdc.gov/ToxProfiles/tp43-c5.pdf.
Hexachloroethane	National Toxicology Program. 1994. Seventh annual report on carcinogens: Hexachloroethane. Available: https://ntp.niehs.nih.gov/ntp/roc/content/profiles/hexachloroethane.pdf. ATSDR. 2015. Toxic substances portal - hexachloroethane. Available: https://www.atsdr.cdc.gov/phs/phs.asp?id=868&tid=169.
Lead	National Center for Environmental Health Division of Emergency and Environmental Health Services. 2014. Lead: Prevention tips. Available: https://www.cdc.gov/nceh/lead/tips.htm. Newell R, Rogers K. 2003. The u.S. Experience with the phasedown of lead in gasoline.Resources for the Future. Fowler T. 2008. A brief history of lead regulation. Science Progress where science, technology, and policy meet. US EPA. 2017. Drinking water contaminants – standards and regulations. Available: https://www.epa.gov/dwstandardsregulations/use-lead-free-pipes-fittings-fixtures-solder-and-flux-drinking-water. US EPA. 2011. Basic questions and answers for the drinking water strategy contaminant groups effort. Available: https://www.epa.gov/dwstandardsregulations.
Mercury	US EPA. 1996. Mercury-containing and rechargeable battery management act of 1996. Available: https://www.congress.gov/104/plaws/publ142/PLAW-104publ142.pdf. Mohapatra SP, Nikolova I, Mitchell A. 2007. Managing mercury in the great lakes: An analytical review of abatement policies. Journal of environmental management 83:80-92.
Methyl parathion	Jaga K, Dharmani C. 2006. Methyl parathion: An organophosphate insecticide not quite forgotten. Reviews on environmental health 21:57-67.
Metiram	Schneider K. 1992. E.P.A., in a reversal, lifts a ban on farm chemicals. New York Times.

Mirex	OEHHA. 2016. Mirex. Available: https://oehha.ca.gov/chemicals/mirex.
Methyl Tertiary Butyl Ether (MTBE)	American Cancer Society. 2014. MTBE and cancer risk. Available: https://www.cancer.org/cancer/cancer-causes/mtbe.html.
Parabens	Franklin K. 2016. Johnson & Johnson meets ingredient commitments. Chemical Watch: Global Risk & Regulation News.
Polybrominated Biphenyls (PBBs)	ATSDR. 2015. Toxic substances portal - polybrominated biphenyls (PBBs). Available: https://www.atsdr.cdc.gov/PHS/PHS.asp?id=527&tid=94.
Polybrominated diphenyl ether (PBDE)	Guardia MJL, Hale RC, Harvey E. 2006. Detailed polybrominated diphenyl ether (pbde) congener composition of the widely used penta-, octa-, and deca-pbde technical flame-retardant mixtures. Environmental science & technology 40:6247-6254. US EPA. 2010. An exposure assessment of polybrominated diphenyl ethers (pbde) (final). Available: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=210404.
Polychlorinated biphenyls (PCBs)	US EPA. 1979. EPA bans PCB manufacture; phases out uses. Available: https://archive.epa.gov/epa/aboutepa/epa-bans-pcb-manufacture-phases-out-uses.html.
Perfluorinated compounds, including Perfluoroalkyl and Polyfluoroalkyl Substances (PFASs)	US EPA. 2016. Risk management for per- and polyfluoroalkyl substances (PFASs) under tsca. Available: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-management-and-polyfluoroalkyl-substances-pfass. Buck RC, Franklin J, Berger U, Conder JM, Cousins IT, de Voogt P, et al. 2011. Perfluoroalkyl and polyfluoroalkyl substances in the environment: Terminology, classification, and origins. Integrated environmental assessment and management 7:513-541. Fromme H, Becher G, Hilger B, Völkel W. 2016. Brominated flame retardants – exposure and risk assessment for the general population. International Journal of Hygiene and Environmental Health 219:1-23. Butenhoff JL, Olsen GW, Pfahles-Hutchens A. 2006. The applicability of biomonitoring data for perfluorooctanesulfonate to the environmental public health continuum. Environmental Health Perspectives 114:1776-1782. US EPA. 2003. Perfluoroalkyl sulfonates; significant new use rule. US EPA. 2002. Perfluoroalkyl sulfonates; significant new use rule; final rule and supplemental proposed rule. Available: https://www.federalregister.gov/documents/2002/12/09/02-31011/perfluoroalkyl-sulfonates-significant-new-use-rule.
Phthalates	US Consumer Product Safety Commission. 2015. Phthalates. Available: https://www.cpsc.gov/Business--Manufacturing/Business-Education/Business-Guidance/Phthalates-Information. Hileman B. 2007. California bans phthalates in toys for children. Chemical & Engineering News 85:12. Szabo L. 2007. Hospitals move to phase out chemical. ABC News.
2,4,5-T	Institute of Medicine Committee to Review the Health Effects in Vietnam Veterans of Exposure to H. 1994. 2, history of the controversy over the use of herbicides. Available: https://www.ncbi.nlm.nih.gov/books/NBK236351/.
2,3,7,8-Tetrachlorodibenzo-p-Dioxin (2,3,7,8,-TCDD)	US EPA. 2000. 2,3,7,8-tetrachlorodibenzo-p-dioxin (2,3,7,8,-tcdd). Available: https://www.epa.gov/sites/production/files/2016-09/documents/2-3-7-8-tetrachlorodibenzo-p-dioxin.pdf.
Tetrachlorvinphos	US EPA. 2017. Tetrachlorvinphos (tcvp). Available: https://www.epa.gov/ingredients-used-pesticide-products/tetrachlorvinphos-tcvp.
Thallium	Staff of the Nonferrous Metals Division. 1972. Thallium. Available: http://digicoll.library.wisc.edu/cgi-bin/EcoNatRes/EcoNatRes-idx?type=goto&id=EcoNatRes.MinYB1972v1&page=1358&isize=XL.
Thimerosal	CDC. 2015. Thimerosal in vaccines. Available: https://www.cdc.gov/vaccinesafety/concerns/thimerosal/index.html.

Triclosan and triclocarbon	FDA. 2016. FDA issues final rule on safety and effectiveness of antibacterial soaps. Available: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm517478.htm.
Trichloroethylene	Gilbert S. 2014. Trichloroethylene. Available: http://www.toxipedia.org/display/toxipedia/Trichloroethylene.
Tungsten	Kiger PJ. 2013. U.S. Phase-out of incandescent light bulbs continues in 2014 with 40-, 60-watt varieties. National Geographic.
Styrene	United Press International. 1988. Berkeley widens ban on foam food containers. Los Angeles Times.

Table A1.7. Latest restriction date, decade, and period by chemicals.

NHANES Codename	Chemical Name (units)	Latest Restriction Date	Latest Restriction Decade	Latest Restriction Period
LBXGLY	Glycideamide (pmoL/G Hb)	#N/A	No Restrictions	No Restrictions
LBXACR	Acrylamide (pmoL/G Hb)	#N/A	No Restrictions	No Restrictions
LBXBB1LA	2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)	1979	1970s	1970-1984
LBXBR2LA	2,4,4'-tribromodiphenyl ether lipid adj (ng/g)	2004	2000s	2000-2014
LBXBR8LA	2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)	2004	2000s	2000-2014
LBXBR6LA	2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)	2004	2000s	2000-2014
LBXBR3LA	2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)	2004	2000s	2000-2014
LBXBR5LA	2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)	2004	2000s	2000-2014
LBXBR7LA	2,2',4,4',5,5'-hexabromodiphenyl lipid adj (ng/g)	2004	2000s	2000-2014
LBXD03LA	1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	1995	1990s	1985-1999
LBXD07LA	1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	1995	1990s	1985-1999
LBXD05LA	1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	1995	1990s	1985-1999
LBXF04LA	1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	1995	1990s	1985-1999
LBXF03LA	2,3,4,7,8-pncdf Lipid Adj (pg/g)	1995	1990s	1985-1999
LBXF08LA	1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	1995	1990s	1985-1999
SSMEL	Melamine (ng/mL)	#N/A	No Restrictions	No Restrictions
SSCYA	Cyanuric acid (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXBPB	Lead (ug/dL)	1995	1990s	1985-1999
LBXBCD	Cadmium (ug/L)	2008	2000s	2000-2014
LBXTHG	Mercury, total (ug/L)	1996	1990s	1985-1999
URXUHG	Mercury, urine (ng/mL)	1996	1990s	1985-1999
URXUBA	Barium, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUCO	Cobalt, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUCS	Cesium, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUMO	Molybdenum, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUPB	Lead, urine (ng/mL)	1995	1990s	1985-1999
URXUSB	Antimony, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUTL	Thallium, urine (ng/mL)	1972	1970s	1970-1984

URXUTU	Tungsten, urine (ng/mL)	2012	2010s	2000-2014
URXUUR	Uranium, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUAS	Urinary total Arsenic (µg/L)	#N/A	No Restrictions	No Restrictions
URXUAB	Urinary Arsenobetaine (µg/L)	#N/A	No Restrictions	No Restrictions
URXUDMA	Urinary Dimethylarsonic acid (µg/L)	#N/A	No Restrictions	No Restrictions
URXUCD	Cadmium, urine (ng/mL)	2008	2000s	2000-2014
LBXBMN	Blood manganese (ug/L)	#N/A	No Restrictions	No Restrictions
LBXSCU	Serum Copper (ug/dL)	#N/A	No Restrictions	No Restrictions
LBXSZN	Serum Zinc (ug/dL)	#N/A	No Restrictions	No Restrictions
LBXBGM	Mercury, methyl (ug/L)	1996	1990s	1985-1999
URXSCN	Urinary thiocyanate (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUIO	Iodine, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXNO3	Urinary nitrate (ng/mL)	#N/A	No Restrictions	No Restrictions
URXUP8	Perchlorate, urine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXTRS	Urinary Triclosan (ng/mL)	2016	No Restrictions	No Restrictions
URXBPS	Urinary Bisphenol S (ug/L)	#N/A	No Restrictions	No Restrictions
URXMPB	Methyl paraben (ng/ml)	2015	No Restrictions	No Restrictions
URXPPB	Propyl paraben (ng/ml)	2015	No Restrictions	No Restrictions
URXBPH	Urinary Bisphenol A (ng/mL)	2008	2000s	2000-2014
URXBPF	Urinary Bisphenol F (ug/L)	#N/A	No Restrictions	No Restrictions
URXBP3	Urinary Benzophenone-3 (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXTNALA	Trans-nonachlor Lipid Adj (ng/g)	1983	1980s	1970-1984
LBXOXYLA	Oxychlordane Lipid Adj (ng/g)	1983	1980s	1970-1984
LBXPDELA	p,p'-DDE Lipid Adj (ng/g)	1959	1950s	Before 1970s
LBXDIELA	Dieldrin Lipid Adj (ng/g)	1974	1970s	1970-1984
LBXBHCLA	Beta-hexachlorocyclohexane Lipid Adj (ng/g)	1976	1970s	1970-1984
URXCPM	3,5,6-trichloropyridinol (ug/L)	2000	2000s	2000-2014
URXOP3	Dimethylthiophosphate (ug/L)	2000	2000s	2000-2014
URXOPM	3-phenoxybenzoic acid (ug/L)	#N/A	No Restrictions	No Restrictions
URX14D	2,5-dichlorophenol (ug/L)	#N/A	No Restrictions	No Restrictions

URXDCB	2,4-dichlorophenol (ug/L)	#N/A	No Restrictions	No Restrictions
URXPAR	Paranitrophenol (ug/L)	1998	1990s	1985-1999
URX24D	2,4-D (ug/L)	#N/A	No Restrictions	No Restrictions
URXDEA	DEET acid (ug/L)	#N/A	No Restrictions	No Restrictions
URXCNP	Mono(carboxynonyl) phthalate (ng/mL)	2008	2000s	2000-2014
SSURHIBP	Mono-2-hydroxy-iso-butyl phthlte (ng/mL)	#N/A	No Restrictions	No Restrictions
URXCOP	Mono(carboxyoctyl) phthalate (ng/mL)	2008	2000s	2000-2014
URXMIB	Mono-isobutyl pthalate (ng/mL)	2009	2000s	2000-2014
URXMHP	Mono-(2-ethyl)-hexyl phthalate (ng/mL)	2007	2000s	2000-2014
URXMOH	Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)	2007	2000s	2000-2014
URXMZP	Mono-benzyl phthalate (ng/mL)	2009	2000s	2000-2014
URXMHH	Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)	2007	2000s	2000-2014
URXECP	Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)	2007	2000s	2000-2014
URXMBP	Mono-n-butyl phthalate (ng/mL)	2009	2000s	2000-2014
URXMEP	Mono-ethyl phthalate (ng/mL)	#N/A	No Restrictions	No Restrictions
URXMNM	Mono-n-methyl phthalate (ng/mL)	#N/A	No Restrictions	No Restrictions
SSURMHBP	Mono-3-hydroxy-n-butyl phthalate (ng/mL)	2009	2000s	2000-2014
URXMC1	Mono-(3-carboxypropyl) phthalate (ng/mL)	#N/A	No Restrictions	No Restrictions
URXDMA	o-Desmethylangolensin (O-DMA) (ng/mL)	#N/A	No Restrictions	No Restrictions
URXETL	Enterolactone (ng/mL)	#N/A	No Restrictions	No Restrictions
URXETD	Enterodiol (ng/mL)	#N/A	No Restrictions	No Restrictions
URXP05	3-phenanthrene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP06	1-phenanthrene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP07	2-phenanthrene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP19	4-phenanthrene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP25	2 & 3-Hydroxyphenanthrene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP10	1-pyrene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP01	1-napthol (ng/L)	#N/A	No Restrictions	No Restrictions
URXP02	2-napthol (ng/L)	#N/A	No Restrictions	No Restrictions
URXP17	9-fluorene (ng/L)	#N/A	No Restrictions	No Restrictions

URXP03	3-fluorene (ng/L)	#N/A	No Restrictions	No Restrictions
URXP04	2-fluorene (ng/L)	#N/A	No Restrictions	No Restrictions
LBD199LA	PCB199 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX180LA	PCB180 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX209LA	PCB209 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX146LA	PCB146 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX170LA	PCB170 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX194LA	PCB194 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX187LA	PCB187 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX153LA	PCB153 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX196LA	PCB196 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX138LA	PCB138 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBXHXCLA	3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	1979	1970s	1970-1984
LBX118LA	PCB118 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX099LA	PCB99 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBXPCBLA	3,3',4,4',5-pcnb Lipid Adj (pg/g)	1979	1970s	1970-1984
LBX074LA	PCB74 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX028LA	PCB28 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX049LA	PCB49 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBX044LA	PCB44 Lipid Adj (ng/g)	1979	1970s	1970-1984
LBXPFDE	Perfluorodecanoic acid (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXMPAH	2-(N-methyl-PFOSA) acetate (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXPFNA	Perfluorononanoic acid (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXPFHS	Perfluorohexane sulfonic acid (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXPFOA	Perfluorooctanoic acid (ng/mL)	2002	2000s	2000-2014
LBXPFOS	Perfluorooctane sulfonic acid (ng/mL)	2002	2000s	2000-2014
LBXCOT	Cotinine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXNAL	NNAL , urine (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXVXY	Blood m-/p-Xylene (ng/ml)	#N/A	No Restrictions	No Restrictions
LBXVDB	Blood 1,4-Dichlorobenzene (ng/ml)	1987	1980s	1985-1999
LBXVBM	Blood Bromodichloromethane (pg/ml)	#N/A	No Restrictions	No Restrictions
URXHEM	N-Ace-S-(2-Hydroxyethyl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXPHG	Phenylglyoxylic acid (ng/mL)	2016	No Restrictions	No Restrictions
URXHP2	N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions

URXMB3	N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URX34M	3-methipurc acd & 4-methipurc acd (ng/mL)	#N/A	No Restrictions	No Restrictions
URXBMA	N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXBPM	N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)	#N/A	No Restrictions	No Restrictions
URXGAM	N-ac-S-(2-carbmo-2-hydxel)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXMAD	Mandelic acid (ng/mL)	2016	No Restrictions	No Restrictions
URXAAM	N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXAMC	N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXATC	2-amnothiazolne-4-carbxylic acid (ng/mL)	#N/A	No Restrictions	No Restrictions
URXCYM	N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXTTC	2-thoxothazlidne-4-carbxylic acid (ng/mL)	#N/A	No Restrictions	No Restrictions
URX2MH	2-Methylhippuric acid (ng/mL)	#N/A	No Restrictions	No Restrictions
URXPMM	N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXCEM	N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXHPM	N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)	#N/A	No Restrictions	No Restrictions
URXDHB	N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)	#N/A	No Restrictions	No Restrictions
LBXVTO	Blood Toluene (ng/ml)	#N/A	No Restrictions	No Restrictions
LBXNM	Blood Nitromethane (pg/mL)	#N/A	No Restrictions	No Restrictions
LBXVCF	Blood Chloroform (pg/ml)	#N/A	No Restrictions	No Restrictions

Table A1.8. $\beta_{age}, \beta_{age^2}$, calculated $s_{children}$, and calculated fold difference when $X_{children}$ = 5 years and $(\text{X\_age})^{-}$ = 31.88 years for all chemical biomarkers, which are ranked by the fold difference of chemical biomarker levels between a child of 5 years and an adult of 31.88 years ($10^{s_{children}}$) in descending order.

Codename	Chemical Name	β_{age}	β_{age^2}	Schildren	Fold Difference ($10^{s_{children}}$)
URXMZP	Mono-benzyl phthalate (ng/mL)	-9.67E-03	2.14E-04	0.415	2.598
URXDMA	o-Desmethylangolensin (O-DMA) (ng/mL)	-7.51E-03	1.88E-04	0.338	2.176
URXMC1	Mono-(3-carboxypropyl) phthalate (ng/mL)	-7.14E-03	2.01E-04	0.337	2.173
URXATC	2-amnothiazolne-4-carbxylic acid (ng/mL)	-6.48E-03	1.90E-04	0.312	2.049
URXUTU	Tungsten, urine (ng/mL)	-6.89E-03	1.48E-04	0.292	1.960
URXCPM	3,5,6-trichloropyridinol (ug/L)	-4.71E-03	2.24E-04	0.288	1.942
SSURHIBP	Mono-2-hydroxy-iso-butyl phthlte (ng/mL)	-6.28E-03	1.61E-04	0.285	1.927
URXMBP	Mono-n-butyl phthalate (ng/mL)	-5.43E-03	1.79E-04	0.275	1.884
LBXBR7LA	2,2',4,4',5,5'-hexabromodiphenyl ether lipid adj (ng/g)	-6.59E-03	1.21E-04	0.264	1.839
URXDEA	DEET acid (ug/L)	-6.42E-03	1.20E-04	0.260	1.818
URXMOH	Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)	-5.64E-03	1.44E-04	0.256	1.803
URXECP	Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)	-5.21E-03	1.46E-04	0.245	1.760
SSURMHBP	Mono-3-hydroxy-n-butyl phthalate (ng/mL)	-4.18E-03	1.71E-04	0.236	1.721
URXUMO	Molybdenum, urine (ng/mL)	-4.68E-03	1.41E-04	0.228	1.689
URXMHH	Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)	-4.89E-03	1.19E-04	0.217	1.650
URXPAR	Paranitrophenol (ug/L)	-3.35E-03	1.70E-04	0.213	1.632
LBXBR5LA	2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)	-4.92E-03	1.05E-04	0.208	1.614
URXUIO	Iodine, urine (ng/mL)	-2.79E-03	1.76E-04	0.202	1.594
URXOP3	Dimethylthiophosphate (ug/L)	-2.70E-03	1.76E-04	0.200	1.584
LBXBR3LA	2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)	-4.67E-03	9.96E-05	0.197	1.576
URXCOP	Mono(carboxyoctyl) phthalate (ng/mL)	-5.01E-03	8.47E-05	0.196	1.570
URXMNM	Mono-n-methyl phthalate (ng/mL)	-3.65E-03	1.32E-04	0.194	1.562
URXMIB	Mono-isobutyl pthalate (ng/mL)	-4.58E-03	9.55E-05	0.192	1.556
LBXPFHS	Perfluorohexane sulfonic acid (ng/mL)	-3.01E-03	1.40E-04	0.182	1.520
LBXMPAH	2-(N-methyl-PFOSA) acetate (ng/mL)	-2.65E-03	1.49E-04	0.179	1.510
URXCNP	Mono(carboxynonyl) phthalate (ng/mL)	-3.93E-03	9.99E-05	0.178	1.506
URXUCO	Cobalt, urine (ng/mL)	-3.64E-03	1.02E-04	0.172	1.485
LBXBR6LA	2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)	-4.17E-03	7.85E-05	0.169	1.475
URXHEM	N-Ace-S-(2-Hydroxyethyl)-L-cys (ng/mL)	-4.59E-03	5.97E-05	0.166	1.467
URXUP8	Perchlorate, urine (ng/mL)	-2.74E-03	1.09E-04	0.153	1.421
URXETL	Enterolactone (ng/mL)	-1.80E-03	1.39E-04	0.149	1.409
URXP10	1-pyrene (ng/L)	-4.12E-03	2.88E-05	0.131	1.354

LBXBR8LA	2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)	-2.94E-03	7.24E-05	0.131	1.353
LBX049LA	PCB49 Lipid Adj (ng/g)	-3.12E-03	6.07E-05	0.128	1.342
URXBPH	Urinary Bisphenol A (ng/mL)	-2.88E-03	6.60E-05	0.125	1.334
URXTTC	2-thoxothazlidne-4-carbxylic acid (ng/mL)	-2.31E-03	8.27E-05	0.122	1.324
LBX044LA	PCB44 Lipid Adj (ng/g)	-2.84E-03	5.97E-05	0.119	1.317
URXUBA	Barium, urine (ng/mL)	-3.42E-03	3.60E-05	0.118	1.312
URXUSB	Antimony, urine (ng/mL)	-2.67E-03	6.28E-05	0.117	1.310
URXDHB	N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)	-1.27E-03	1.12E-04	0.115	1.304
URXNAL	NNAL , urine (ng/mL)	-2.24E-03	7.03E-05	0.111	1.291
LBXF08LA	1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	-2.15E-03	7.07E-05	0.109	1.284
URX14D	2,5-dichlorophenol (ug/L)	-4.78E-04	1.31E-04	0.107	1.281
URXOPM	3-phenoxybenzoic acid (ug/L)	-1.74E-03	6.93E-05	0.097	1.250
URXDCB	2,4-dichlorophenol (ug/L)	-7.03E-04	1.04E-04	0.094	1.242
URXNO3	Urinary nitrate (ng/mL)	-2.66E-03	2.96E-05	0.093	1.238
URX24D	2,4-D (ug/L)	-1.22E-03	7.55E-05	0.087	1.222
LBX028LA	PCB28 Lipid Adj (ng/g)	-1.10E-03	7.63E-05	0.085	1.215
LBXGLY	Glycideamide (pmoL/G Hb)	-2.61E-03	1.53E-05	0.081	1.206
URXUUR	Uranium, urine (ng/mL)	-9.01E-04	7.21E-05	0.076	1.192
URXUTL	Thallium, urine (ng/mL)	-1.83E-03	2.40E-05	0.067	1.166
URXBMA	N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)	-3.74E-04	7.75E-05	0.066	1.164
URXP05	3-phenanthrene (ng/L)	-9.75E-04	5.23E-05	0.064	1.159
LBXVDB	Blood 1,4-Dichlorobenzene (ng/ml)	-3.98E-04	4.82E-05	0.046	1.111
LBXBR2LA	2,4,4'-tribromodiphenyl ether lipid adj (ng/g)	-4.72E-04	4.37E-05	0.044	1.107
URXUCS	Cesium, urine (ng/mL)	-3.65E-04	4.58E-05	0.043	1.104
URXAAM	N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)	-1.08E-03	1.74E-05	0.042	1.100
URXMHP	Mono-(2-ethyl)-hexyl phthalate (ng/mL)	-2.18E-03	-2.45E-05	0.041	1.099
URXETD	Enterodiol (ng/mL)	-3.72E-04	4.14E-05	0.040	1.096
URXUPB	Lead, urine (ng/mL)	1.51E-03	1.10E-04	0.039	1.095
LBXPFOA	Perfluorooctanoic acid (ng/mL)	-5.80E-05	2.97E-05	0.023	1.054
URX34M	3-methipurc acd & 4-methipurc acd (ng/mL)	7.96E-04	5.81E-05	0.021	1.049
URXBPF	Urinary Bisphenol F (ug/L)	-1.08E-04	2.16E-05	0.019	1.044
SSCYA	Cyanuric acid (ng/mL)	8.50E-04	5.46E-05	0.017	1.039
LBXBMN	Blood manganese (ug/L)	-8.32E-04	-9.03E-06	0.016	1.037
LBXVCF	Blood Chloroform (pg/ml)	-3.96E-04	5.94E-06	0.015	1.035
URXPHG	Phenylglyoxylic acid (ng/mL)	5.81E-04	4.01E-05	0.013	1.031
LBXACR	Acrylamide (pmoL/G Hb)	-7.18E-04	-1.19E-05	0.011	1.025
LBXSCU	Serum Copper (ug/dL)	1.19E-04	7.37E-06	0.002	1.005
LBXSZN	Serum Zinc (ug/dL)	-7.65E-05	-1.16E-06	0.001	1.003
URXP03	3-fluorene (ng/L)	-1.19E-04	-1.11E-05	-0.005	0.989
URXSCN	Urinary thiocyanate (ng/mL)	-1.09E-03	-6.86E-05	-0.020	0.955
URXP19	4-phenanthrene (ng/L)	5.15E-04	-1.63E-05	-0.026	0.943

URXUDMA	Urinary Dimethylarsonic acid (µg/L)	1.26E-03	8.43E-06	-0.028	0.938
URXP06	1-phenanthrene (ng/L)	1.11E-03	-7.23E-06	-0.035	0.922
LBXBPB	Lead (ug/dL)	3.89E-03	8.72E-05	-0.042	0.909
URXCEM	N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)	3.01E-03	5.26E-05	-0.043	0.906
SSMEL	Melamine (ng/mL)	3.01E-03	3.89E-05	-0.053	0.886
LBXNM	Blood Nitromethane (pg/mL)	1.93E-03	-2.03E-06	-0.053	0.885
LBXPFOS	Perfluorooctane sulfonic acid (ng/mL)	2.89E-03	3.27E-05	-0.054	0.883
URXP04	2-fluorene (ng/L)	1.40E-03	-2.36E-05	-0.055	0.882
URXBPS	Urinary Bisphenol S (ug/L)	2.08E-03	-2.58E-06	-0.058	0.875
URXPMM	N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)	2.85E-03	2.23E-05	-0.060	0.870
URX2MH	2-Methylhippuric acid (ng/mL)	1.60E-03	-2.60E-05	-0.062	0.867
LBXVXY	Blood m-/p-Xylene (ng/ml)	1.52E-03	-3.36E-05	-0.065	0.861
LBXVBM	Blood Bromodichloromethane (pg/ml)	1.23E-03	-4.78E-05	-0.068	0.856
URXP02	2-napthol (ng/L)	1.54E-03	-4.08E-05	-0.071	0.849
LBXPFNA	Perfluorononanoic acid (ng/mL)	2.27E-03	-2.83E-05	-0.081	0.829
URXP01	1-napthol (ng/L)	4.09E-03	3.82E-05	-0.082	0.827
URXP25	2 & 3-Hydroxyphenanthrene (ng/L)	1.53E-03	-5.82E-05	-0.083	0.826
URXBP3	Urinary Benzophenone-3 (ng/mL)	-6.64E-04	-1.50E-04	-0.091	0.811
LBXVTO	Blood Toluene (ng/ml)	2.57E-03	-3.24E-05	-0.092	0.808
URXP17	9-fluorene (ng/L)	2.86E-03	-2.95E-05	-0.098	0.797
URXUAS	Urinary total Arsenic (µg/L)	3.37E-03	-1.06E-05	-0.098	0.797
URXP07	2-phenanthrene (ng/L)	2.59E-03	-5.18E-05	-0.107	0.782
LBXPFDE	Perfluorodecanoic acid (ng/mL)	3.45E-03	-5.77E-05	-0.135	0.734
LBXF04LA	1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	5.84E-03	2.91E-05	-0.136	0.731
URXGAM	N-ac-S-(2-carbmo-2-hydxel)-L-cys (ng/mL)	3.27E-03	-7.50E-05	-0.142	0.721
URXCYM	N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)	3.68E-03	-7.80E-05	-0.155	0.700
URXTRS	Urinary Triclosan (ng/mL)	2.96E-03	-1.16E-04	-0.163	0.686
URXMAD	Mandelic acid (ng/mL)	4.00E-03	-8.60E-05	-0.170	0.676
URXMPB	Methyl paraben (ng/ml)	4.50E-03	-7.57E-05	-0.176	0.667
URXHPM	N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)	4.78E-03	-7.39E-05	-0.182	0.658
LBXDIELA	Dieldrin Lipid Adj (ng/g)	6.70E-03	-1.99E-05	-0.195	0.639
URXMEP	Mono-ethyl phthalate (ng/mL)	4.95E-03	-8.60E-05	-0.195	0.638
LBXF03LA	2,3,4,7,8-pncdf Lipid Adj (pg/g)	8.48E-03	4.00E-05	-0.199	0.632
URXUHG	Mercury, urine (ng/mL)	4.61E-03	-1.09E-04	-0.202	0.627
URXAMC	N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)	6.61E-03	-6.21E-05	-0.222	0.599
LBXBCD	Cadmium (ug/L)	7.31E-03	-4.97E-05	-0.232	0.586
URXPPB	Propyl paraben (ng/ml)	4.92E-03	-1.41E-04	-0.234	0.584
URXBPM	N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)	5.23E-03	-1.63E-04	-0.259	0.551
LBX099LA	PCB99 Lipid Adj (ng/g)	1.00E-02	1.22E-05	-0.260	0.550
URXMB3	N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)	7.49E-03	-8.33E-05	-0.262	0.548
LBXPCBLA	3,3',4,4',5-pcnb Lipid Adj (pg/g)	1.01E-02	-4.80E-06	-0.276	0.530

URXUAB	Urinary Arsenobetaine (µg/L)	7.96E-03	-9.59E-05	-0.283	0.521
LBX118LA	PCB118 Lipid Adj (ng/g)	1.19E-02	4.10E-05	-0.290	0.513
LBXD05LA	1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	9.44E-03	-5.05E-05	-0.290	0.513
LBX074LA	PCB74 Lipid Adj (ng/g)	1.25E-02	6.26E-05	-0.290	0.512
URXHP2	N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)	7.77E-03	-1.52E-04	-0.318	0.480
LBXTHG	Mercury, total (ug/L)	8.76E-03	-1.83E-04	-0.368	0.429
LBXD07LA	1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	1.19E-02	-6.70E-05	-0.369	0.428
LBXOXYLA	Oxychlordane Lipid Adj (ng/g)	1.41E-02	1.02E-05	-0.370	0.426
LBXHXCLA	3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	1.40E-02	3.19E-06	-0.373	0.423
LBX209LA	PCB209 Lipid Adj (ng/g)	1.80E-02	1.16E-04	-0.400	0.398
LBXBHCLA	Beta-hexachlorocyclohexane Lipid Adj (ng/g)	1.53E-02	8.75E-06	-0.405	0.394
URXUCD	Cadmium, urine (ng/mL)	1.44E-02	-9.23E-05	-0.452	0.353
LBXBGM	Mercury, methyl (ug/L)	1.10E-02	-2.23E-04	-0.457	0.349
LBXD03LA	1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	1.62E-02	-5.10E-05	-0.471	0.338
LBXTNALA	Trans-nonachlor Lipid Adj (ng/g)	1.75E-02	-5.37E-05	-0.508	0.310
LBXPDELA	p,p'-DDE Lipid Adj (ng/g)	1.84E-02	-1.19E-04	-0.580	0.263
LBX138LA	PCB138 Lipid Adj (ng/g)	1.98E-02	-1.11E-04	-0.613	0.244
LBX153LA	PCB153 Lipid Adj (ng/g)	2.15E-02	-1.28E-04	-0.670	0.214
LBX146LA	PCB146 Lipid Adj (ng/g)	2.26E-02	-1.31E-04	-0.703	0.198
LBXBB1LA	2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)	2.10E-02	-2.71E-04	-0.761	0.174
LBX187LA	PCB187 Lipid Adj (ng/g)	2.60E-02	-1.88E-04	-0.836	0.146
LBX180LA	PCB180 Lipid Adj (ng/g)	2.79E-02	-2.21E-04	-0.908	0.124
LBX170LA	PCB170 Lipid Adj (ng/g)	2.84E-02	-2.52E-04	-0.946	0.113
LBX196LA	PCB196 Lipid Adj (ng/g)	3.07E-02	-2.68E-04	-1.020	0.096
LBXCOT	Cotinine (ng/mL)	1.61E-02	-8.17E-04	-1.024	0.095
LBD199LA	PCB199 Lipid Adj (ng/g)	3.34E-02	-2.61E-04	-1.086	0.082
LBX194LA	PCB194 Lipid Adj (ng/g)	3.67E-02	-3.34E-04	-1.228	0.059

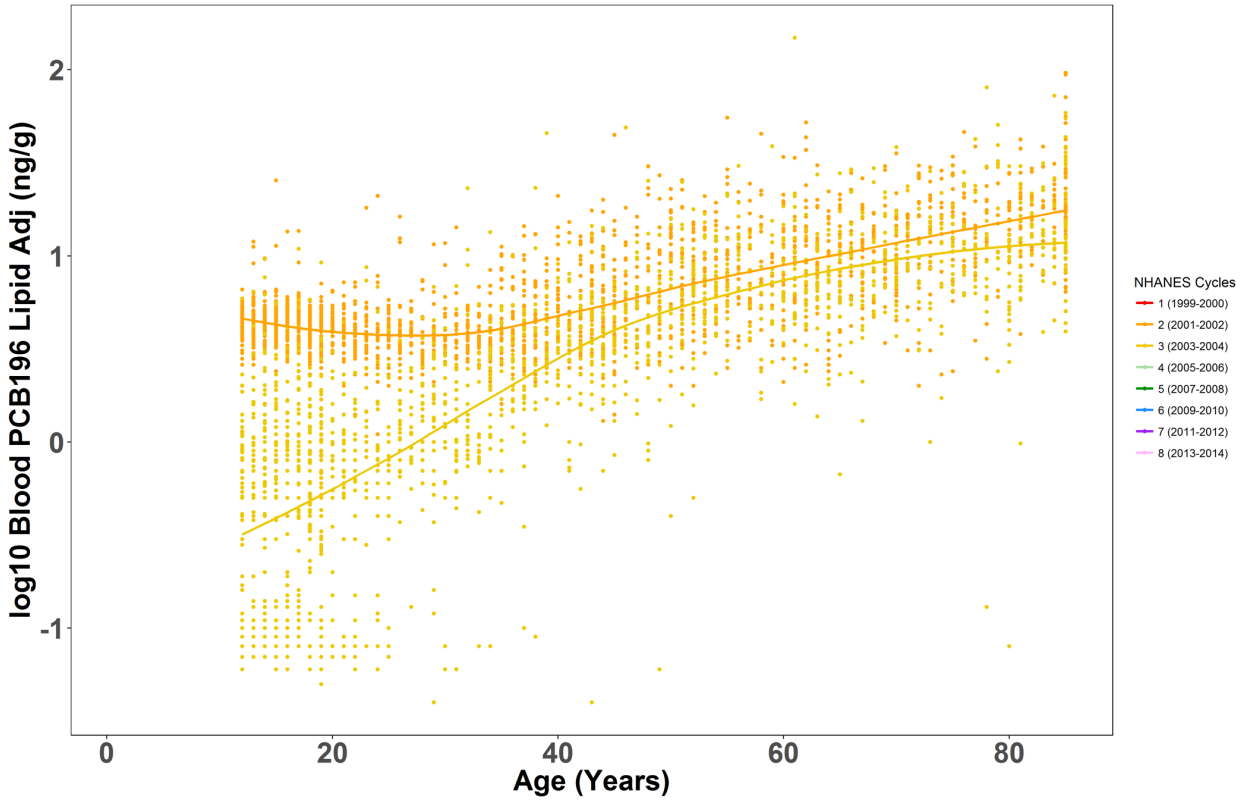

Figure A1.1. PCB 196 concentrations across the life-stages stratified by NHANEs cycles for Cycle 2 and 3.

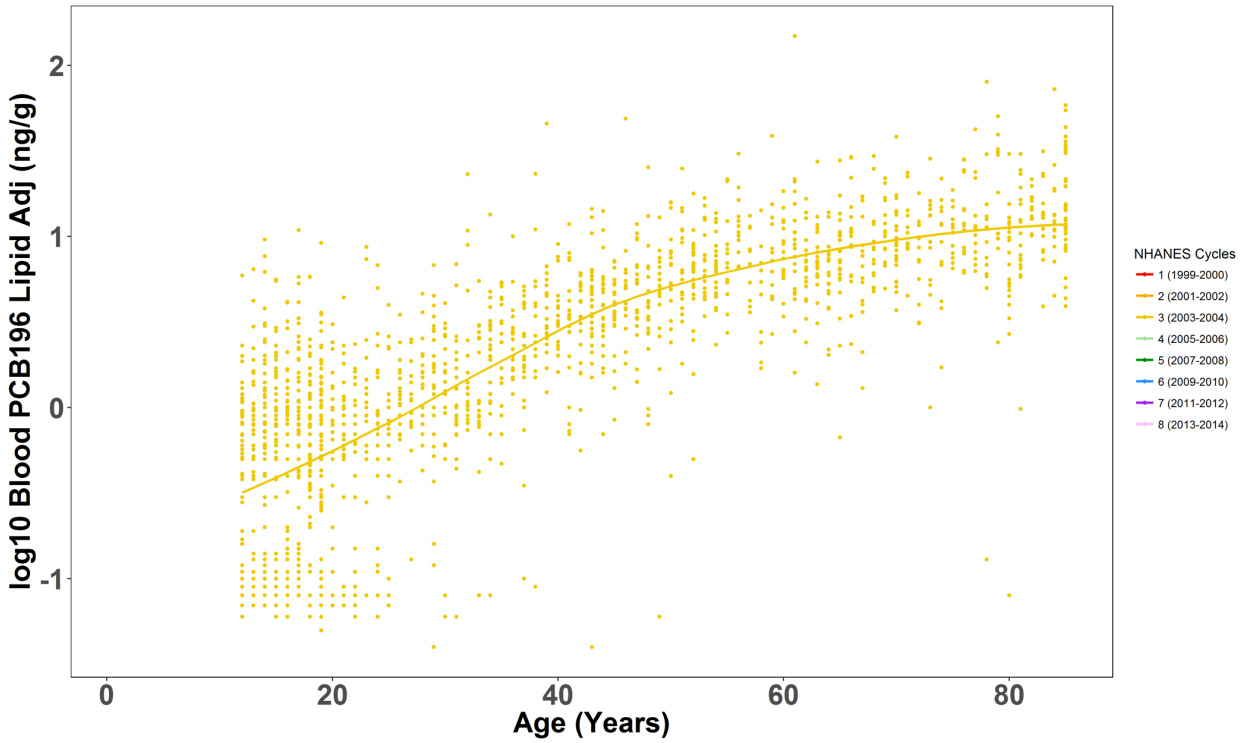

Figure A1.2. PCB 196 concentrations across the life-stages stratified for only Cycle 3.

Text A1.2. Discussion of regression coefficients for NHANES cycles.

Figure A1.3 summarize the ranges of average log change in chemical biomarker levels over the NHANES cycles, termed β_{cycle} throughout the analysis and discussion. These values are interpreted as the log change in chemical biomarker concentration for a 2-year cycle increase and represent an overall trajectory across the eight NHANES cycles. The distributions of β_{cycle}*'s* are discretized into 5 trend trajectories: highly decreasing (≤ -0.30), moderately decreasing (> -0.30 and ≤ -0.125), slightly decreasing (> -0.125 and ≤ -0.045), stable (> -0.045 and ≤ 0.041), and increasing (> 0.041). The majority of chemical biomarkers have β_{cycle}*'s* between -0.045 and 0.041, implying little or no variation over time. The majority of pesticides and PFASs have negative β_{cycle}*'s*, demonstrating a decrease in chemical biomarker levels over time, while a few pesticides, phthalates, and PAHs have high positive β_{cycle}*'s*, reflecting increasing time trends.

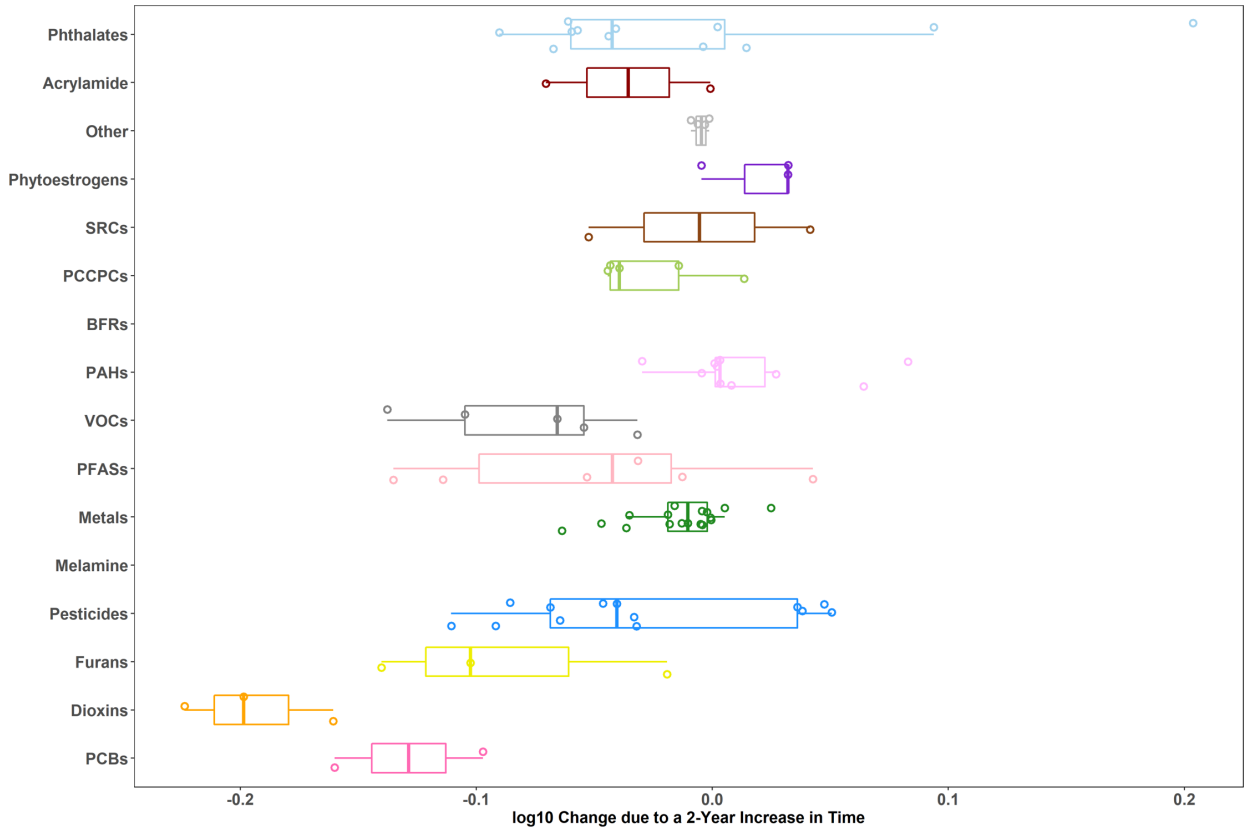

Figure A1.3. Characteristics of the 141 NHANES chemical exposure biomarkers from 16 classes for ranges of cycle coefficients, defined as the percent change in chemical concentration due to a two-year (one NHANES cycle) increase in time. The classes are ranked by the means of class-specific age percent differences (Figure 2C). Colors are used to differentiate the chemical classes. BFRs, Brominated Flame Retardants; SRCs, Smoking Related Compounds; PAHs, Polycyclic Aromatic Hydrocarbons; PCCPCs, Personal Care and Consumer Product Compounds; VOCs, Volatile Organic Compounds; PFCs, Perfluoroalkyl Chemicals; PCBs, Polychlorinated Biphenyls

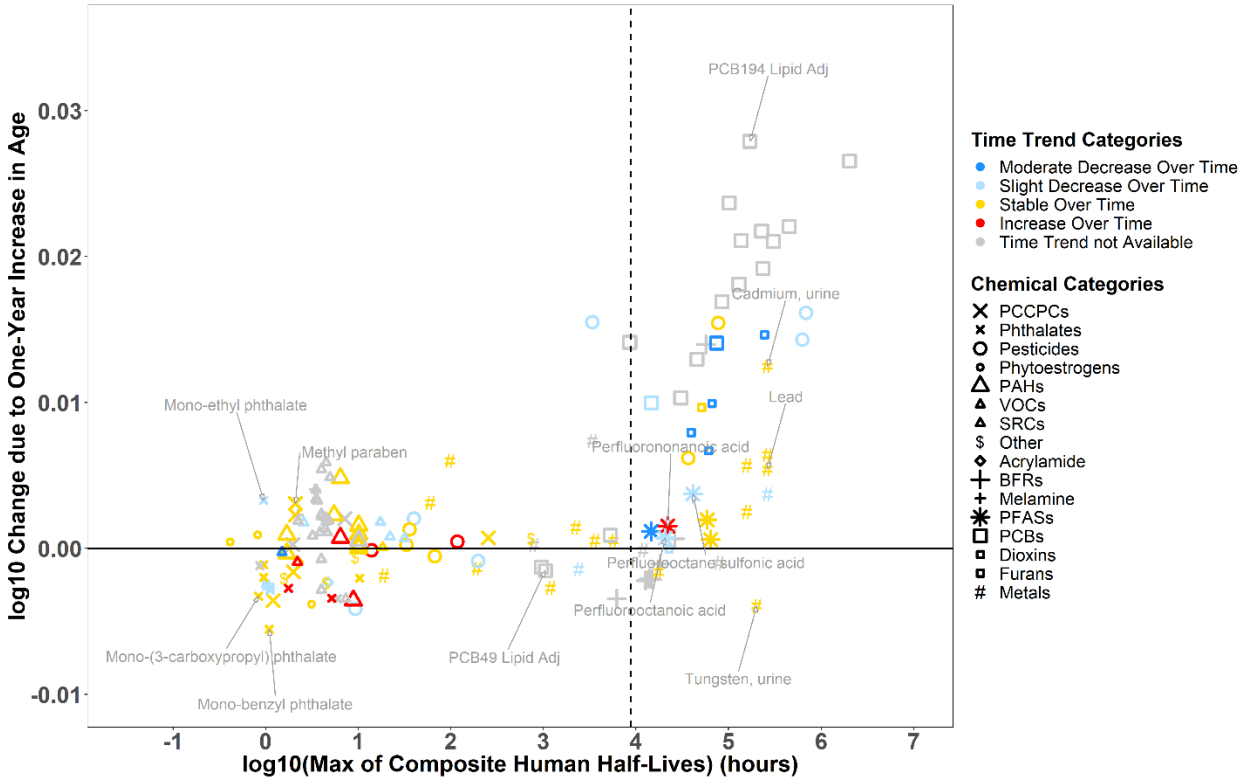

Figure A1.4. Association between linear age coefficients and chemical persistency in the human body for 144 substances with colors indicating the time trend trajectories and symbols indicating the different chemical classes. The same abbreviations for the chemical classes are used as those in Figure S3.

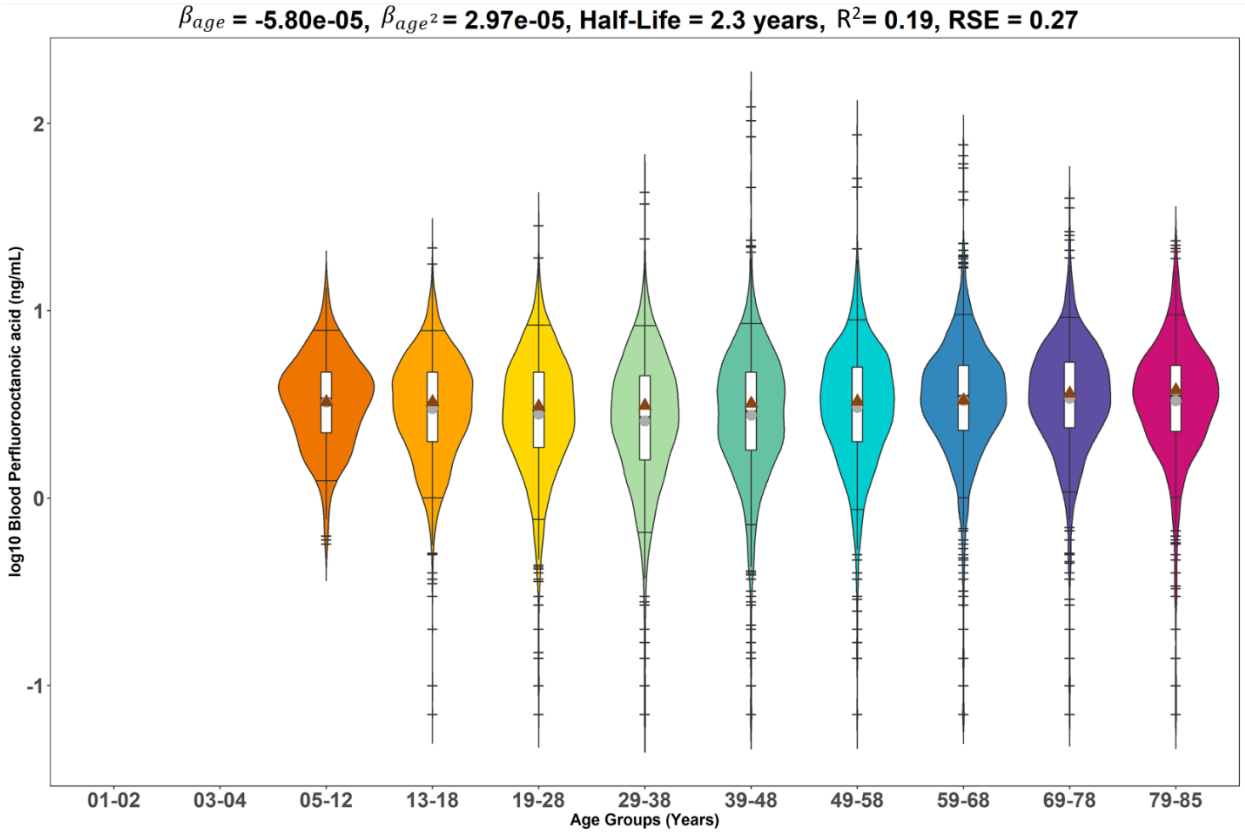

Figure A1.5. Violin plots of PFOA concentrations partitioned by age groups to display the distribution with the 5th, 25th, 50th, 75th, and 95th percentiles as indicated by the superimposed boxplot and show the frequency of the urinary cadmium biomarker levels represented by the width of the violins.

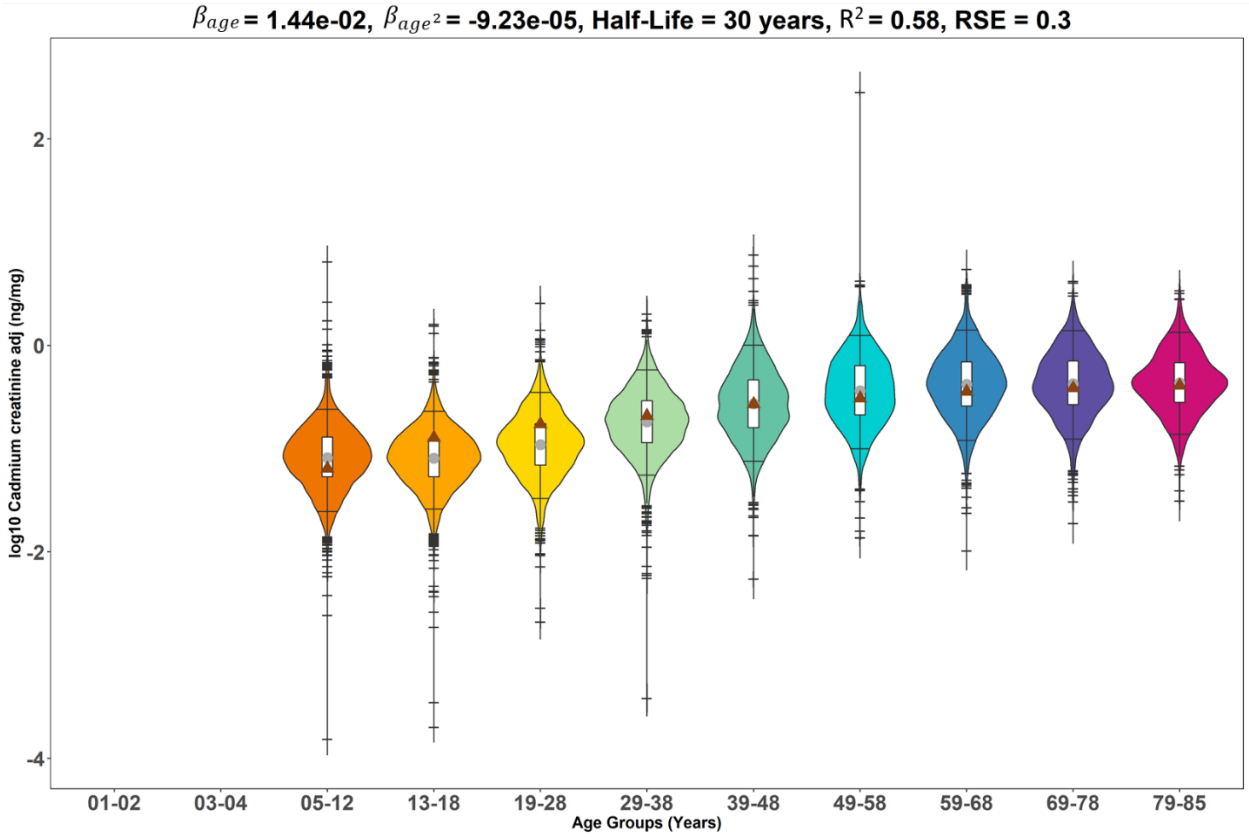

Figure A1.6. Violin plots of urinary cadmium concentrations partitioned by age groups to display the distribution with the 5th, 25th, 50th, 75th, and 95th percentiles as indicated by the superimposed boxplot and show the frequency of the urinary cadmium biomarker levels represented by the width of the violins.

Appendix 2. Racial Disparities in Chemical Biomarker Concentrations in United States Women

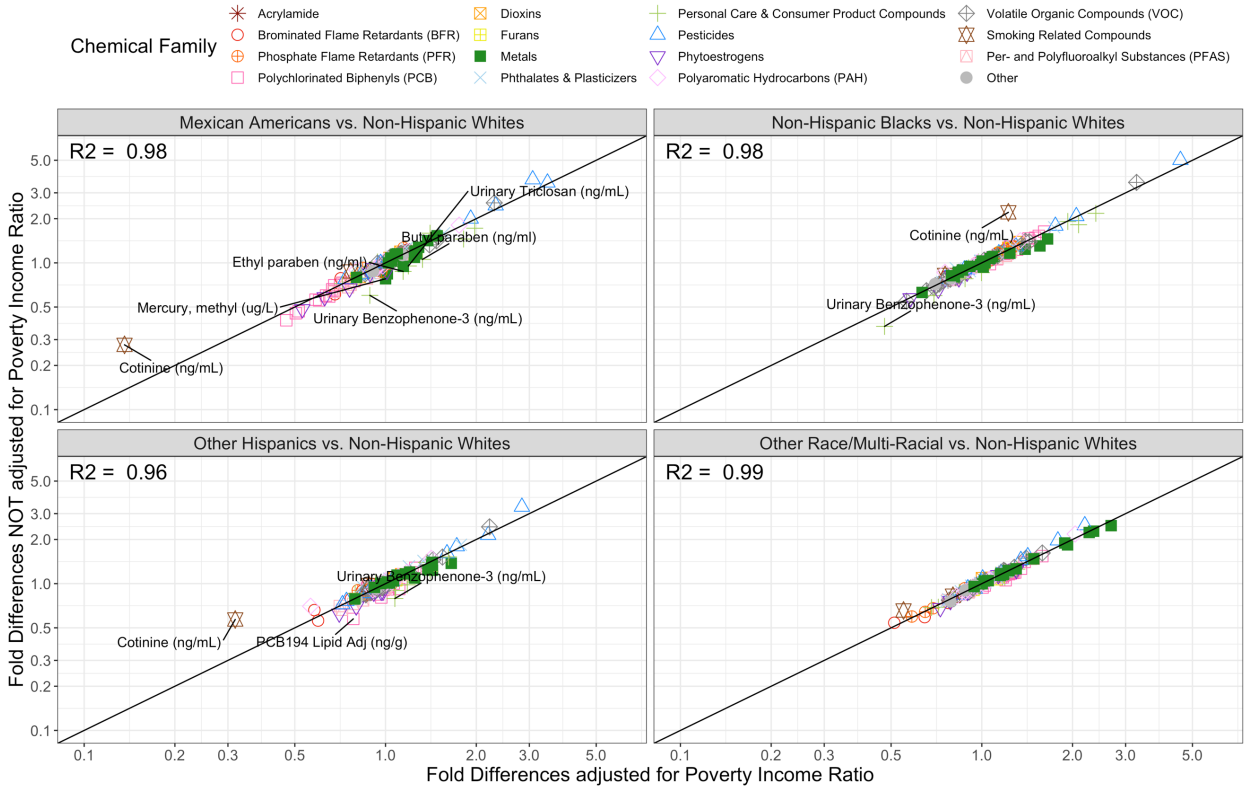

Figure A2.1. Panel of correlation plots comparing fold differences for race that adjusted for poverty income ratio (PIR) with those that excluded PIR from the regression models. Colors and shapes represent the different chemical families. Chemicals were labeled if fold differences changed by greater than 25% when PIR was considered a covariate in the regression models.

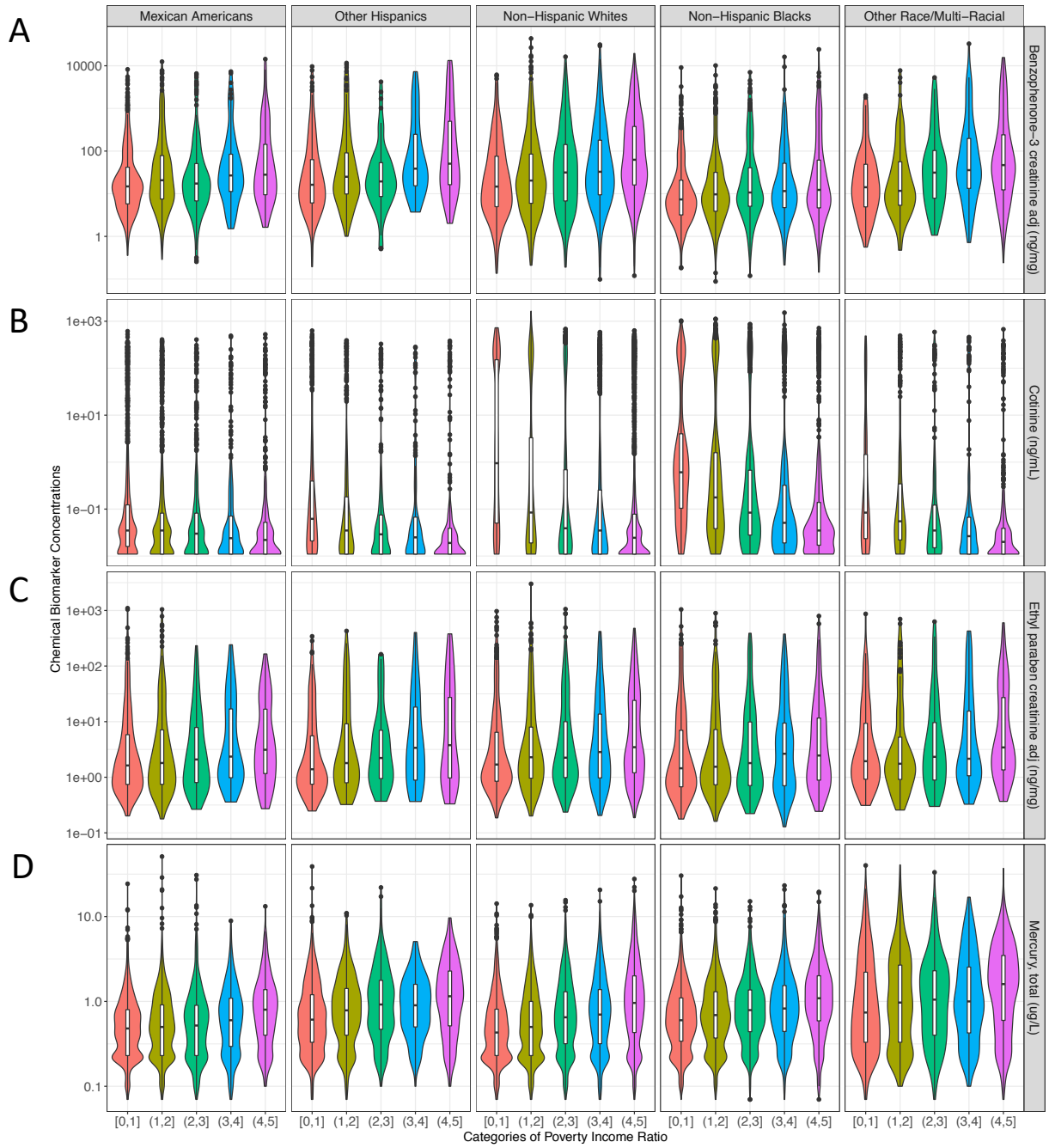

Figure A2.2. Panel of violin plots showing the distribution of chemical biomarker levels changes across categories of poverty income ratio (PIR) and stratified by race for A) an indicator of sunscreen use, benzophenone-3, B) a biomarker of smoking, cotinine, C) a chemical used in personal care products, ethyl paraben, and D) methyl mercury. Colors represent different categories of PIR.

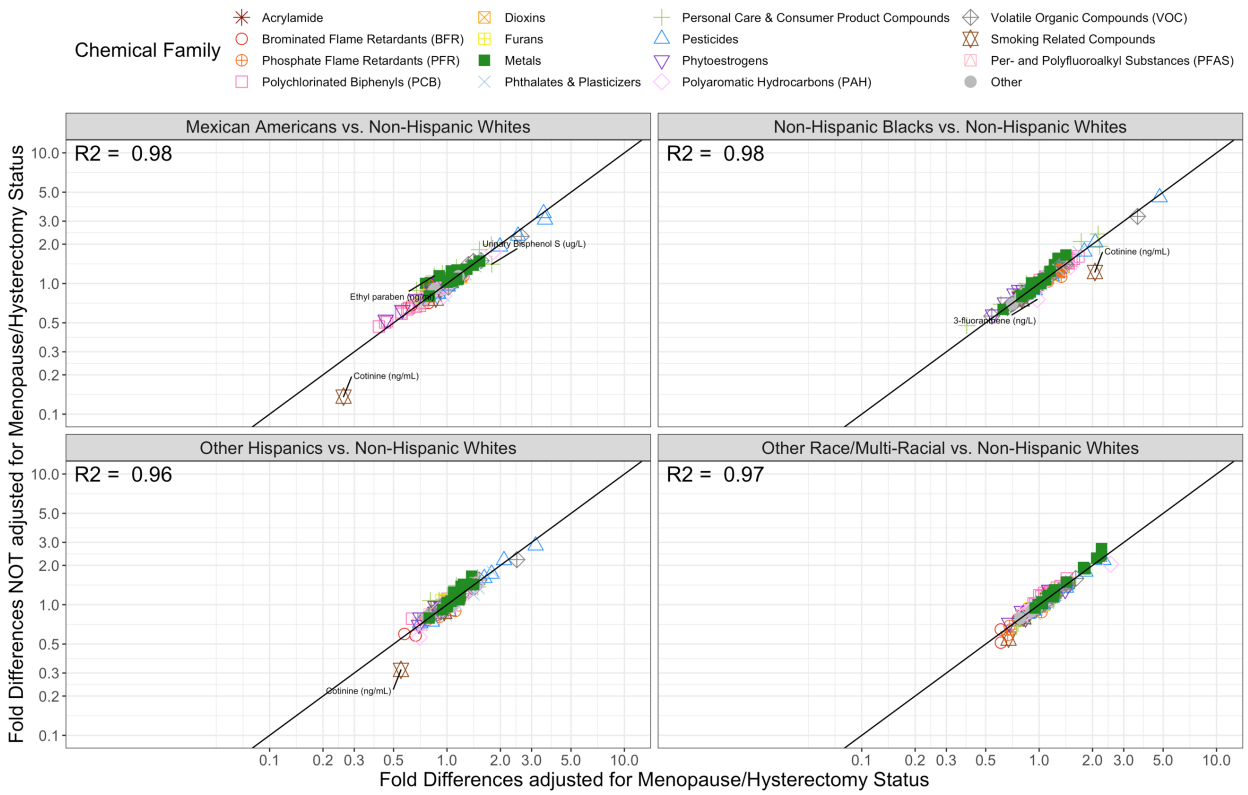

Figure A2.3. Panel of correlation plots comparing fold differences for race that adjusted for menopause/hysterectomy status with those that excluded this reproductive health variable from the regression models. Colors and shapes represent the different chemical families. Chemicals were labeled if fold differences changed by greater than 25% when reasons of having irregular periods was considered a covariate in the regression models.

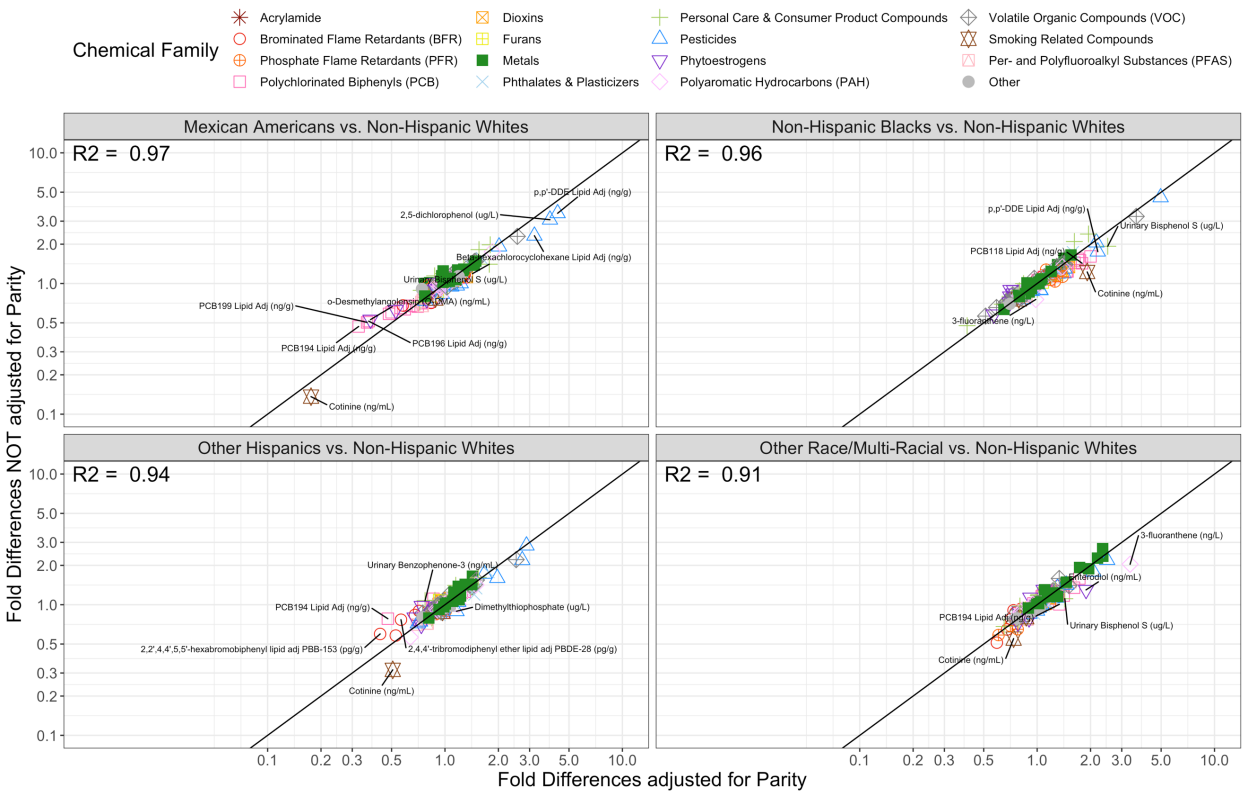

Figure A2.4. Panel of correlation plots comparing fold differences for race that adjusted for parity with those that excluded parity from the regression models. Colors and shapes represent the different chemical families. Chemicals were labeled if fold differences changed by greater than 25% when parity was considered a covariate in the regression models.

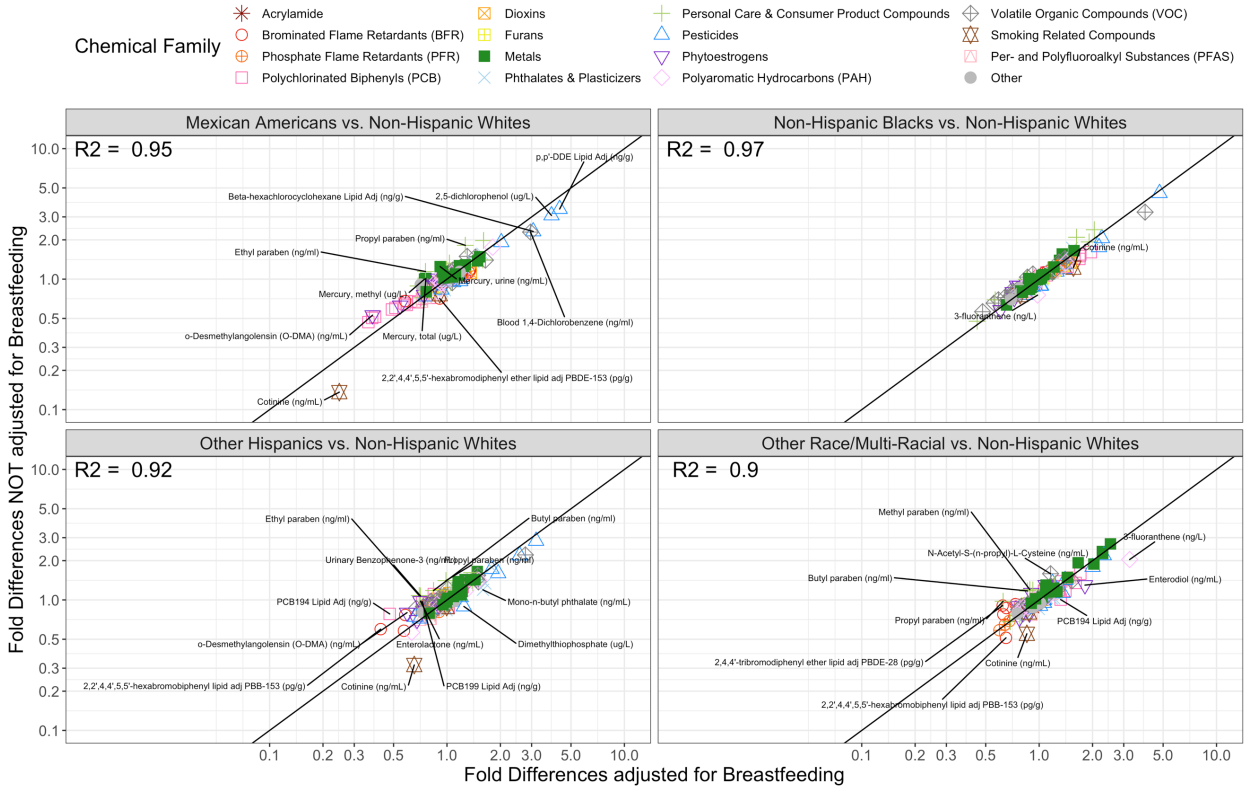

Figure A2.5. Panel of correlation plots comparing fold differences for race that adjusted for breastfeeding for at least a month with those that excluded this reproductive health variable from the regression models. Colors and shapes represent the different chemical families. Chemicals were labeled if fold differences changed by greater than 25% when breastfeeding was considered a covariate in the regression models.

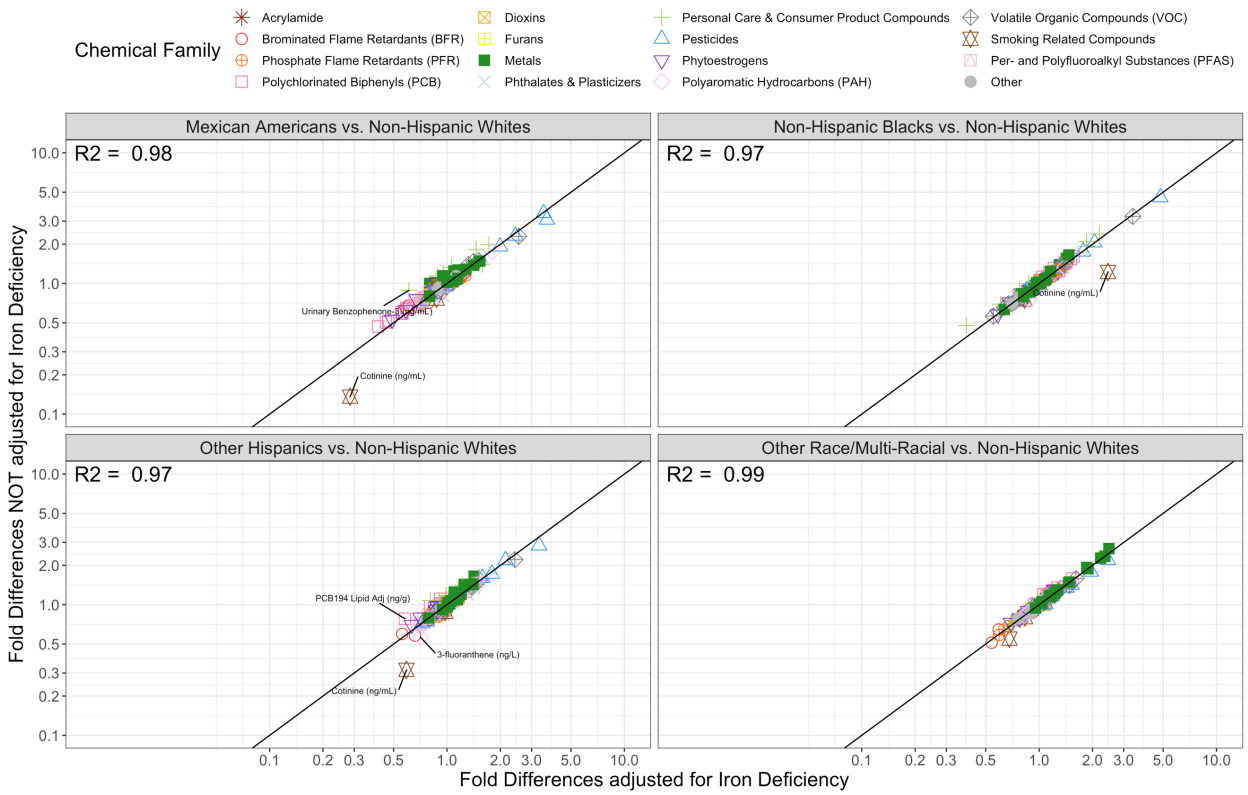

Figure A2.6. Panel of correlation plots comparing fold differences for race that adjusted for iron deficiency with those that excluded this nutritional factor from the regression models. Colors and shapes represent the different chemical families. Chemicals were labeled if fold differences changed by greater than 25% when breastfeeding was considered a covariate in the regression models.

Table A2.1. Indicator (marked by an "X") of which chemical biomarker measurements for a given NHANES cycle was excluded from analysis.

Chemical name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8
Acrylamide (pmoL/G Hb)								
Glycideamide (pmoL/G Hb)								
2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)								
2,2',4-tribromodiphenyl ether lipid adj (ng/g)								
2,4,4'-tribromodiphenyl ether lipid adj (ng/g)								
2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)								
2,2',3,4,4'-pentabromodiphenyl ether lipid adj (ng/g)								
2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)								
2,3',4,4'-tetrabromodiphenyl lipid adj (ng/g)								
2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)								
2,2',4,4',5,5'-hexabromodiphenyl lipid adj (ng/g)								
2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)								
2,2',3,4,4',5',6-heptabromodiphenyl ether lipid adj (ng/g)								
1,2,3,7,8-pncdd Lipid Adj (pg/g)								
1,2,3,4,7,8-hxcdd Lipid Adj (pg/g)								
1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	X							
1,2,3,7,8,9-hxcdd Lipid Adj (pg/g)								
1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	X							
1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	X							
2,3,7,8-tcdd Lipid Adj (pg/g)								
2,3,7,8-tcdf Lipid Adj (pg/g)								
1,2,3,7,8-pncdf Lipid Adj (pg/g)								
2,3,4,7,8-pncdf Lipid Adj (pg/g)	X							
1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	X							
1,2,3,6,7,8-hxcdf Lipid Adj (pg/g)								
1,2,3,7,8,9-hxcdf Lipid Adj (pg/g)								

2,3,4,6,7,8-hxcdf Lipid Adj (pg/g)								
1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	X							
1,2,3,4,7,8,9-hpcdf Lipid Adj (pg/g)								
1,2,3,4,6,7,8,9-ocdf Lipid Adj (pg/g)								
Cyanuric acid (ng/mL)								
Melamine (ng/mL)								
Cadmium (ug/L)								
Mercury, ethyl (ug/L)								
Mercury, methyl (ug/L)								
Blood manganese (ug/L)								
Lead (ug/dL)								
Mercury, Inorganic (ug/L)								
Serum Copper (ug/dL)								
Serum Zinc (ug/dL)								
Mercury, total (ug/L)								
Urinary Arsenobetaine (µg/L)								
Urinary Arsenocholine (µg/L)								
Urinary total Arsenic (µg/L)								
Urinary Arsenous acid (µg/L)								
Urinary Arsenic acid (µg/L)								
Barium, urine (ng/mL)								
Beryllium, urine (ng/mL)								
Cadmium, urine (ng/mL)								
Cobalt, urine (ng/mL)								
Cesium, urine (ng/mL)								
Urinary Dimethylarsonic acid (µg/L)								
Mercury, urine (ng/mL)								
Urinary Monomethylacrsonic acid (µg/L)								
Molybdenum, urine (ng/mL)								
Lead, urine (ng/mL)								

Platinum, urine (ng/mL)								
Antimony, urine (ng/mL)								
Thallium, urine (ng/mL)								
Urinary Trimethylarsine Oxide (µg/L)								
Tungsten, urine (ng/mL)								
Uranium, urine (ng/mL)								
Urinary nitrate (ng/mL)								
Urinary thiocyanate (ng/mL)								
Iodine, urine (ng/mL)								
Perchlorate, urine (ng/mL)								
Mono-2-hydroxy-iso-butyl phthlte (ng/mL)								
Mono-3-hydroxy-n-butyl phthalate (ng/mL)								
Mono(carboxynonyl) phthalate (ng/mL)								
Mono(carboxyoctyl) phthalate (ng/mL)								
Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)								
Mono-n-butyl phthalate (ng/mL)								
Mono-(3-carboxypropyl) phthalate (ng/mL)								
Mono-cyclohexyl phthalate (ng/mL)								
Mono-ethyl phthalate (ng/mL)								
Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)								
MHNCH (ng/mL)								
Mono-(2-ethyl)-hexyl phthalate (ng/mL)								
Mono-isobutyl pthalate (ng/mL)								
Mono-n-methyl phthalate (ng/mL)								
Mono-isononyl phthalate (ng/mL)								
Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)								
Mono-n-octyl phthalate (ng/mL)								
Mono-benzyl phthalate (ng/mL)								
1-napthol (ng/L)								
2-napthol (ng/L)		X						

3-fluorene (ng/L)								
2-fluorene (ng/L)								
3-phenanthrene (ng/L)								
1-phenanthrene (ng/L)								
2-phenanthrene (ng/L)								
3-fluoranthene (ng/L)								
1-pyrene (ng/L)								X
9-fluorene (ng/L)						X	X	
4-phenanthrene (ng/L)								
2 & 3-Hydroxyphenanthrene (ng/L)								
PCB199 Lipid Adj (ng/g)		X						
PCB28 Lipid Adj (ng/g)								
PCB44 Lipid Adj (ng/g)								
PCB49 Lipid Adj (ng/g)								
PCB52 Lipid Adj (ng/g)								
PCB66 Lipid Adj (ng/g)								
PCB74 Lipid Adj (ng/g)	X	X						
PCB87 Lipid Adj (ng/g)								
PCB99 Lipid Adj (ng/g)	X	X						
PCB101 Lipid Adj (ng/g)								
PCB105 Lipid Adj (ng/g)								
PCB110 Lipid Adj (ng/g)								
PCB118 Lipid Adj (ng/g)	X	X						
PCB128 Lipid Adj (ng/g)								
PCB138 Lipid Adj (ng/g)	X	X						
PCB146 Lipid Adj (ng/g)								
PCB149 Lipid Adj (ng/g)								
PCB151 Lipid Adj (ng/g)								
PCB153 Lipid Adj (ng/g)	X	X						
PCB156 Lipid Adj (ng/g)								

PCB157 Lipid Adj (ng/g)								
PCB167 Lipid Adj (ng/g)								
PCB170 Lipid Adj (ng/g)	X	X						
PCB172 Lipid Adj (ng/g)								
PCB177 Lipid Adj (ng/g)								
PCB178 Lipid Adj (ng/g)								
PCB180 Lipid Adj (ng/g)	X	X						
PCB183 Lipid Adj (ng/g)								
PCB187 Lipid Adj (ng/g)	X	X						
PCB189 Lipid Adj (ng/g)								
PCB194 Lipid Adj (ng/g)		X						
PCB195 Lipid Adj (ng/g)								
PCB196 Lipid Adj (ng/g)		X						
PCB206 Lipid Adj (ng/g)								
PCB209 Lipid Adj (ng/g)								
3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	X							
3,3',4,4',5-pcnb Lipid Adj (pg/g)	X							
3,4,4',5-tcb Lipid Adj (pg/g)								
Urinary 4-tert-octylphenol (ng/mL)								
Urinary Benzophenone-3 (ng/mL)								
Urinary Bisphenol F (ug/L)								
Urinary Bisphenol A (ng/mL)								
Urinary Bisphenol S (ug/L)								
Butyl paraben (ng/ml)								
Ethyl paraben (ng/ml)								
Methyl paraben (ng/ml)								
Propyl paraben (ng/ml)								
Urinary Triclocarban (ng/mL)								
Urinary Triclosan (ng/mL)								
Aldrin Lipid Adj (ng/g)								

Beta-hexachlorocyclohexane Lipid Adj (ng/g)								
Dieldrin Lipid Adj (ng/g)								
Endrin Lipid Adj (ng/g)								
G-hexachlorocyclohexane Lipid Adj (ng/g)								
Hexachlorobenzene Lipid Adj (ng/g)								
Heptachlor Epoxide Lipid Adj (ng/g)								
Mirex Lipid Adj (ng/g)								
o,p'-DDT Lipid Adj (ng/g)								
Oxychlordane Lipid Adj (ng/g)								
p,p'-DDE Lipid Adj (ng/g)								
p,p'-DDT Lipid Adj (ng/g)								
Trans-nonachlor Lipid Adj (ng/g)								
2,5-dichlorophenol (ug/L)								
2,4,5-trichlorophenol (ug/L)								
2,4-D (ug/L)								
2,4,5-T (ug/L)								
2,4,6-trichlorophenol (ug/L)								
4-fluoro-3-phenoxybenzoic acid (ug/L)								
Atrazine (ug/L)								
Acetochlor mercapturate (ug/L)								
Alachor mercapturate (ug/L)								
Acephate (ug/L)								
Atrazine mercapturate (ug/L)								
Bensulfuron methyl (ug/L)								
dibromovinyl-dimeth prop carboacid (ug/L)								
Carbofuranphenol (ug/L)								
cis dichlorovnl-dimeth carboacid (ug/L)								
Chlorsulfuron (ug/L)								
chloro-hydro-meth-chromen-one/ol (ug/L)								
3,5,6-trichloropyridinol (ug/L)								

2,4-dichlorophenol (ug/L)								
Diaminochloroatrazine (ug/L)								
DEET acid (ug/L)								
DEET (ug/L)								
Desethyl hydroxy DEET (ug/L)								
diethylaminomethylpyrimidinol/one (ug/L)								
Desethyl atrazine (ug/L)								
Ethametsulfuron methyl (ug/L)								
Ethylenethio urea (ug/L)								
Foramsulfuron (ug/L)								
Halosulfuron (ug/L)								
Malathion diacid (ug/L)								
Metolachlor mercapturate (ug/L)								
Methamidaphos (ug/L)								
Mesosulfuron methyl (ug/L)								
Metsulfuron methyl (ug/L)								
Dimethoate (ug/L)								
Nicosulfuron (ug/L)								
O-methoate (ug/L)								
Dimethylphosphate (ug/L)								
Diethylphosphate (ug/L)								
Dimethylthiophosphate (ug/L)								
Diethylthiophosphate (ug/L)								
Dimethyldithiophosphate (ug/L)								
Diethyldithiophosphate (ug/L)								
3-phenoxybenzoic acid (ug/L)								
O-Phenyl phenol (ug/L)								
Oxasulfuron (ug/L)								
Oxypyrimidine (ug/L)								
Paranitrophenol (ug/L)	X							

Pentachlorophenol (ug/L)								
Primisulfuron methyl (ug/L)								
2-isopropoxyphenol (ug/L)								
Prosulfuron (ug/L)								
Propylenethio urea (ug/L)								
Rimsulfuron (ug/L)								
Desisopropyl atrazine (ug/L)								
Desisopropyl atrazine mercapturate (ug/L)								
Sulfometuron methyl (ug/L)								
Sulfosulfuron (ug/L)								
trans dichlorovnl-dimeth carboacid (ug/L)								
Thifensulfuron methyl (ug/L)								
Triasulfuron (ug/L)								
Triflusulfuron methyl (ug/L)								
2-(N-ethyl-PFOSA) acetate (ng/mL)								
Perfluorobutane sulfonic acid (ng/mL)								
2-(N-methyl-PFOSA) acetate (ng/mL)								
Perfluorodecanoic acid (ng/mL)								
Perfluorododecanoic acid (ng/mL)								
Perfluoroheptanoic acid (ng/mL)								
Perfluorohexane sulfonic acid (ng/mL)								
Perfluorononanoic acid (ng/mL)								
Perfluorooctanoic acid (ng/mL)								
Perfluorooctane sulfonic acid (ng/mL)								
Perfluorooctane sulfonamide (ng/mL)								
Perfluoroundecanoic acid (ng/mL)								
Br. iso of perfluorooctanoate (ug/L)								
Monomethyl branched iso of PFOS (ug/L)								
Linear perfluorooctanoate (ug/L)								
Linear perfluorooctane sulfonate (ug/L)								

Bis(2-chloroethyl) phosphate (ug/L)								
Bis(1-chloro-2-propyl) phosphate (ug/L)								
Bis(1,3-dichloro-2-propyl) phosphate (ug/L)								
Dibutyl phosphate (ug/L)								
Dibenzyl phosphate (ug/L)								
Di-o-cresyl phosphate (ug/L)								
Di-p-cresyl phosphate (ug/L)								
Diphenyl phosphate (ug/L)								
2,3,4,5-tetrabromobenzoic acid (ug/L)								
Daidzein (ng/mL)		X						
o-Desmethylangolensin (O-DMA) (ng/mL)								
Equol (ng/mL)	X	X						
Enterodiol (ng/mL)	X	X	X					
Enterolactone (ng/mL)	X	X						
Genistein (ng/mL)				X				
Cotinine (ng/mL)								
Hydroxycotinine, Serum (ng/mL)								
NNAL , urine (ng/mL)								
Blood 2,5-Dimethylfuran (ng/mL)								
Blood 1,1,1,2-tetrachloroethane (ng/mL)								
Blood Nitromethane (pg/mL)								
Blood Hexane (ng/mL)								
Blood Heptane (ng/mL)								
Blood Octane (ng/mL)								
Blood 1,1-Dichloroethane (ng/mL)								
Blood 1,2-Dichlorobenzene (ng/mL)								
Blood 1,1-Dichloroethene (ng/mL)								
Blood 1,2-Dichloroethane (ng/mL)								
Blood cis-1,2-Dichloroethene (ng/mL)								
Blood 1,1,2-Trichloroethane (ng/mL)								

Blood 1,2-Dibromo-3-chloropropane (ng/mL)								
Blood trans-1,2-Dichloroethene (ng/mL)								
Blood 1,3-Dichlorobenzene (ng/mL)								
Blood Tetrachloroethene (ng/ml)								
Blood 1,1,2,2-Tetrachloroethane (ng/mL)								
Blood Bromoform (pg/ml)								
Blood Bromodichloromethane (pg/ml)	X							X
Blood Benzene (ng/ml)								
Blood Cyclohexane (ng/mL)								
Blood Chlorobenzene (ng/mL)								
Blood Chloroform (pg/ml)	X							X
Blood Dibromochloromethane (pg/ml)								
Blood Carbon Tetrachloride (ng/ml)								
Blood 1,4-Dichlorobenzene (ng/ml)	X	X	X					
Blood 1,2-dibromoethane (ng/ml)								
Blood Diethyl Ether (ng/mL)								
Blood Dibromomethane (ng/mL)								
Blood 1,2-Dichloropropane (ng/mL)								
Blood 1,4-Dioxane (ng/mL)								
Blood Ethyl Acetate (ng/mL)								
Blood Ethylbenzene (ng/ml)								
Blood Chloroethane (ng/mL)								
Blood furan (ng/ml)								
Blood Hexachloroethane (ng/mL)								
Blood isopropylbenzene (ng/ml)								
Blood Methylene Chloride (ng/mL)								
Blood Methylcyclopentane (ng/mL)								
Blood MTBE (pg/ml)								
Blood Nitrobenzene (ng/mL)								
Blood o-Xylene (ng/ml)								

Blood Styrene (ng/ml)								
Blood Trichloroethene (ng/ml)								
Blood 1,1,1-Trichloroethane (ng/mL)								
Blood aaa-Trifluorotoluene (ng/mL)								
Blood Tetrahydrofuran (ng/mL)								
Blood Toluene (ng/ml)								
Blood 1,2,3-trichloropropane (ng/ml)								
Blood Vinyl Bromide (ng/mL)								
Blood m-/p-Xylene (ng/ml)								
N-acel-S-(1,2-dichlorovinl)-L-cys (ng/mL)								
N-Acel-S-(2,2-Dichlorvinyl)-L-cys (ng/mL)								
2-Methylhippuric acid (ng/mL)								
3-methipurc acd & 4-methipurc acd (ng/mL)								
N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)								
N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)								
2-amnothiazolne-4-carbxylic acid (ng/mL)								
N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)								
N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)								
N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)								
N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)								
N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)								
N-Ace-S-(dimethylphenyl)-L-Cys (ng/mL)								
N-ac-S-(2-carbmo-2-hydxel)-L-cys (ng/mL)								
N-Ace-S-(2-Hydroxyethyl)-L-cys (ng/mL)								
N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)								
N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)								
Mandelic acid (ng/mL)								
N-A-S-(1-HydrxMet)-2-Prpn)-L-Cys (ng/mL)								
N-Ac-S-(2-Hydrxy-3-butnyl)-L-Cys (ng/mL)								
N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)								

N-ace-S-(phenl-2-hydxyetl)-L-cys (ng/mL)								
Phenylglyoxylic acid (ng/mL)								
N-Acetyl-S-(phenyl)-L-cysteine (ng/mL)								
N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)								
N-Acetyl-S-(trichlorovinyl)-L-cys (ng/mL)								
2-thoxothazlidne-4-carbxylic acid (ng/mL)								
PCB199 (ng/g)								
PCB28 (ng/g)								
PCB44 (ng/g)								
PCB49 (ng/g)								
PCB52 (ng/g)								
PCB66 (ng/g)								
PCB74 (ng/g)								
PCB87 (ng/g)								
PCB99 (ng/g)								
PCB101 (ng/g)								
PCB105 (ng/g)								
PCB110 (ng/g)								
PCB118 (ng/g)								
PCB128 (ng/g)								
PCB138 (ng/g)								
PCB146 (ng/g)								
PCB149 (ng/g)								
PCB151 (ng/g)								
PCB153 (ng/g)								
PCB156 (ng/g)								
PCB157 (ng/g)								
PCB167 (ng/g)								
PCB170 (ng/g)								
PCB172 (ng/g)								

PCB177 (ng/g)								
PCB178 (ng/g)								
PCB180 (ng/g)								
PCB183 (ng/g)								
PCB187 (ng/g)								
PCB189 (ng/g)								
PCB194 (ng/g)								
PCB195 (ng/g)								
PCB196 (ng/g)								
PCB206 (ng/g)								
PCB209 (ng/g)								
Aldrin (ng/g)								
2,2',4,4',5,5'-hexabromobiphenyl (pg/g)								
Beta-hexachlorocyclohexane (ng/g)								
2,2',4-tribromodiphenyl ether (pg/g)								
2,4,4'-tribromodiphenyl ether (pg/g)								
2,2',4,4'-tetrabromodiphenyl ether (pg/g)								
2,2',3,4,4'-pentabromodiphenyl ether (pg/g)								
2,2',4,4',5-pentabromodiphenyl ether (pg/g)								
2,2',4,4',6-pentabromodiphenyl ether (pg/g)								
2,3',4,4'-tetrabromodiphenyl ether (pg/g)								
2,2',4,4',5,5'-hexabromodiphenyl ether (pg/g)								
2,2',4,4',5,6'-hexabromodiphenyl ether (pg/g)								
2,2',3,4,4',5',6-heptabromodiphenyl ether (pg/g)								
1,2,3,7,8-pncdd (fg/g)								
1,2,3,4,7,8-hxcdd (fg/g)								
1,2,3,6,7,8-hxcdd (fg/g)								
1,2,3,7,8,9-hxcdd (fg/g)								
1,2,3,4,6,7,8-hpcdd (fg/g)								
1,2,3,4,6,7,8,9-ocdd (fg/g)								

Dieldrin (ng/g)								
Endrin (ng/g)								
2,3,7,8-tcdf (fg/g)								
1,2,3,7,8-pncdf (fg/g)								
2,3,4,7,8-pncdf (fg/g)								
1,2,3,4,7,8-hxcdf (fg/g)								
1,2,3,6,7,8-hxcdf (fg/g)								
1,2,3,7,8,9-hxcdf (fg/g)								
2,3,4,6,7,8-hxcdf (fg/g)								
1,2,3,4,6,7,8-hpcdf (fg/g)								
1,2,3,4,7,8,9-hpcdf (fg/g)								
1,2,3,4,6,7,8,9-ocdf (fg/g)								
Gamma-hexachlorocyclohexane (ng/g)								
Hexachlorobenzene (ng/g)								
Heptachlor Epoxide (ng/g)								
3,3',4,4',5,5'-hxcb (fg/g)								
Mirex (ng/g)								
o,p'-DDT (ng/g)								
Oxychlordane (ng/g)								
3,3',4,4',5-pcnb (pg/g)								
p,p'-DDE (ng/g)								
p,p'-DDT (ng/g)								
3,4,4',5-tcb (fg/g)								
2,3,7,8-tcdd (fg/g)								
Trans-nonachlor (ng/g)								

Table A2.2. Included chemical biomarkers for analysis with corresponding CAS No. and chemical family classification.

Chemical Name	CAS NO.	Chemical Family
Glycideamide (pmoL/G Hb)	5694-00-8	Acrylamide
Acrylamide (pmoL/G Hb)	79-06-1	Acrylamide
2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)	59080-40-9	Brominated Flame Retardants (BFR)
2,4,4'-tribromodiphenyl ether lipid adj (ng/g)	41318-75-6	Brominated Flame Retardants (BFR)
2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)	207122-15-4	Brominated Flame Retardants (BFR)
2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)	189084-64-8	Brominated Flame Retardants (BFR)
2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)	5436-43-1	Brominated Flame Retardants (BFR)
2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)	60348-60-9	Brominated Flame Retardants (BFR)
2,2',4,4',5,5'-hexabromodiphenyl lipid adj (ng/g)	68631-49-2	Brominated Flame Retardants (BFR)
Bis(2-chloroethyl) phosphate (ug/L)	3040-56-0	Phosphate Flame Retardants (PFR)
Bis(1-chloro-2-propyl) phosphate (ug/L)	789440-10-4	Phosphate Flame Retardants (PFR)
Bis(1,3-dichloro-2-propyl) phosphate (ug/L)	72236-72-7	Phosphate Flame Retardants (PFR)
Dibutyl phosphate (ug/L)	107-66-4	Phosphate Flame Retardants (PFR)
Diphenyl phosphate (ug/L)	838-85-7	Phosphate Flame Retardants (PFR)
1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	57653-85-7	Dioxins
1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	3268-87-9	Dioxins
1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	35822-46-9	Dioxins
1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	70648-26-9	Furans
2,3,4,7,8-pncdf Lipid Adj (pg/g)	57117-31-4	Furans
1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	67562-39-4	Furans
Lead (ug/dL)	7439-92-1	Metals
Cadmium (ug/L)	7440-43-9	Metals
Mercury, total (ug/L)	7439-97-6	Metals
Mercury, urine (ng/mL)	7439-97-6	Metals
Barium, urine (ng/mL)	7440-39-3	Metals
Cobalt, urine (ng/mL)	7440-48-4	Metals
Cesium, urine (ng/mL)	7440-46-2	Metals

Molybdenum, urine (ng/mL)	7439-98-7	Metals
Lead, urine (ng/mL)	7439-92-1	Metals
Antimony, urine (ng/mL)	7440-36-0	Metals
Thallium, urine (ng/mL)	7440-28-0	Metals
Tungsten, urine (ng/mL)	7440-33-7	Metals
Uranium, urine (ng/mL)	7440-61-1	Metals
Urinary total Arsenic (µg/L)	7440-38-2	Metals
Urinary Arsenobetaine (µg/L)	64436-13-1	Metals
Urinary Dimethylarsonic acid (µg/L)	75-60-5	Metals
Cadmium, urine (ng/mL)	7440-43-9	Metals
Blood manganese (ug/L)	7439-96-5	Metals
Serum Copper (ug/dL)	7440-50-8	Metals
Serum Zinc (ug/dL)	7440-66-6	Metals
Mercury, methyl (ug/L)	22967-92-6	Metals
Urinary thiocyanate (ng/mL)	302-04-5	Other
Iodine, urine (ng/mL)	7553-56-2	Other
Urinary nitrate (ng/mL)	14797-55-8	Other
Perchlorate, urine (ng/mL)	14797-73-0	Other
Urinary Triclosan (ng/mL)	3380-34-5	Personal Care & Consumer Product Compounds
Urinary Bisphenol S (ug/L)	80-09-1	Personal Care & Consumer Product Compounds
Butyl paraben (ng/ml)	94-26-8	Personal Care & Consumer Product Compounds
Ethyl paraben (ng/ml)	120-47-8	Personal Care & Consumer Product Compounds
Methyl paraben (ng/ml)	99-76-3	Personal Care & Consumer Product Compounds
Propyl paraben (ng/ml)	94-13-3	Personal Care & Consumer Product Compounds
Urinary Bisphenol A (ng/mL)	80-05-7	Personal Care & Consumer Product Compounds
Urinary Bisphenol F (ug/L)	620-92-8	Personal Care & Consumer Product Compounds
Urinary Benzophenone-3 (ng/mL)	131-57-7	Personal Care & Consumer Product Compounds
Trans-nonachlor Lipid Adj (ng/g)	39765-80-5	Pesticides
Oxychlordane Lipid Adj (ng/g)	27304-13-8	Pesticides
p,p'-DDE Lipid Adj (ng/g)	72-55-9	Pesticides

Dieldrin Lipid Adj (ng/g)	60-57-1	Pesticides
Beta-hexachlorocyclohexane Lipid Adj (ng/g)	319-85-7	Pesticides
3,5,6-trichloropyridinol (ug/L)	6515-38-4	Pesticides
Dimethylthiophosphate (ug/L)	1112-38-5	Pesticides
3-phenoxybenzoic acid (ug/L)	3739-38-6	Pesticides
2,5-dichlorophenol (ug/L)	583-78-8	Pesticides
2,4-dichlorophenol (ug/L)	120-83-2	Pesticides
Paranitrophenol (ug/L)	100-02-7	Pesticides
DEET acid (ug/L)	72236-23-8	Pesticides
Mono(carboxynonyl) phthalate (ng/mL)	26761-40-0	Phthalates & Plasticizers
Mono(carboxyoctyl) phthalate (ng/mL)	898544-09-7	Phthalates & Plasticizers
Mono-isobutyl pthalate (ng/mL)	30833-53-5	Phthalates & Plasticizers
Mono-(2-ethyl)-hexyl phthalate (ng/mL)	4376-20-9	Phthalates & Plasticizers
Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)	40321-98-0	Phthalates & Plasticizers
Mono-benzyl phthalate (ng/mL)	2528-16-7	Phthalates & Plasticizers
Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)	40321-99-1	Phthalates & Plasticizers
Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)	40809-41-4	Phthalates & Plasticizers
Mono-n-butyl phthalate (ng/mL)	131-70-4	Phthalates & Plasticizers
Mono-ethyl phthalate (ng/mL)	2306-33-4	Phthalates & Plasticizers
Mono-n-methyl phthalate (ng/mL)	4376-18-5	Phthalates & Plasticizers
Mono-(3-carboxypropyl) phthalate (ng/mL)	66851-46-5	Phthalates & Plasticizers
o-Desmethylangolensin (O-DMA) (ng/mL)	21255-69-6	Phytoestrogens
Enterolactone (ng/mL)	78473-71-9	Phytoestrogens
Enterodiol (ng/mL)	80226-00-2	Phytoestrogens
Daidzein (ng/mL)	486-66-8	Phytoestrogens
Equol (ng/mL)	531-95-3	Phytoestrogens
Genistein (ng/mL)	446-72-0	Phytoestrogens
3-phenanthrene (ng/L)	605-87-8	Polyaromatic Hydrocarbons (PAH)
1-phenanthrene (ng/L)	2433-56-9	Polyaromatic Hydrocarbons (PAH)
2-phenanthrene (ng/L)	605-55-0	Polyaromatic Hydrocarbons (PAH)

4-phenanthrene (ng/L)	7651-86-7	Polyaromatic Hydrocarbons (PAH)
2 & 3-Hydroxyphenanthrene (ng/L)	605-55-0	Polyaromatic Hydrocarbons (PAH)
1-pyrene (ng/L)	129-00-0	Polyaromatic Hydrocarbons (PAH)
1-napthol (ng/L)	90-15-3	Polyaromatic Hydrocarbons (PAH)
2-napthol (ng/L)	135-19-3	Polyaromatic Hydrocarbons (PAH)
9-fluorene (ng/L)	484-17-3	Polyaromatic Hydrocarbons (PAH)
3-fluorene (ng/L)	6344-67-8	Polyaromatic Hydrocarbons (PAH)
2-fluorene (ng/L)	2443-58-5	Polyaromatic Hydrocarbons (PAH)
3-fluoranthene (ng/L)	205-82-3	Polyaromatic Hydrocarbons (PAH)
PCB199 Lipid Adj (ng/g)	52663-75-9	Polychlorinated Biphenyls (PCB)
PCB180 Lipid Adj (ng/g)	35065-29-3	Polychlorinated Biphenyls (PCB)
PCB209 Lipid Adj (ng/g)	2051-24-3	Polychlorinated Biphenyls (PCB)
PCB170 Lipid Adj (ng/g)	35065-30-6	Polychlorinated Biphenyls (PCB)
PCB194 Lipid Adj (ng/g)	35694-08-7	Polychlorinated Biphenyls (PCB)
PCB187 Lipid Adj (ng/g)	52663-68-0	Polychlorinated Biphenyls (PCB)
PCB153 Lipid Adj (ng/g)	35065-27-1	Polychlorinated Biphenyls (PCB)
PCB196 Lipid Adj (ng/g)	42740-50-1	Polychlorinated Biphenyls (PCB)
PCB138 Lipid Adj (ng/g)	35065-28-2	Polychlorinated Biphenyls (PCB)
3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	32774-16-6	Polychlorinated Biphenyls (PCB)
PCB118 Lipid Adj (ng/g)	31508-00-6	Polychlorinated Biphenyls (PCB)
PCB99 Lipid Adj (ng/g)	38380-01-7	Polychlorinated Biphenyls (PCB)
3,3',4,4',5-pcnb Lipid Adj (pg/g)	57465-28-8	Polychlorinated Biphenyls (PCB)
PCB74 Lipid Adj (ng/g)	32690-93-0	Polychlorinated Biphenyls (PCB)
PCB49 Lipid Adj (ng/g)	41464-40-8	Polychlorinated Biphenyls (PCB)
PCB44 Lipid Adj (ng/g)	41464-39-5	Polychlorinated Biphenyls (PCB)
Perfluorodecanoic acid (ng/mL)	335-76-2	Per- and Polyfluoroalkyl Substances (PFAS)
2-(N-methyl-PFOSA) acetate (ng/mL)	2355-31-9	Per- and Polyfluoroalkyl Substances (PFAS)
Perfluorononanoic acid (ng/mL)	375-95-1	Per- and Polyfluoroalkyl Substances (PFAS)
Perfluorohexane sulfonic acid (ng/mL)	355-46-4	Per- and Polyfluoroalkyl Substances (PFAS)
Perfluorooctanoic acid (ng/mL)	335-67-1	Per- and Polyfluoroalkyl Substances (PFAS)

Perfluorooctane sulfonic acid (ng/mL)	1763-23-1	Per- and Polyfluoroalkyl Substances (PFAS)
Cotinine (ng/mL)	486-56-6	Smoking Related Compounds
NNAL , urine (ng/mL)	76014-81-8	Smoking Related Compounds
Blood m-/p-Xylene (ng/ml)	108-38-3/106-42-3	Volatile Organic Compounds (VOC)
Blood 1,4-Dichlorobenzene (ng/ml)	106-46-7	Volatile Organic Compounds (VOC)
Blood Bromodichloromethane (pg/ml)	75-27-4	Volatile Organic Compounds (VOC)
Phenylglyoxylic acid (ng/mL)	611-73-4	Volatile Organic Compounds (VOC)
N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)	75-56-9	Volatile Organic Compounds (VOC)
N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)	106-99-0	Volatile Organic Compounds (VOC)
3-methipurc acd & 4-methipurc acd (ng/mL)	27115-49-7	Volatile Organic Compounds (VOC)
N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)	19542-77-9	Volatile Organic Compounds (VOC)
N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)	106-94-5	Volatile Organic Compounds (VOC)
Mandelic acid (ng/mL)	90-64-2	Volatile Organic Compounds (VOC)
N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)	81690-92-8	Volatile Organic Compounds (VOC)
N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)	103974-29-4	Volatile Organic Compounds (VOC)
2-amnothiazolne-4-carbxylic acid (ng/mL)	16899-18-6	Volatile Organic Compounds (VOC)
N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)	74514-75-3	Volatile Organic Compounds (VOC)
2-thoxothazlidne-4-carbxylic acid (ng/mL)	20933-67-9	Volatile Organic Compounds (VOC)
2-Methylhippuric acid (ng/mL)	42013-20-7	Volatile Organic Compounds (VOC)
N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)	33164-70-4	Volatile Organic Compounds (VOC)
N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)	51868-61-2	Volatile Organic Compounds (VOC)
N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)	23127-40-4	Volatile Organic Compounds (VOC)
N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)	144889-50-9	Volatile Organic Compounds (VOC)
Blood Toluene (ng/ml)	108-88-3	Volatile Organic Compounds (VOC)
Blood Nitromethane (pg/mL)	75-52-5	Volatile Organic Compounds (VOC)
Blood Chloroform (pg/ml)	67-66-3	Volatile Organic Compounds (VOC)

Table A2.3. NHANES codenames for survey weights used for a given chemical biomarker and NHANES cycle.

Chemical name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8
Glycideamide (pmoL/G Hb)	NA	NA	WTMEC2YR	WTMEC2YR	NA	NA	NA	NA
Acrylamide (pmoL/G Hb)	NA	NA	WTMEC2YR	WTMEC2YR	NA	NA	NA	NA
2,2',4,4',5,5'-hexabromobiphenyl lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,4,4'-tribromodiphenyl ether lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',4,4',5,6'-hexabromodiphenyl ether lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',4,4',6-pentabromodiphenyl lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',4,4'-tetrabromodiphenyl ether lipid ad (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',4,4',5-pentabromodiphenyl lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',3,4,4',5',6-heptabromodiphenyl ether lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',4,4',5,5'-hexabromodiphenyl lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,3',4,4'-tetrabromodiphenyl lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',3,4,4'-pentabromodiphenyl ether lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
2,2',4-tribromodiphenyl ether lipid adj (ng/g)	NA	NA	WTSB2YR	NA	NA	NA	NA	NA
1,2,3,6,7,8-hxcdd Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,7,8-hxcdd Lipid Adj (pg/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,7,8-pncdd Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
2,3,7,8-tcdd Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,6,7,8,9-ocdd Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,7,8,9-hxcdd Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,6,7,8-hpcdd Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,6,7,8-hxcdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,7,8-hxcdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
2,3,4,7,8-pncdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
2,3,4,6,7,8-hxcdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,6,7,8-hpcdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,6,7,8,9-ocdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
Melamine (ng/mL)	NA	NA	WTSBMEL	NA	NA	NA	NA	NA

Cyanuric acid (ng/mL)	NA	NA	WTSBMEL	NA	NA	NA	NA	NA
Lead (ug/dL)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTSH2YR
Cadmium (ug/L)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTSH2YR
Mercury, total (ug/L)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTSH2YR
Mercury, Inorganic (ug/L)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTSH2YR
Mercury, urine (ng/mL)	WTMEC4YR	WTMEC4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Barium, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Cobalt, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Cesium, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Molybdenum, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Lead, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Platinum, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	NA	NA
Antimony, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Thallium, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Tungsten, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Uranium, urine (ng/mL)	NA	WTSHM2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary total Arsenic (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary Arsenous acid (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary Arsenobetaine (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary Arsenocholine (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary Dimethylarsonic acid (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary Monomethylacrsonic acid (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Cadmium, urine (ng/mL)	WTSHM4YR	WTSHM4YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Blood manganese (ug/L)	NA	NA	NA	NA	NA	NA	WTMEC2YR	WTSH2YR
Serum Copper (ug/dL)	NA	NA	NA	NA	NA	NA	WTSA2YR	WTSA2YR
Serum Zinc (ug/dL)	NA	NA	NA	NA	NA	NA	WTSA2YR	WTSA2YR
Mercury, methyl (ug/L)	NA	NA	NA	NA	NA	NA	WTMEC2YR	WTSH2YR
Urinary Triclocarban (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Urinary Triclosan (ng/mL)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Urinary Bisphenol S (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR

Urinary 4-tert-octylphenol (ng/mL)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	NA	NA
Butyl paraben (ng/ml)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Ethyl paraben (ng/ml)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Methyl paraben (ng/ml)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Propyl paraben (ng/ml)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Urinary Bisphenol A (ng/mL)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Urinary Bisphenol F (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Urinary Benzophenone-3 (ng/mL)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mirex Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
Trans-nonachlor Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
Oxychlordane Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
Heptachlor Epoxide Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
p,p'-DDE Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
Dieldrin Lipid Adj (ng/g)	NA	WTSPO2YR	WTSB2YR	NA	NA	NA	NA	NA
Endrin Lipid Adj (ng/g)	NA	WTSPO2YR	WTSB2YR	NA	NA	NA	NA	NA
Aldrin Lipid Adj (ng/g)	NA	WTSPO2YR	WTSB2YR	NA	NA	NA	NA	NA
p,p'-DDT Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
Hexachlorobenzene Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
2,4,5-trichlorophenol (ug/L)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	NA	NA
2,4,6-trichlorophenol (ug/L)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	NA	NA
Beta-hexachlorocyclohexane Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
Pentachlorophenol (ug/L)	NA	NA	WTSC2YR	NA	NA	NA	NA	NA
Dimethylphosphate (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Diethylthiophosphate (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Diethylphosphate (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
3,5,6-trichloropyridinol (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA
Dimethylthiophosphate (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
cis dichlorovnl-dimeth carboacid (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	NA	NA	NA	NA
trans dichlorovnl-dimeth carboacid (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA
3-phenoxybenzoic acid (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA

2,5-dichlorophenol (ug/L)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
2,4-dichlorophenol (ug/L)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Paranitrophenol (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA
2,4-D (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	NA	WTSC2YR	WTSC2YR	NA	NA
diethylaminomethylpyrimidinol/one (ug/L)	NA	WTSPP2YR	NA	NA	NA	NA	NA	NA
Oxypyrimidine (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA
Dimethyldithiophosphate (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
DEET acid (ug/L)	NA	NA	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	WTSC2YR
Desethyl hydroxy DEET (ug/L)	NA	NA	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	WTSC2YR
DEET (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	WTSC2YR
Alachor mercapturate (ug/L)	WTSPP2YR	NA	NA	NA	NA	NA	NA	NA
Malathion diacid (ug/L)	WTSPP2YR	NA	NA	NA	WTSC2YR	WTSC2YR	NA	NA
O-Phenyl phenol (ug/L)	NA	NA	WTSC2YR	WTSB2YR	WTSB2YR	WTSB2YR	NA	NA
Ethylenethio urea (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
MHNCH (ng/mL)	NA	NA	NA	NA	NA	NA	WTSA2YR	WTSB2YR
Mono-isononyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono(carboxynonyl) phthalate (ng/mL)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-2-hydroxy-iso-butyl phthlate (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Mono(carboxyoctyl) phthalate (ng/mL)	NA	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-isobutyl phthalate (ng/mL)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-(2-ethyl)-hexyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-(2-ethyl-5-oxohexyl) phthalate (ng/mL)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-benzyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-(2-ethyl-5-hydroxyhexyl) phthalate (ng/mL)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-2-ethyl-5-carboxypentyl phthalate (ng/mL)	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-n-butyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-ethyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR
Mono-n-methyl phthalate (ng/mL)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA
Mono-3-hydroxy-n-butyl phthalate (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Mono-(3-carboxypropyl) phthalate (ng/mL)	NA	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSB2YR

Mono-cyclohexyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	NA	NA
o-Desmethylangolensin (O-DMA) (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA	NA
Enterolactone (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA	NA
Enterodiol (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA	NA
3-phenanthrene (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA
1-phenanthrene (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSA2YR
2-phenanthrene (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA
4-phenanthrene (ng/L)	NA	NA	WTSB2YR	WTSB2YR	NA	NA	WTSA2YR	NA
2 & 3-Hydroxyphenanthrene (ng/L)	NA	NA	NA	NA	NA	NA	NA	WTSA2YR
1-pyrene (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSA2YR
1-napthol (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSA2YR
2-napthol (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSA2YR
9-fluorene (ng/L)	NA	NA	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA
3-fluorene (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSA2YR
2-fluorene (ng/L)	NA	WTSPH2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	WTSA2YR
3-fluoranthene (ng/L)	WTSPH4YR	WTSPH4YR	NA	NA	NA	NA	NA	NA
PCB199 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB172 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB206 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB180 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB209 Lipid Adj (ng/g)	NA	NA	WTSC2YR	NA	NA	NA	NA	NA
PCB189 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB146 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB156 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB170 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB194 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB157 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB195 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB187 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB153 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA

PCB167 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB196 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB138 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB177 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
3,3',4,4',5,5'-hxcb Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB183 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB118 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB128 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB99 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB105 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB178 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
3,3',4,4',5-pcnb Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB74 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
3,4,4',5-tcb Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB28 Lipid Adj (ng/g)	WTSPO2YR	NA	WTSC2YR	NA	NA	NA	NA	NA
PCB52 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB149 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB151 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB101 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB87 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB66 Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
PCB110 Lipid Adj (ng/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
PCB49 Lipid Adj (ng/g)	NA	NA	WTSC2YR	NA	NA	NA	NA	NA
PCB44 Lipid Adj (ng/g)	NA	NA	WTSC2YR	NA	NA	NA	NA	NA
Perfluoroundecanoic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluorododecanoic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluorodecanoic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluorooctane sulfonamide (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	NA
2-(N-methyl-PFOSA) acetate (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluorononanoic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR

Linear perfluorooctane sulfonate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Monomethyl branched iso of PFOS (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Linear perfluorooctanoate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Br. iso of perfluorooctanoate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSB2YR
Perfluorohexane sulfonic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluorooctanoic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluorooctane sulfonic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Perfluoroheptanoic acid (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Cotinine (ng/mL)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR
NNAL , urine (ng/mL)	NA	NA	NA	NA	WTMEC2YR	WTMEC2YR	WTMEC2YR	NA
Blood Tetrachloroethene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood o-Xylene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood m-/p-Xylene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Benzene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood 1,4-Dichlorobenzene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Bromodichloromethane (pg/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
N-ace-S-(phenl-2-hydxyetl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Ace-S-(2-Hydroxyethyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Phenylglyoxylic acid (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood Dibromochloromethane (pg/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
N-Ace-S-(2-hydroxypropyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-A-S-(4-hydrxy-2butn-l-yl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSVOC2Y	WTSA2YR
Blood 2,5-Dimethylfuran (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
3-methipurc acd & 4-methipurc acd (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Acetyl-S-(benzyl)-L-cysteine (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Acetyl-S-(n-propyl)-L-cysteine (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-ac-S-(2-carbmo-2-hydxel)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Acetyl-S-(phenyl)-L-cysteine (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood MTBE (pg/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Mandelic acid (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR

N-Ace-S-(2-carbamoylethyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Ace-S-(N-methlcarbamoyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
2-amnothiazolne-4-carbxylic acid (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-acetyl-S-(2-cyanoethyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
2-thoxothazlidne-4-carbxylic acid (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
2-Methylhippuric acid (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Ac-S-(2-Hydrxy-3-butnyl)-L-Cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-A-S-(3-hydrxprpl-1-metl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Acetyl-S-(2-Carbxyethyl)-L-Cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Ace-S-(3-Hydroxypropyl)-L-Cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Ace-S-(3,4-Dihidxybutl)-L-Cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood Ethylbenzene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Styrene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA	NA
Blood Toluene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA	WTSVOC2Y
Blood Bromoform (pg/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Nitromethane (pg/mL)	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood furan (ng/ml)	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Chloroform (pg/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Urinary thiocyanate (ng/mL)	NA	WTMEC2YR	NA	WTMEC2YR	WTMEC2YR	WTSA2YR	WTSA2YR	WTSA2YR
Iodine, urine (ng/mL)	NA	WTUIO2YR	WTSC2YR	WTSC2YR	WTMEC2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary nitrate (ng/mL)	NA	WTMEC2YR	NA	WTMEC2YR	WTMEC2YR	WTSA2YR	WTSA2YR	WTSA2YR
Perchlorate, urine (ng/mL)	NA	WTMEC2YR	WTSC2YR	WTMEC2YR	WTMEC2YR	WTSA2YR	WTSA2YR	WTSA2YR
1,2,3,7,8,9-hxcdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,4,7,8,9-hpcdf Lipid Adj (pg/g)	NA	WTSPO2YR	WTSC2YR	NA	NA	NA	NA	NA
1,2,3,7,8-pncdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
2,3,7,8-tcdf Lipid Adj (pg/g)	WTSPO4YR	WTSPO4YR	WTSC2YR	NA	NA	NA	NA	NA
Beryllium, urine (ng/mL)	WTSHM2YR	WTSHM2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	NA	NA
Urinary Arsenic acid (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR
Urinary Trimethylarsine Oxide (µg/L)	NA	NA	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	WTSA2YR	NA
Mercury, ethyl (ug/L)	NA	NA	NA	NA	NA	NA	WTMEC2YR	WTSH2YR

o,p'-DDT Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
G-hexachlorocyclohexane Lipid Adj (ng/g)	WTSPO4YR	WTSPO4YR	WTSB2YR	NA	NA	NA	NA	NA
4-fluoro-3-phenoxybenzoic acid (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA
2,4,5-T (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	NA	WTSC2YR	WTSC2YR	NA	NA
Prosulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Carbofuranphenol (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	NA	NA	NA	NA	NA
chloro-hydro-meth-chromen-one/ol (ug/L)	NA	WTSPP2YR	NA	NA	NA	NA	NA	NA
Oxasulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Desethyl atrazine (ug/L)	NA	NA	NA	NA	WTSC2YR	NA	NA	NA
Desisopropyl atrazine mercapturate (ug/L)	NA	NA	NA	NA	WTSC2YR	NA	NA	NA
Diaminochloroatrazine (ug/L)	NA	NA	NA	NA	WTSC2YR	NA	NA	NA
Desisopropyl atrazine (ug/L)	NA	NA	NA	NA	WTSC2YR	NA	NA	NA
Atrazine mercapturate (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	NA	NA	NA
Diethyldithiophosphate (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Chlorsulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Sulfosulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
dibromovinyl-dimeth prop carboacid (ug/L)	WTSPP4YR	WTSPP4YR	NA	NA	WTSC2YR	WTSC2YR	NA	NA
Metolachlor mercapturate (ug/L)	NA	WTSPP2YR	NA	NA	NA	NA	NA	NA
Acetochlor mercapturate (ug/L)	NA	WTSPP2YR	NA	NA	NA	NA	NA	NA
2-isopropoxyphenol (ug/L)	WTSPP4YR	WTSPP4YR	WTSC2YR	NA	NA	NA	NA	NA
Methamidaphos (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
O-methoate (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Acephate (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Dimethoate (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Ethametsulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Metsulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Nicosulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Propylenethio urea (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Mono-n-octyl phthalate (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSB2YR	NA	NA
2-(N-ethyl-PFOSA) acetate (ng/mL)	WTMEC2YR	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	NA

Perfluorobutane sulfonic acid (ng/mL)	NA	NA	WTSA2YR	WTSA2YR	WTSC2YR	WTSC2YR	WTSA2YR	WTSB2YR
Blood Hexachloroethane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood Carbon Tetrachloride (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
N-Acetyl-S-(trichlorovinyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood 1,1,2,2-Tetrachloroethane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood 1,1,1-Trichloroethane (ng/mL)	WTSVOC4Y	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
N-Ace-S-(dimethylphenyl)-L-Cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood Nitrobenzene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
N-acel-S-(1,2-dichlorovinl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
N-Acel-S-(2,2-Dichlorvinyl)-L-cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood 1,2,3-trichloropropane (ng/ml)	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Trichloroethene (ng/ml)	WTSVOC4Y	WTSVOC4Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood 1,2-Dichlorobenzene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood 1,2-Dichloropropane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood 1,2-Dichloroethane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Methylene Chloride (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood 1,3-Dichlorobenzene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Chlorobenzene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
N-A-S-(1-HydrxMet)-2-Prpn)-L-Cys (ng/mL)	NA	NA	NA	WTSVOC2Y	NA	NA	WTSA2YR	WTSA2YR
Blood isopropylbenzene (ng/ml)	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood Dibromomethane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood Hexane (ng/mL)	NA	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood trans-1,2-Dichloroethene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood Heptane (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Octane (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Cyclohexane (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Diethyl Ether (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Ethyl Acetate (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Chloroethane (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Methylcyclopentane (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y

Blood Tetrahydrofuran (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Sulfometuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Atrazine (ug/L)	NA	NA	NA	NA	WTSC2YR	NA	NA	NA
Triasulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Halosulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Triflusulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Primisulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Rimsulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Thifensulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Bensulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Foramsulfuron (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Mesosulfuron methyl (ug/L)	NA	NA	WTSC2YR	WTSC2YR	WTSC2YR	NA	NA	NA
Blood 1,1,1,2-tetrachloroethane (ng/mL)	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood 1,1,2-Trichloroethane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood 1,1-Dichloroethane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood 1,1-Dichloroethene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood cis-1,2-Dichloroethene (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA
Blood 1,2-Dibromo-3-chloropropane (ng/mL)	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	NA	NA
Blood 1,2-dibromoethane (ng/ml)	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood 1,4-Dioxane (ng/mL)	NA	NA	NA	NA	NA	WTSVOC2Y	WTSVOC2Y	WTSVOC2Y
Blood aaa-Trifluorotoluene (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Blood Vinyl Bromide (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTSVOC2Y
Albumin, urine (ug/mL)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR
Albumin, urine (mg/L)	NA	NA	NA	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR
Creatinine, urine (mg/dL)	WTMEC4YR	WTMEC4YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR
Creatinine, urine (umol/L)	NA	NA	NA	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR	WTMEC2YR
Bis(2-chloroethyl) phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Bis(1-chloro-2-propyl) phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Bis(1,3-dichloro-2-propyl) phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Dibutyl phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR

Dibenzyl phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Di-o-cresyl phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Di-p-cresyl phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Diphenyl phosphate (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
2,3,4,5-tetrabromobenzoic acid (ug/L)	NA	NA	NA	NA	NA	NA	NA	WTSC2YR
Daidzein (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA	NA
Equol (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA	NA
Genistein (ng/mL)	WTSPH4YR	WTSPH4YR	WTSB2YR	WTSB2YR	WTSB2YR	WTSA2YR	NA	NA
Hydroxycotinine, Serum (ng/mL)	NA	NA	NA	NA	NA	NA	NA	WTMEC2YR

Table A2.4. NHANES codenames for survey weights used in children aged 3-11 years old for a given Per- and Polyfluoroalkyl Substance (PFAS) and NHANES cycle.

Chemical name	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8
Perfluoroundecanoic acid (ng/mL)								WTSS2YR
Perfluorododecanoic acid (ng/mL)								WTSS2YR
Perfluorodecanoic acid (ng/mL)								WTSS2YR
Perfluorooctane sulfonamide (ng/mL)								WTSS2YR
2-(N-methyl-PFOSA) acetate (ng/mL)								WTSS2YR
Perfluorononanoic acid (ng/mL)								WTSS2YR
Linear perfluorooctane sulfonate (ug/L)								WTSS2YR
Monomethyl branched iso of PFOS (ug/L)								WTSS2YR
Linear perfluorooctanoate (ug/L)								WTSS2YR
Br. iso of perfluorooctanoate (ug/L)								WTSS2YR
Perfluorohexane sulfonic acid (ng/mL)								WTSS2YR
Perfluoroheptanoic acid (ng/mL)								WTSS2YR
2-(N-ethyl-PFOSA) acetate (ng/mL)								WTSS2YR
Perfluorobutane sulfonic acid (ng/mL)								WTSS2YR

Table A2.5. Number of participants by race and menopause/hysterectomy status

	Menopause – No Regular Periods	Menopause – Yes	Menopause – No Irregular Periods
Mexican Americans	3338	1566	466
Non-Hispanic Blacks	3338	2070	313
Non-Hispanic Whites	4937	5122	611
Other Hispanics	1098	741	99
Other Race/Multi-Racial	1108	512	84

Table A2.6. Number of participants by race and number of pregnancies resulting in live births

Number of live births	Mexican Americans	Non-Hispanic Blacks	Non-Hispanic Whites	Other Hispanics	Other Race/Multi-Racial
0	81	164	199	31	22
1	575	803	1507	258	240
2	709	952	2510	384	317
3	717	667	1742	284	176
4	431	381	775	152	88
5	238	197	330	92	34
6	144	113	173	33	25
7	101	82	69	10	10
8	70	39	37	9	7
9	38	17	24	7	7
10	52	17	12	5	1
11	44	16	9	5	3
12	7	9	2	3	0
13	4	0	0	3	0
15	4	0	0	0	0
17	0	1	0	0	0

Table A2.7. Number of participants by breastfeeding status and race.

	Breastfed – No Do not have children	Breastfeed – Yes	Breastfed – No Have children
Mexican Americans	81	1925	821
Non-Hispanic Blacks	164	1223	1640
Non-Hispanic Whites	199	3516	2668
Other Hispanics	31	728	311
Other Race/Multi-Racial	22	454	187

Table A2.8. Number of participants by race and iron deficiency status.

	Iron deficient - No	Iron deficient – Yes
Mexican Americans	7100	812
Non-Hispanic Blacks	5991	1989
Non-Hispanic Whites	12187	911
Other Hispanics	2289	358
Other Race/Multi-Racial	2137	318

Appendix 3. Biomarker-Based Occupational Exposome

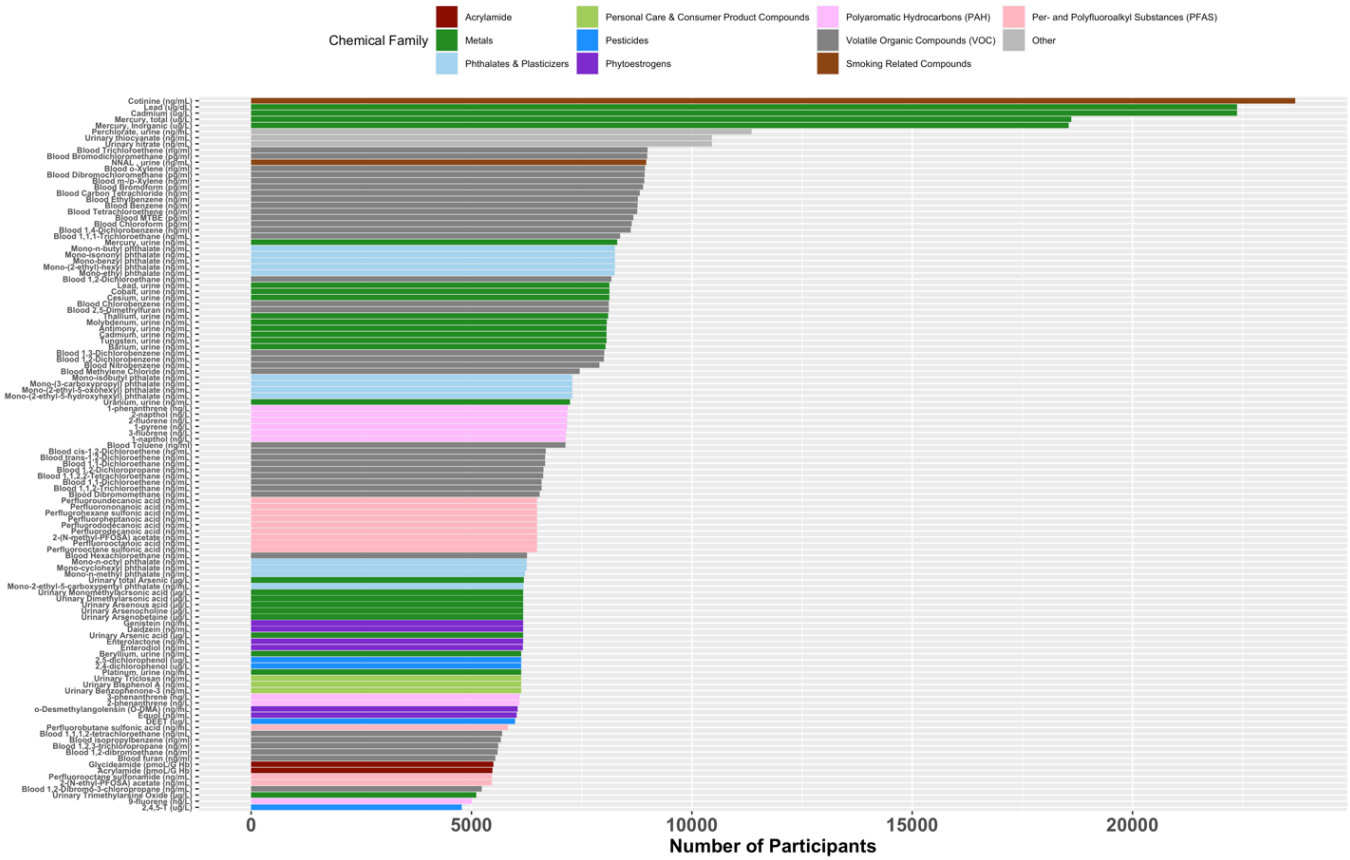

Figure A3.1. Bar plot showing number of participants by included chemical. Participants have data available for age, sex, race, industrial sector, collar category, and given chemical.

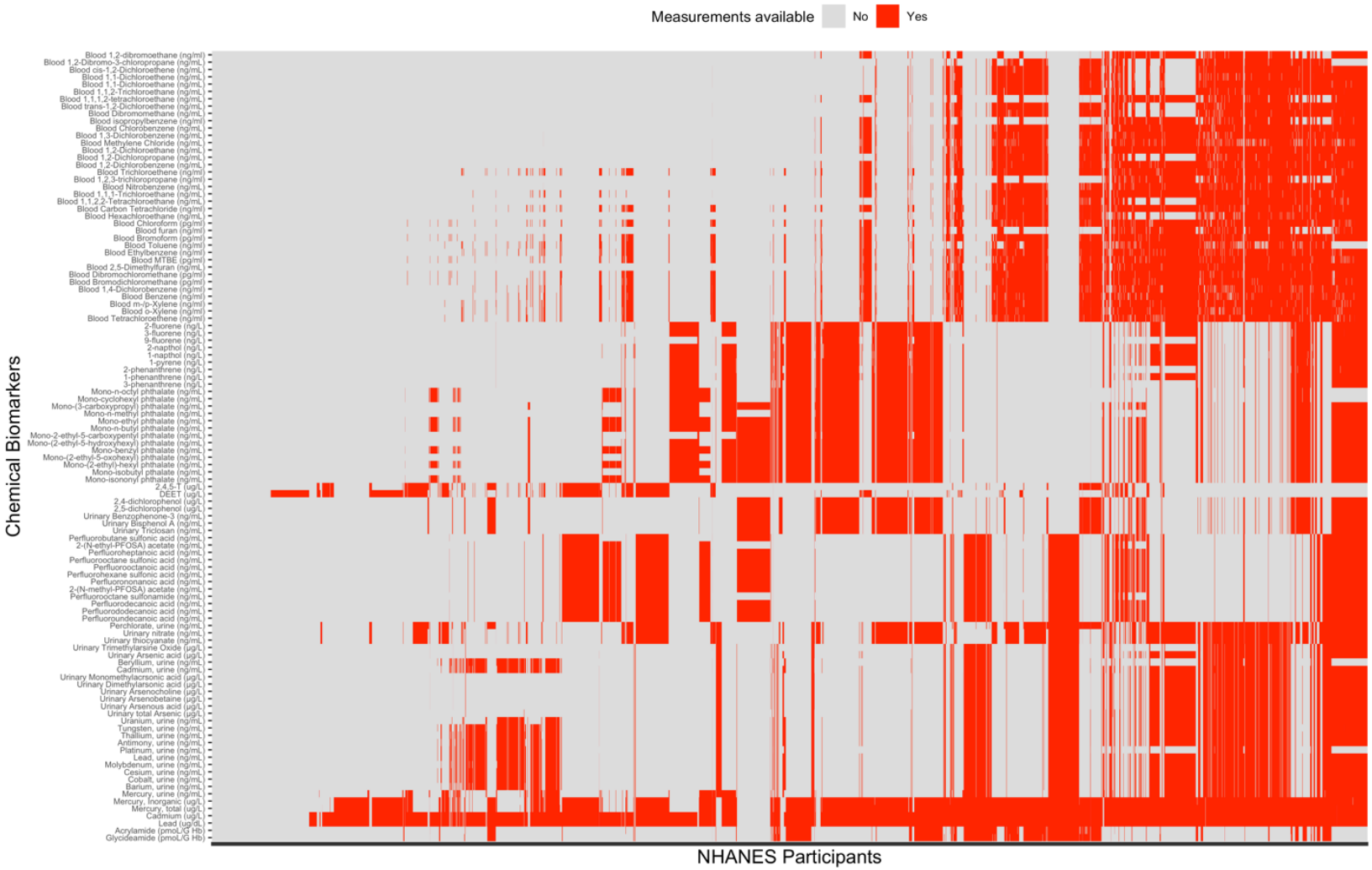

Figure A3.2. Heatmap of dichotomized biomarker measurements by participants and included chemicals to show sparsity of the chemical biomarker dataset. NHANES participants have data available for age, sex, race, industrial sector, and occupational title. Chemical biomarkers are grouped by chemical class. Participants are order by number of measured chemicals, which ranges from 0 to 98 chemicals, e.g. participants to the far-right were measured for 98 chemical biomarkers.

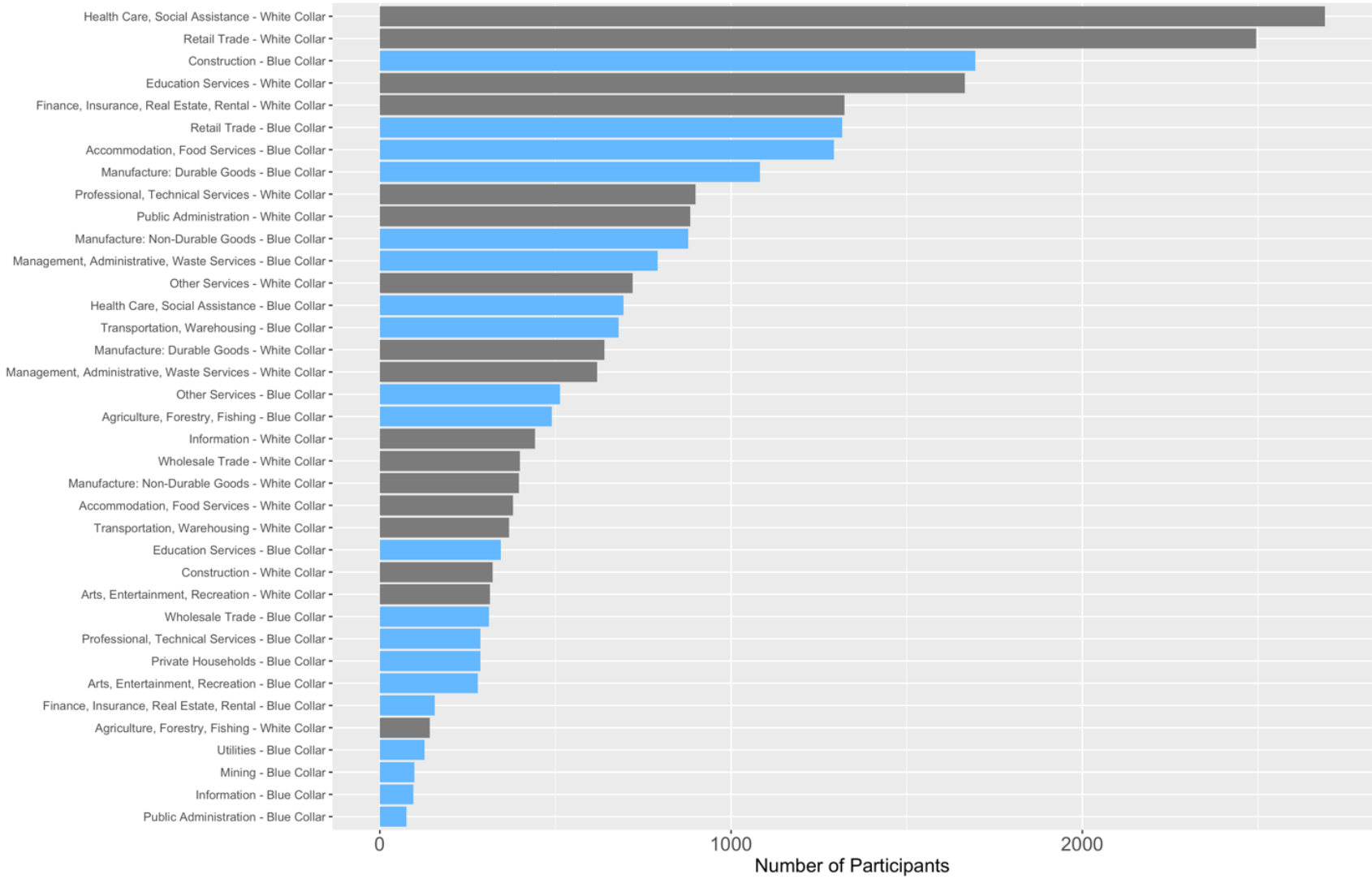

Figure A3.3. Bar plot showing the number of participants by each sector-collar combination. The sector-collar combinations are ordered from highest number of participants to lowest.

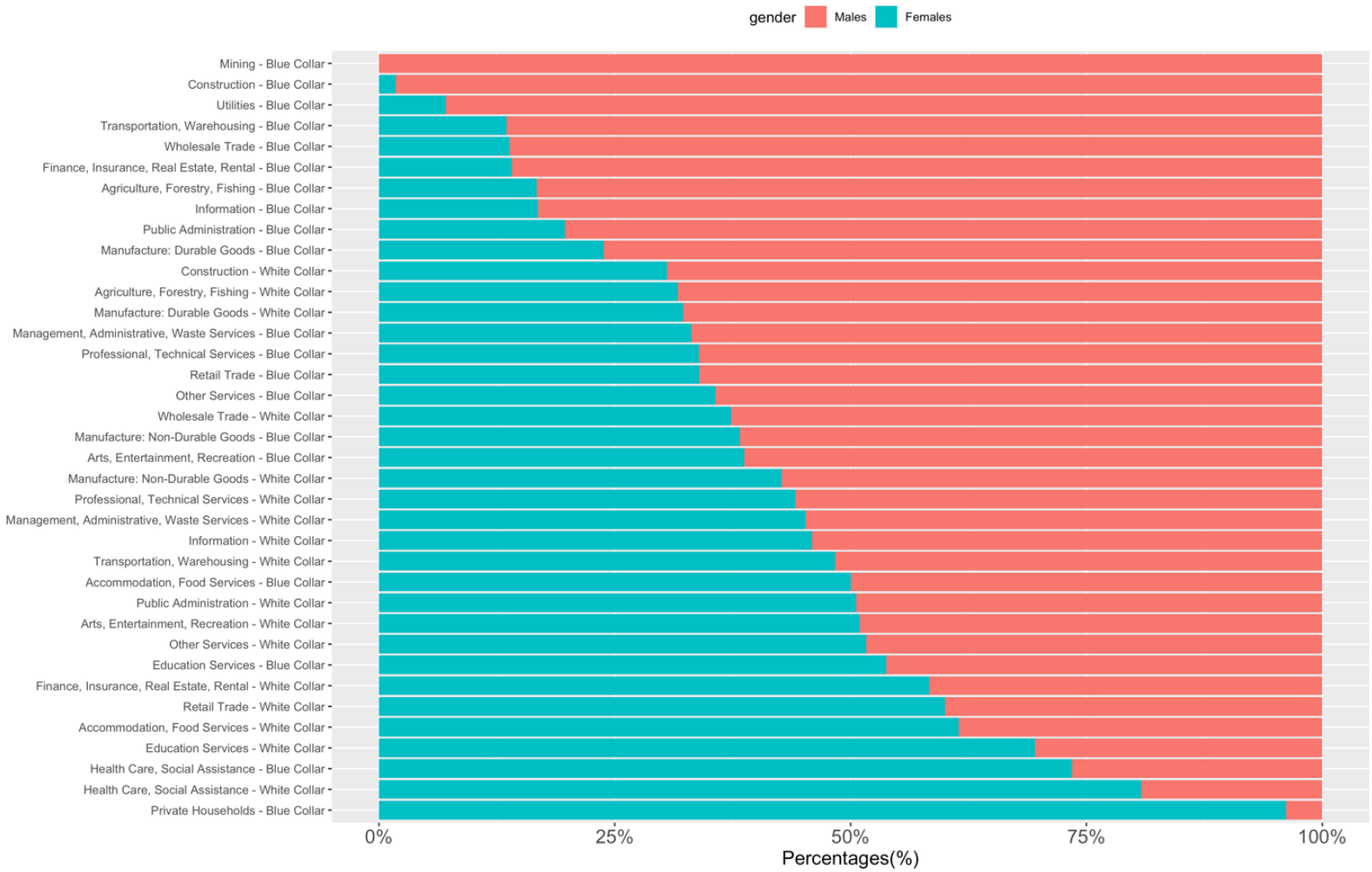

Figure A3.4. Bar plot showing the percentage of male versus female participants for each sector-collar combination. The sector-collar combinations are ordered from highest percentage of males to lowest percentage.

Figure A3.5. Bar plot showing the percentage of Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, and Other Race/Multi-Racial participants for each sector-collar combination. The sector-collar combinations are ordered from highest percentage of Non-Hispanic White participants to lowest.

Figure A3.6. Bar plot showing the percentage of poverty income categories for each sector-collar combination. The sector-collar combinations are ordered from highest percentages of the lowest PIR category ([0,1] as shown in red) to the lowest percentages.

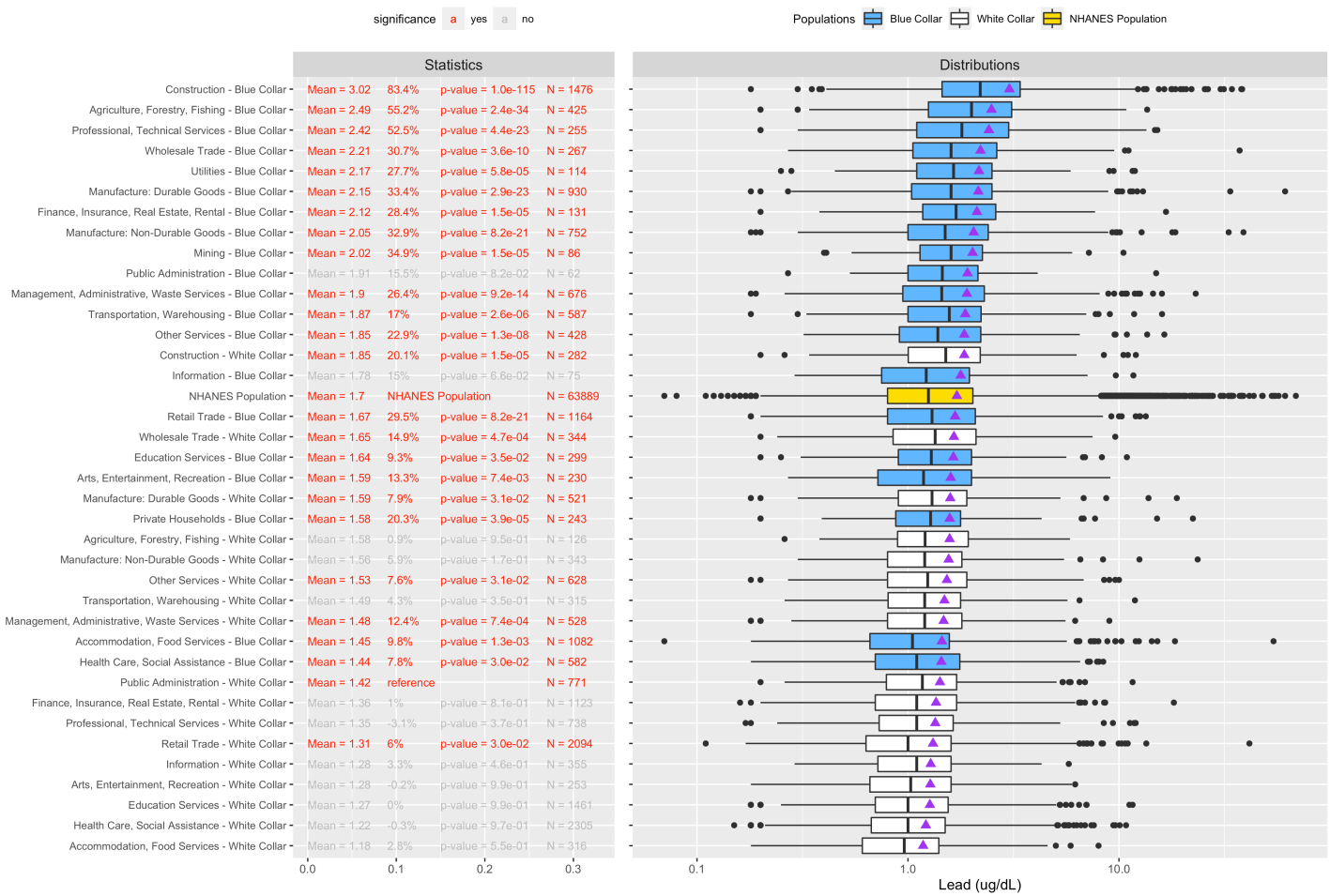

Figure A3.7. Box plot of distribution of lead in blood. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for lead, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of total lead for a given sector-collar combination. Results are adjusted for age, sex, and race.

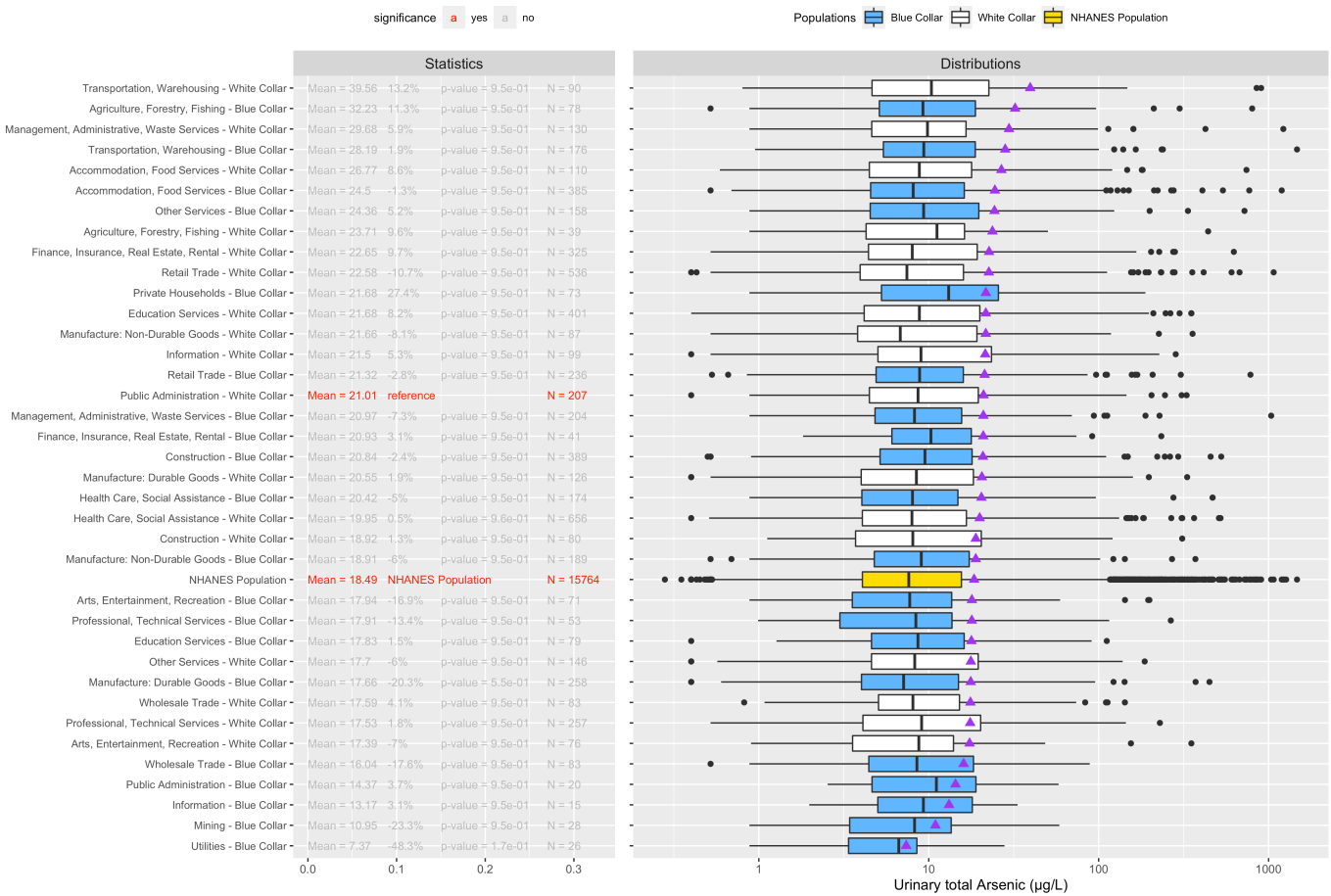

Figure A3.8. Box plot of distribution of total arsenic in urine. Far-left statistics are the mean chemical biomarker concentration. The middle-left statistics are the percent differences except for the "reference" group of "Public Administration – White Collars" and the "NHANES population". The NHANES population includes all participants with measurements for total arsenic, including the sector-collar combinations. The middle-right statistics are the p-values corrected for multiple comparison with the Benjamini and Hochberg FDR procedure of 5%. Far-right statistics are the sample size of each sector-collar combinations. Purple triangle represents the mean concentration of total arsenic for a given sector-collar combination. Results are adjusted for age, sex, and race.

Table A3.1. Job occupation description with corresponding collar category.

Job Occupation Description	Collar Category
Building & Grounds Cleaning, Maintenance Occupations	Blue-collar
Food Preparation, Serving Occupations	Blue-collar
Personal Care, Service Occupations	Blue-collar
Farming, Fishing, Forestry Occupations	Blue-collar
Installation, Maintenance, Repair Occupations	Blue-collar
Construction, Extraction Occupations	Blue-collar
Production Occupations	Blue-collar
Transportation, Material Moving Occupations	Blue-collar
Armed Forces	Blue-collar
Management Occupations	White-collar
STEM, Social, and Legal Services Occupations	White-collar
Healthcare Practitioner, Technical Occupations	White-collar
Education, Training, Library Occupations	White-collar
Arts, Design, Entertainment, Sports, Media Occupations	White-collar
Healthcare Support Occupations	White-collar
Sales & Related Occupations	White-collar
Office, Administrative Support Occupations	White-collar
Protective Service Occupations	White-collar

Table A3.2. List of excluded tobacco and soy metabolites and their corresponding chemical class.

Chemical	Chemical Class
Cotinine (ng/mL)	Smoking Related Compounds
NNAL , urine (ng/mL)	Smoking Related Compounds
o-Desmethylangolensin (O-DMA) (ng/mL)	Soy Metabolites
Enterolactone (ng/mL)	Soy Metabolites
Enterodiol (ng/mL)	Soy Metabolites
Daidzein (ng/mL)	Soy Metabolites
Equol (ng/mL)	Soy Metabolites
Genistein (ng/mL)	Soy Metabolites

Table A3.3. Sample size required to detect significant differences. x% power mean that a study has a x% chance of ending up with a p-value of less than 5% in a statistical test when the effect size (or regression coefficient) is a corresponding value if the sample size is a particular number. For example, a study has a 70% chance of detecting a significance difference with a regression coefficient of 0.37 when the sample size is 90 participants.

Effect Size (or Regression Coefficients)	90% Power	80% Power	70% Power
0.01	209952	156800	123008
0.02	52488	39200	30752
0.03	23328	17422	13668
0.04	13122	9800	7688
0.05	8398	6272	4920
0.06	5832	4356	3417
0.07	4285	3200	2510
0.08	3281	2450	1922
0.09	2592	1936	1519
0.1	2100	1568	1230
0.11	1735	1296	1017
0.2	525	392	308
0.3	233	174	137
0.31	218	163	128
0.32	205	153	120
0.33	193	144	113
0.34	182	136	106
0.35	171	128	100
0.36	162	121	95
0.37	153	115	90
0.38	145	109	85
0.39	138	103	81
0.4	131	98	77
0.5	84	63	49

0.6	58	44	34
0.7	43	32	25
0.8	33	25	19
0.9	26	19	15
1	21	16	12

Table A3.4. Table of description, clinical thresholds, and sample size of each physiological indicator. Lower bound of the threshold implies that values below the threshold are unfavorable or indicative of high risk. Upper bound of the threshold implies that values above the threshold are unfavorable or indicative of high risk.

			Clinical Thresholds					
	Sample Size	Description	Lower - All	Upper - All	Lower - Males	Upper - Males	Lower - Females	Upper - Females
Body Composition								
Body Mass Index (kg/m**2)	23834	Measure of body fat based on height and weight. This measure is used to screen for weight categories associated with health problems. Lower measurements indicate underweighted, whereas higher measurements indicate overweight or obesity.	18.5-19	24.9-25				
Standing Height (cm)	23931	Distance between the lowest and highest points of a person standing upright.						
Subscapular Skinfold (mm)	14381	Subscapular skinfold is skinfold thickness measured under the lowest point of the shoulder blade. It also serves as a measure for subcutaneous fat.						
Triceps Skinfold (mm)	16305	Triceps skinfold is skinfold thickness measured back side middle upper arm. It also serves as a measure for subcutaneous fat.						
Waist Circumference (cm)	23255	A measurement taken around the abdomen at the level of the umbilicus (belly button). This measure is used to screen for patients with weight-related health problems. Higher measurements may indicate higher risk for heart disease and type 2 diabetes.				90-102		80-88
Weight (kg)	23858	Quantifier of heaviness. It serves as a nutrition screen tool and is valuable in						

		monitoring fluid balance and calculating medication doses.						
Relative Fat Mass Index (-)	23240	The relative fat mass formula is a new index for measuring body fatness that can be easily accessible to health practitioners trying to treat overweight patients who often face serious health consequences like diabetes, high blood pressure and heart disease. Higher index indicates higher body fat, which is associated with increased risk of poor health and early mortality.			8-13	19-25	21-24	33-36
Cardiovascular System								
60 sec. pulse (30 sec. pulse * 2):	23123	Normal resting heart rate is the number of heart beats per minute. It is calculated by multiplying the number of beats in 30 seconds by 2. Higher pulse may indicate lack of arterial blood flow or heart disease.	50-60	80-100				
Diastolic: Average blood pressure (mm Hg)	22990	A measure of the force exerted by blood against artery walls during contraction or beating of the heart. Higher measurements indicate hypertension, while lower measurements indicate hypotension.	60	80-90				
Systolic: Average blood pressure (mm Hg)	23045	A measure of the force exerted by blood when heart relaxes between beats. Higher measurements indicate hypertension, while lower measurements indicate hypotension.	90	130-140				
Direct HDL-Cholesterol (mg/dL)	22772	High-density lipoprotein (HDL) is often called “good” cholesterol. It removes cholesterol from the bloodstream. Higher levels of HDL are associated with lower risk of heart disease.			40		50	
LDL-cholesterol (mg/dL)	10696	Low-density lipoprotein (LDL) is often called “bad” cholesterol. It delivers cholesterol to the body. Higher levels are		100-130				

		associated with increased risk in heart disease.						
Ratio of LDL to HDL Cholesterol (-)	10696	This ratio is calculated by dividing the LDL cholesterol number by the HDL cholesterol number. Higher ratios are indicative of higher risk of heart disease.		1.4				
Ratio of Total to HDL Cholesterol (-)	22771	This ratio is calculated by dividing the total cholesterol number by the HDL cholesterol number. Higher ratios are indicative of higher risk of heart disease.				5		4.4
Total cholesterol (mg/dL)	22772	A measure of the total amount of cholesterol in your blood. It includes both low-density lipoprotein (LDL) cholesterol and high-density lipoprotein (HDL) cholesterol. Higher measurements are associated with increased risk in heart diseases.		200				
Triglycerides (mg/dL)	22722	Triglycerides are another type of fat in your blood. When you eat more calories than your body can use, it turns the extra calories into triglycerides. Higher measurements are associated with increased risk in heart diseases.		150				
Immune System								
C-reactive protein (mg/dL)	17059	C-reactive protein (CRP) is a substance produced by the liver in response to inflammation. A high level of CRP in the blood is a marker of inflammation. It can be caused by a wide variety of conditions, from infection to cancer. Higher measurements are indicative of inflammation.		0.1-0.2				
White blood cell count (1000 cells/uL)	22977	Cells of the immune system that are involved in protecting the body against infection and foreign invaders. The count is	4-15	10-11				

		the number of white blood cells in a sample of blood to serve as a general evaluation of health. A higher count may indicate inflammation, an infection, or disease of bone marrow.						
Metabolic System								
Alkaline phosphatase (U/L)	19819	Alkaline phosphatase (ALP) is an enzyme found in several tissues throughout the body. The highest concentrations of ALP are present in the cells that comprise bone and the liver. Elevated levels of ALP in the blood are most commonly caused by liver disease or bone disorders.	20-44	116-147				
Glucose, plasma (mg/dL)	11202	Fasting blood sugar provides vital clues about how the body is managing blood sugar levels. High levels of fasting blood sugar suggest that the body has been unable to lower the levels of sugar in the blood. This points to either insulin resistance or inadequate insulin production, and in some cases, both.	60-70	99-126				
Glycohemoglobin:(%)	22957	A glycohemoglobin test, or hemoglobin A1c, is a blood test that checks the amount of sugar (glucose) bound to the hemoglobin in the red blood cells. The level of glycohemoglobin is increased in the red blood cells of persons with poorly controlled diabetes mellitus.		5.7-7				
Homeostatic Model Assessment of Insulin Resistance (-)	9579	HOMA-IR stands for Homeostatic Model Assessment of Insulin Resistance. The meaningful part of the acronym is “insulin resistance”. It marks for both the presence and extent of any insulin resistance. It is a way to reveal the dynamic between baseline (fasting) blood sugar and the responsive		1.4-2.9				

		hormone insulin. Higher indices indicate increase resistance to insulin.						
Ratio of Insulin to Glucose	9579	This measure is calculated by dividing the concentration of insulin in blood with the concentration of blood glucose. It serves as a measure of insulin resistance. Higher indices indicate increase resistance to insulin.		0.2-0.3				
Nephrology								
Albumin (g/dL)	22733	A protein made in the liver that helps keep fluid in the bloodstream, so it doesn't leak into other tissues. It also carries nutrients throughout the body. Low albumin levels can indicate liver or kidney damage, while higher measurements may indicate dehydration or diarrhea.	3.4-3.5	5.4-5.5				
Blood urea nitrogen (mg/dL)	22732	This test measures the amount of nitrogen in the patient's blood that comes from the waste product urea. Urea is made when proteins break down in the body. Urea is made in the liver and eliminated in the urine. When kidneys are not healthy, they have trouble removing blood urea nitrogen, leading to elevated concentrations in the blood. Lower levels may indicate a diet low in proteins, malnutrition, or liver damage.	6-10	20-21				
Creatinine (mg/dL)	19821	Creatinine is a chemical waste product of creatine, an amino acid made by the liver and stored in the liver. Creatinine is the result of normal muscle metabolism. Low creatinine levels may indicate muscle disease, liver disease, or excessive water loss. Higher levels. May indicate kidney damage, kidney infection, reduced circulation to the kidneys, and dehydration.			0.6-0.9	1.2-1.3	0.5-0.6	1-1.1

Estimated Glomerular Filtration Rate (mL/min/1.73 m^2)	19821	An estimated GFR (eGFR) calculated from serum creatinine using an isotope dilution mass spectrometry (IDMS) traceable equation is a simple and effective way to detect chronic kidney disease. Lower rates indicate kidney dysfunction.	60-90					

Table A3.5. Number of participants with a given number of measured chemicals. For example, 297 participants have measurements available for 98 chemical biomarkers.

Number of measured chemicals	Number of participants
0	1332
1	869
2	227
3	69
4	1025
5	788
6	68
7	538
8	24
9	242
10	199
11	78
12	289
13	684
14	643
15	133
16	303
17	374
18	885
19	81
20	530
21	967
22	693
23	296
24	213

25	1100
26	65
27	248
28	689
29	23
30	135
31	939
32	198
33	89
34	743
35	95
36	700
37	152
38	280
39	934
40	445
41	523
42	743
43	74
44	185
45	259
46	5
47	40
48	14
49	31
50	37
51	77
52	81

53	92
54	126
55	165
56	327
57	310
58	817
59	114
60	136
61	180
62	226
63	447
64	540
65	509
66	100
67	168
68	184
69	68
70	70
71	129
72	27
73	64
74	113
78	1
85	1
86	1
88	3
89	3
90	4

91	18
92	16
93	15
94	11
95	47
96	175
97	228
98	297

Table A3.6. Description of each hierarchical clustering method.

Linkage Method	Description
Single (nearest neighbor)	Distance between two clusters is determined by a pair of elements that are closest to each other. At each step, if a pair of elements, with the shortest distance, do not belong in the same cluster, then the two clusters will be combined.
Complete (farthest neighbor)	Distance between two clusters is determined by a pair of elements that are farthest from each other. At each step, if a pair of elements, with the shortest distance, do not belong in the same cluster, then the two clusters will be combined.
Average (unweighted pair group method with arithmetic mean or UPGMA)	Distance between two clusters is determined by the average of all distances between each pairs of distance in either clusters.
McQuitty (Weighted Pair Group Method with Arithmetic Mean or WPGMA)	Distance of two nearest clusters (*i* and *j*) to another cluster (*k*) is the arithmetic mean of average distances between elements of *k* and *i* along with *k* and *j*.

Table A3.7. Cophenetic correlation coefficients by linkage methods for clustering of the sector-collar combinations based on chemical exposure profiles.

Linkage Method	Cophenetic Correlation Coefficients
Single (nearest neighbor)	0.2828052
Complete (farthest neighbor)	0.5536748
Average (unweighted pair group method with arithmetic mean or UPGMA)	0.5671026
McQuitty (Weighted Pair Group Method with Arithmetic Mean or WPGMA)	0.5626430

Table A3.8. Cophenetic correlation coefficients by linkage methods for clustering of the sector-collar combinations based on physiological response profiles.

Linkage Method	Cophenetic Correlation Coefficients
Single (nearest neighbor)	0.1859953
Complete (farthest neighbor)	0.644285
Average (unweighted pair group method with arithmetic mean or UPGMA)	0.6781713
McQuitty (Weighted Pair Group Method with Arithmetic Mean or WPGMA)	0.6678879

Appendix 4. Characterization of Linear and Non-linear Associations between Physiological Indicators and All-Cause Mortality

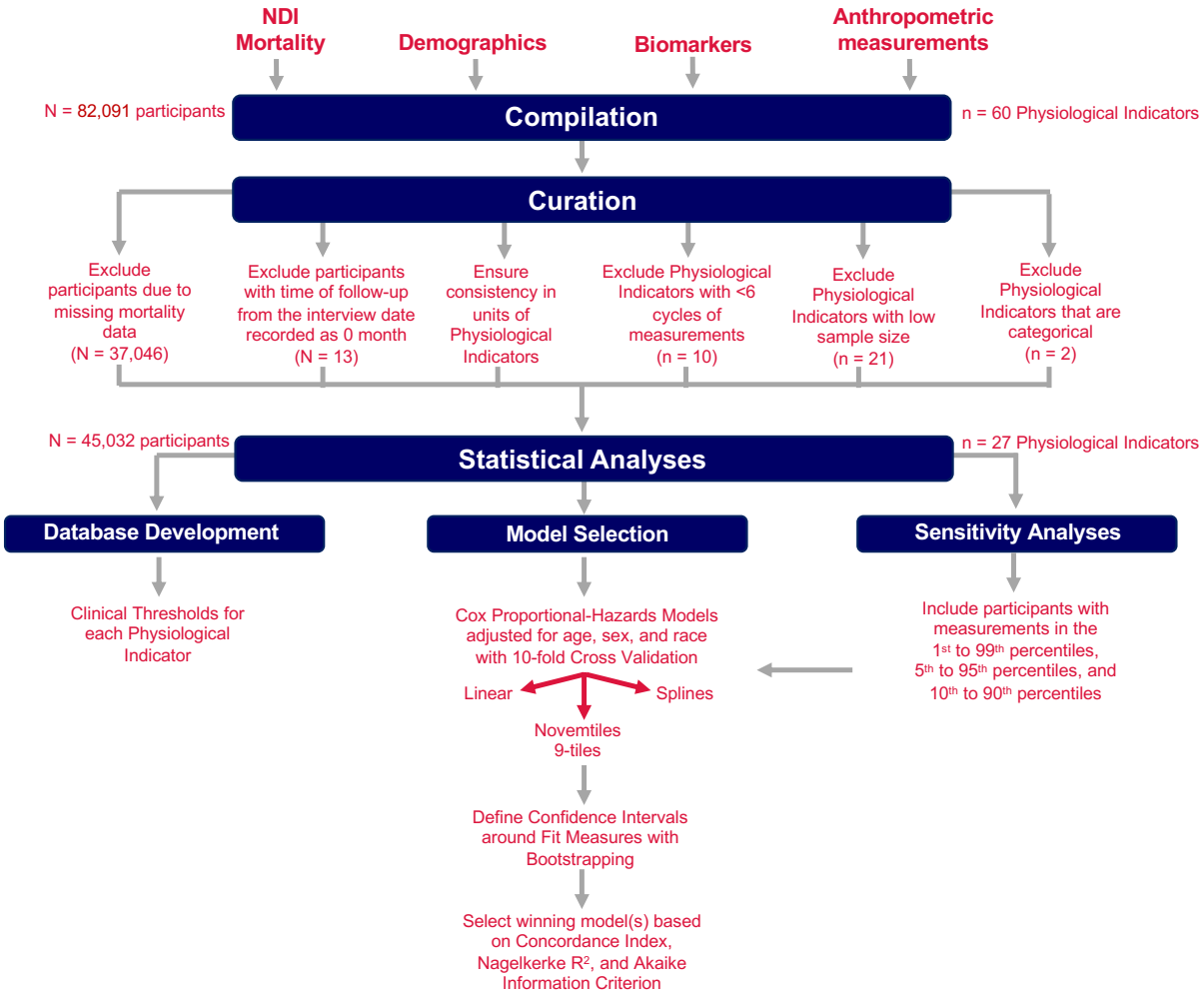

Figure A4.1. Schematic description of curation process and analytical methods. Schematic description of the process to curate the physiological measurements and of the analytical methods used to characterize associations between these measurements and mortality.

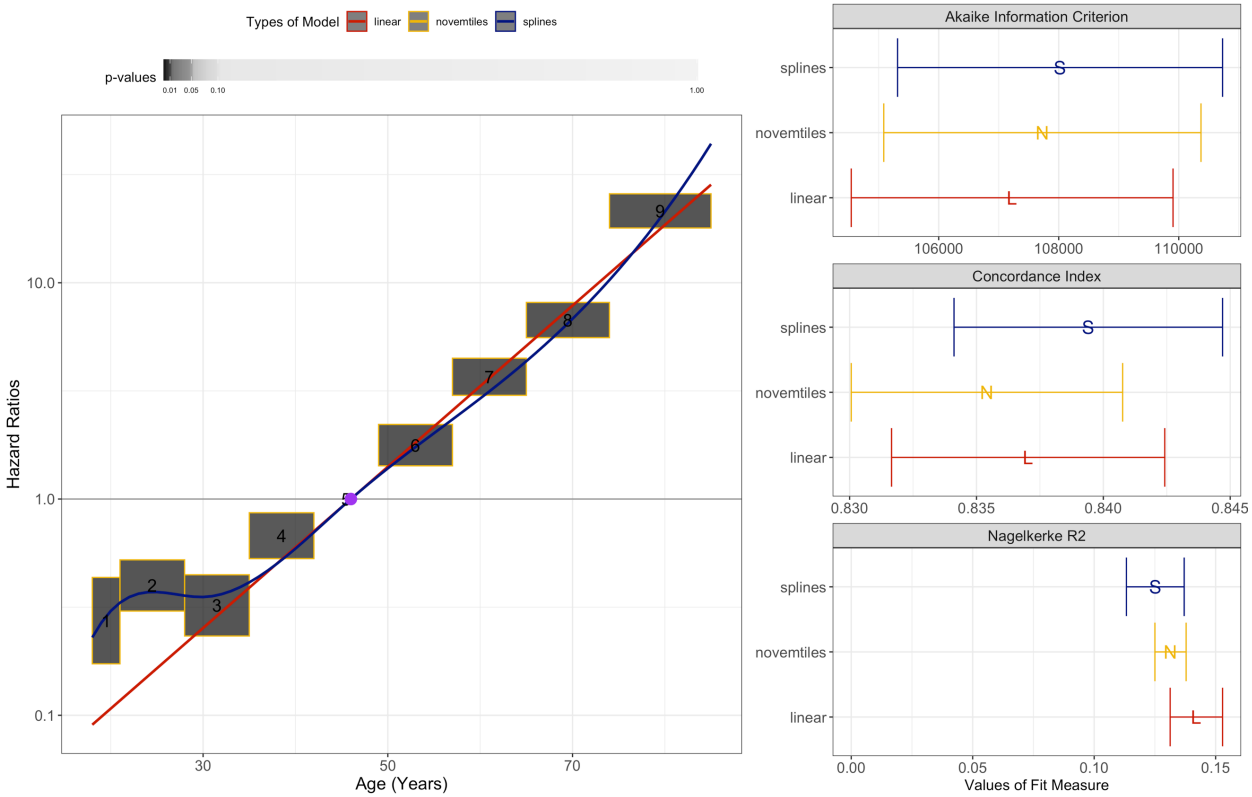

Figure A4.2. Stairway plot of hazard ratios displaying the relative mortality risk for age and alphabet soup plot of prediction performance for linear and non-linear models. Results were adjusted for sex and race/ethnicity.

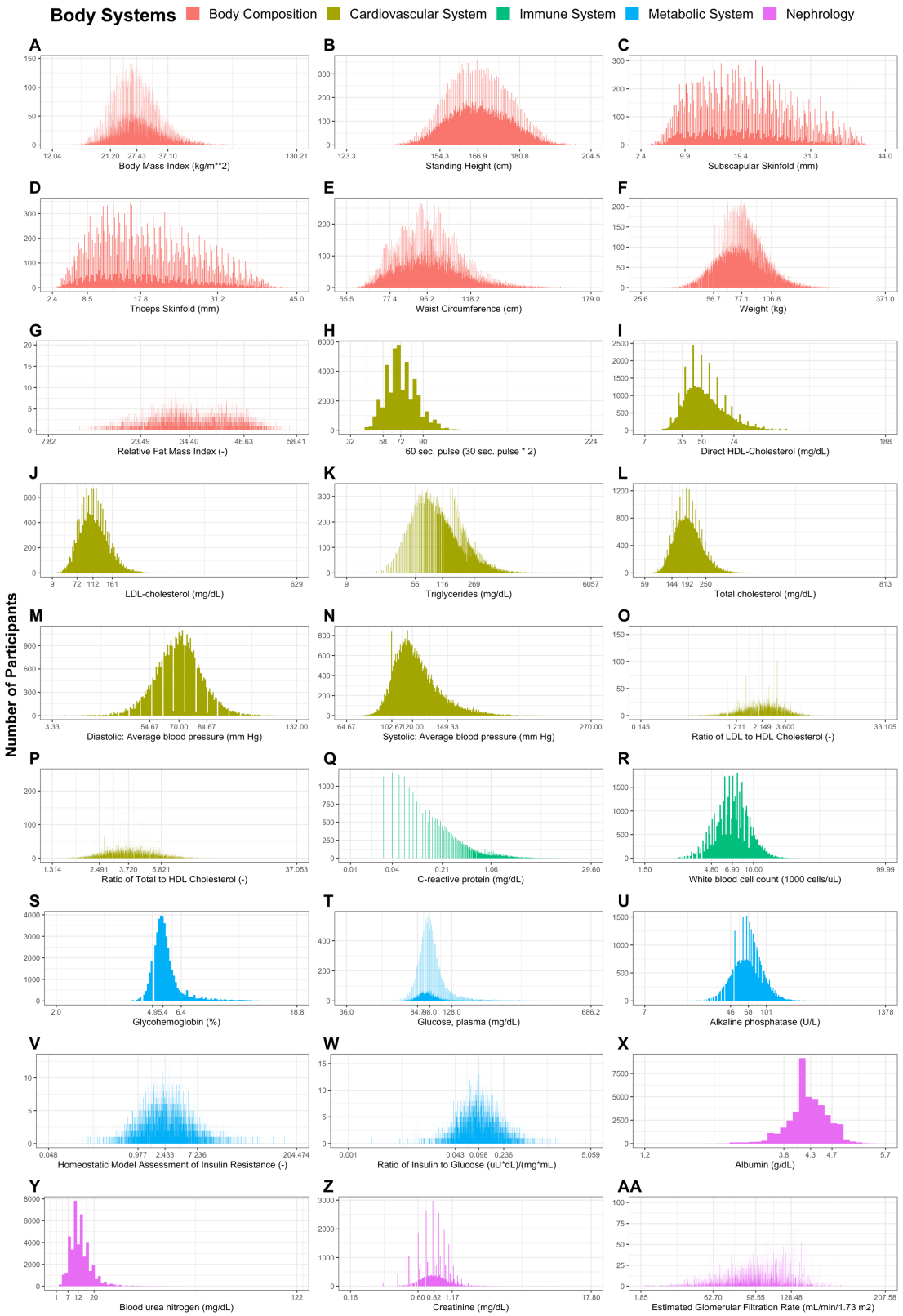

Figure A4.3. Histogram of measurements for each physiological indicator in all participants. Labels for tick marks are provided for the 0th, 10th, 50th, 90th, and 100th percentiles.

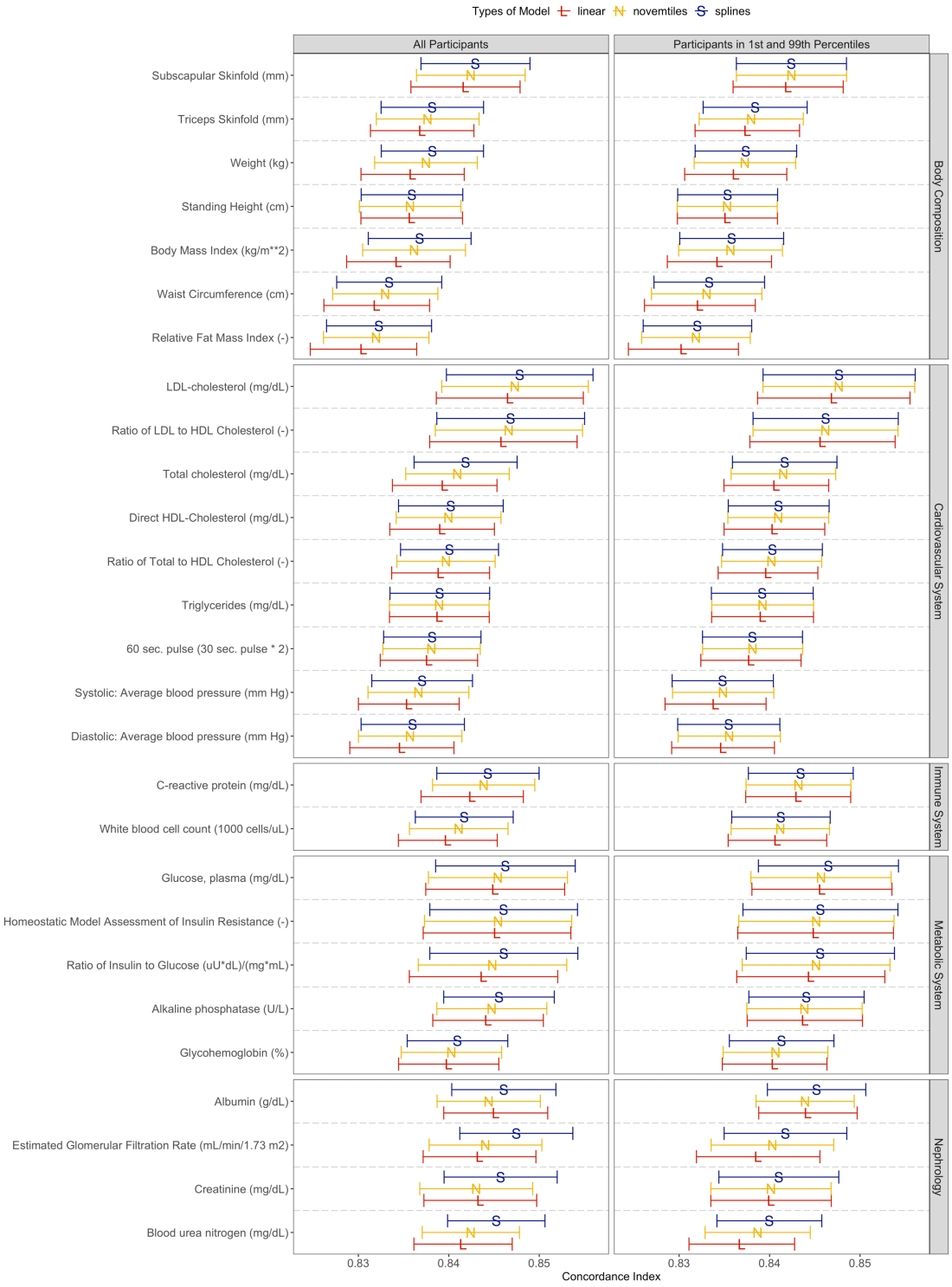

Figure A4.4. Alphabet soup plot displaying the Concordance Index for the associations with all-cause mortality for all physiological indicators across all studied models. Results are adjusted for age, sex, and race/ethnicity. Error bars represent the 95% Confidence Intervals defined through bootstrapping for 1000 replicates.

Figure A4.5. Alphabet soup plot displaying the AIC for the associations with all-cause mortality for all physiological indicators across all studied models. Results are adjusted for age, sex, and race/ethnicity. Error bars represent the 95% Confidence Intervals defined through bootstrapping for 1000 replicates.

Figure A4.6. Scatterplot of the sample size and prediction performance displayed for the A) AIC, B) Concordance Index, and C) Nagelkerke R2. Results are adjusted for age, sex, and race/ethnicity.

Figure A4.7. Alphabet soup plot of *Nagelkerke* R^2 on all participants (0) and participants within the 1st to 99th (1), 5th to 95th (2), and 10th to 90th percentiles (3). Results are adjusted for age, sex, and race/ethnicity. Error bars represent the 95% Confidence Intervals defined through bootstrapping for 1000 replicates.

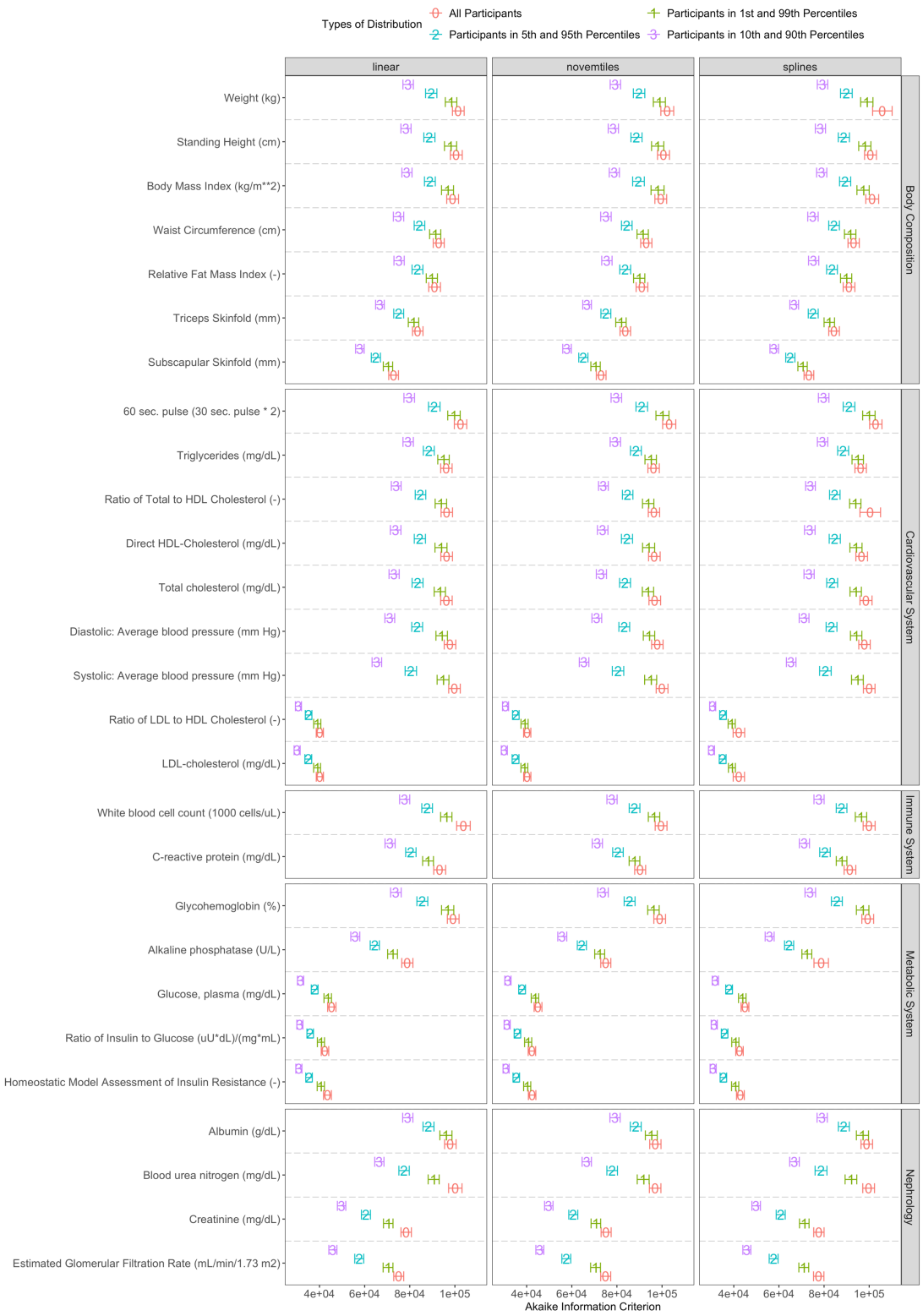

Figure A4.8. Alphabet soup plot of the *AIC* on all participants (0) and participants within the 1st to 99th (1), 5th to 95th (2), and 10th to 90th percentiles (3). Results are adjusted for age, sex, and race/ethnicity. Error bars represent the 95% Confidence Intervals defined through bootstrapping for 1000 replicates.

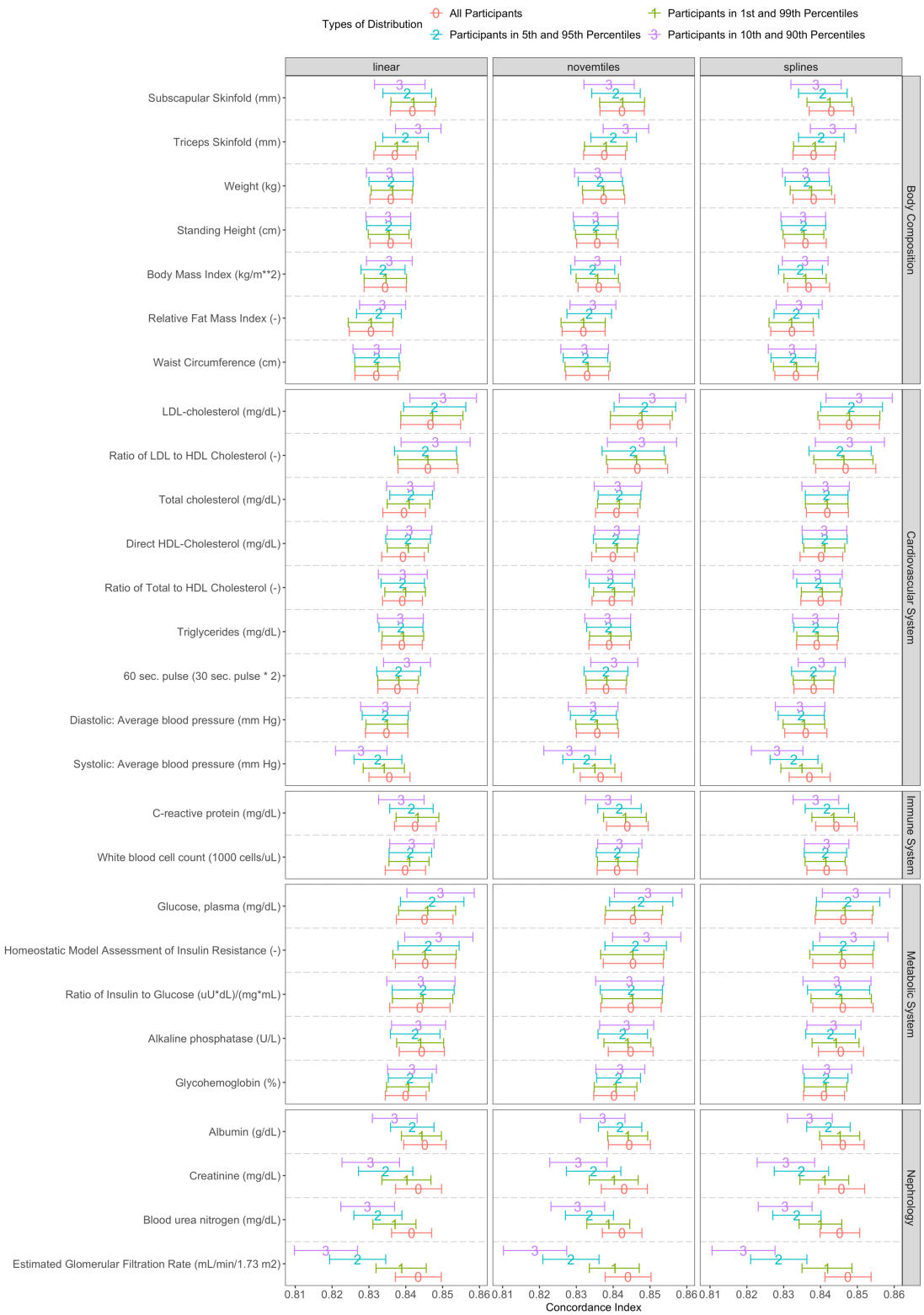

Figure A4.9. Alphabet soup plot of the *Concordance Index* on all participants (0) and participants within the 1st to 99th (1), 5th to 95th (2), and 10th to 90th percentiles (3). Results are adjusted for age, sex, and race/ethnicity. Error bars represent the 95% Confidence Intervals defined through bootstrapping for 1000 replicates.

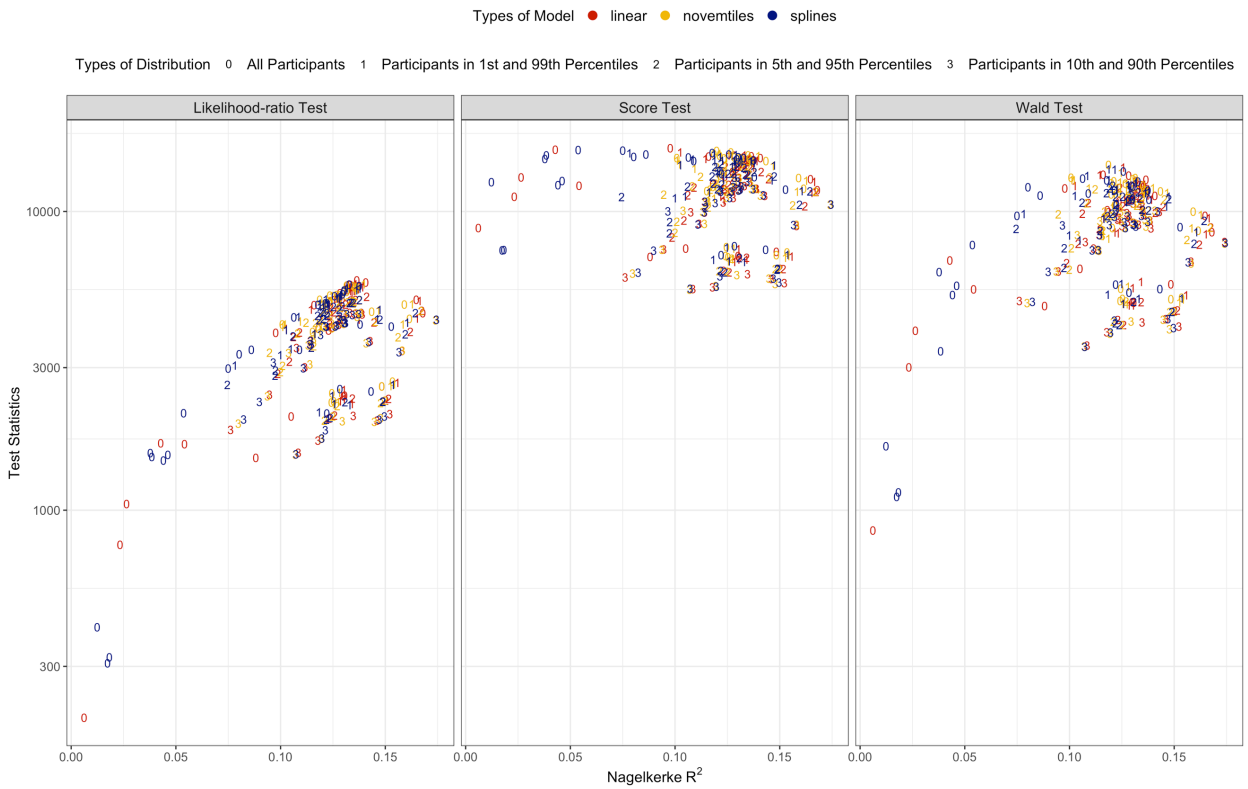

Figure A4.10. Volcano Plots of Nagelkerke R2 and test statistics used to indicate statistical significance of the model compared to a null model. Results are adjusted for age, sex, and race/ethnicity.

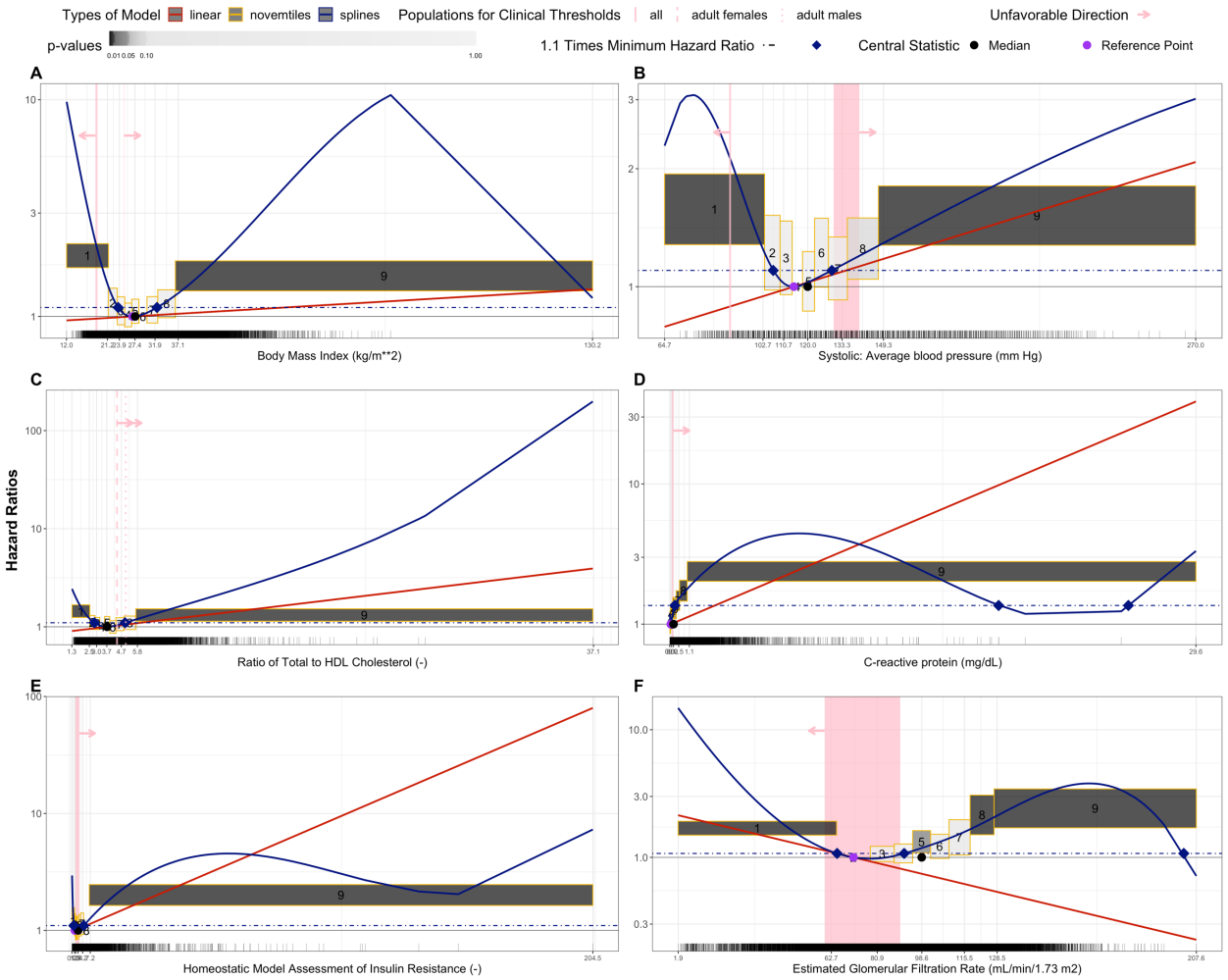

Figure A4.11. Stairway hazard ratios across all models to describe the relative mortality risk for A) Body Mass Index, B) Average Systolic Blood Pressure, C) Ratio of Total to HDL Cholesterol, D) C-Reactive Proteins, E) Homeostatic Model Assessment of Insulin Resistance, and F) Glomerular Filtration Rate when all participants are included. Relative risks for mortality from the novemtiles model are represented by the boxes with the width representing the range of a novemtile and the height representing the 95% Confidence Interval of the hazard ratio. The mean hazard ratio for each novemtile is presented by a digit. The hazard compares participants in a novemtile to those in the reference group at the 5th novemtile. The red and blue lines represent the relative mortality risk with respect to median of a physiological indicator for the linear and spline models, respectively. The dashed navy line represents when the hazard ratio is 10% higher than the minimum hazard ratio. The navy diamonds indicate the concentration at which the hazard ratio shows a 10% increase from the minimum hazard ratio. The purple dot represents the median for a physiological indicator. The pink lines and rectangles represent the values of the clinical thresholds with the width of the rectangles representing the ranges of the threshold. The set of tick marks along the base of the plot represent the distribution of a physiological indicator with increased opacity implying increased number of participants. Results were adjusted for age, sex, and race/ethnicity.

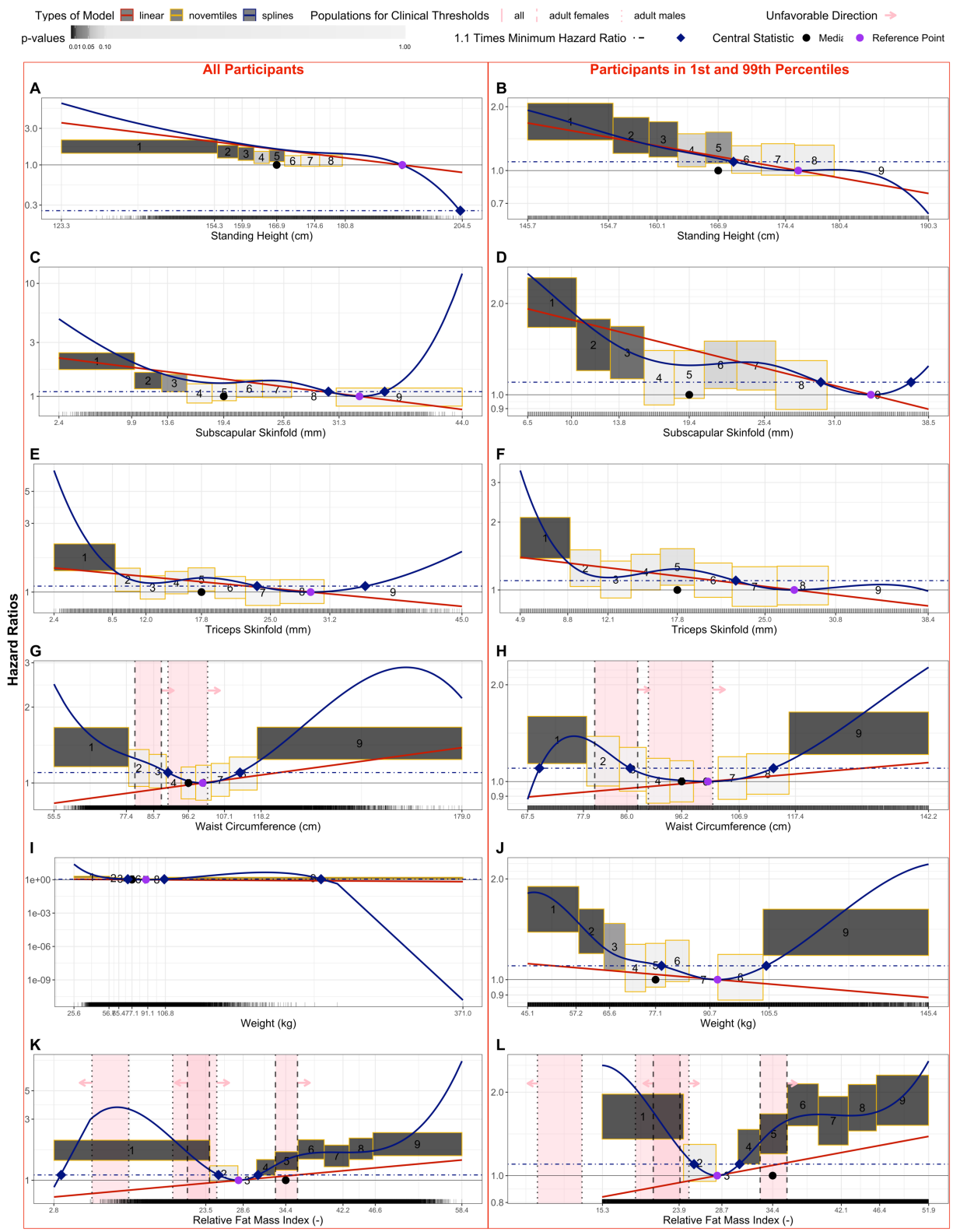

Figure A4.7. Stairway plots of hazard ratios across all models to describe the relative mortality risk for physiological indicators of body composition.

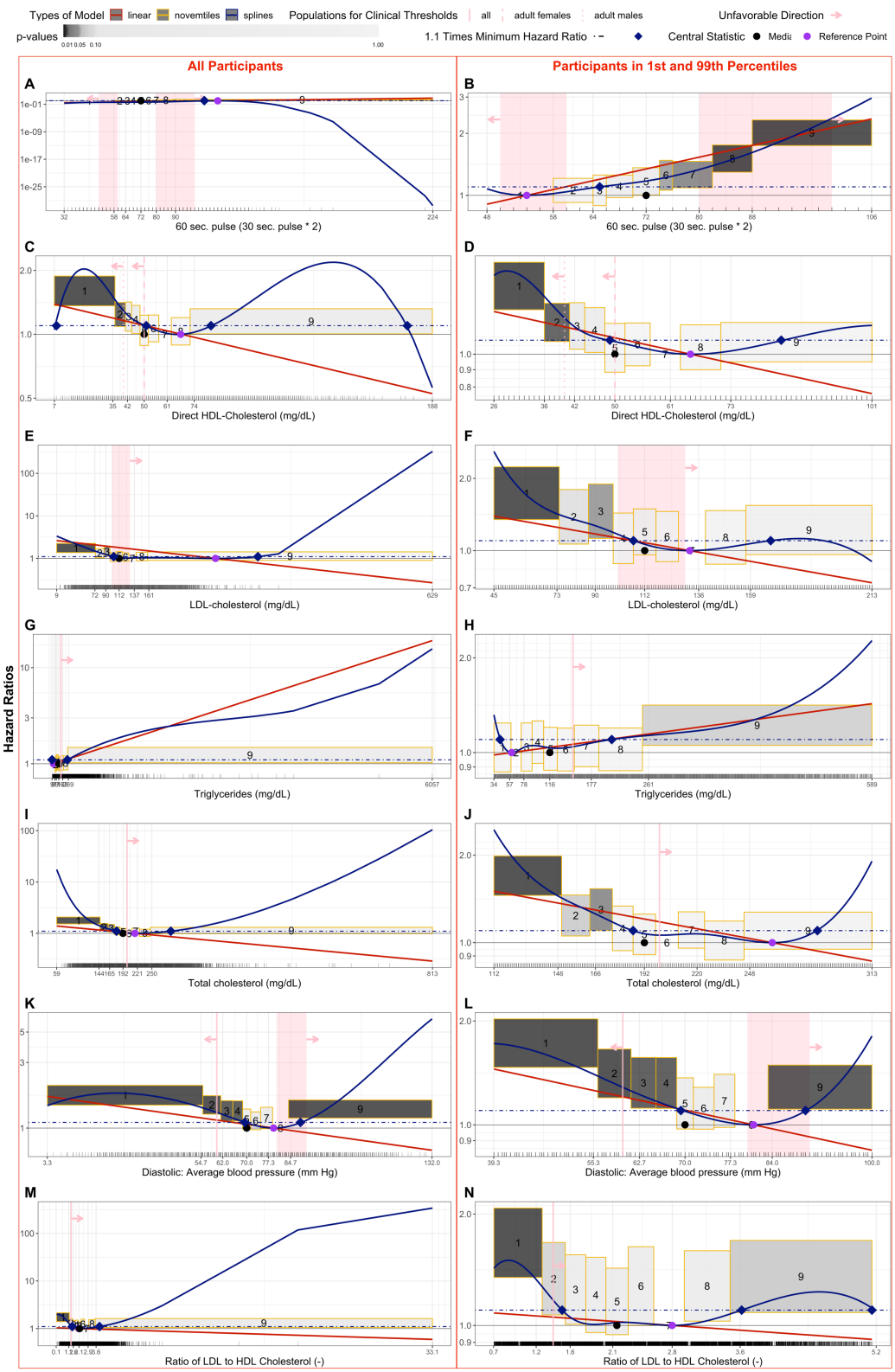

Figure A4.8. Stairway plots of hazard ratios across all models to describe the relative mortality risk for physiological indicators of the cardiovascular system.

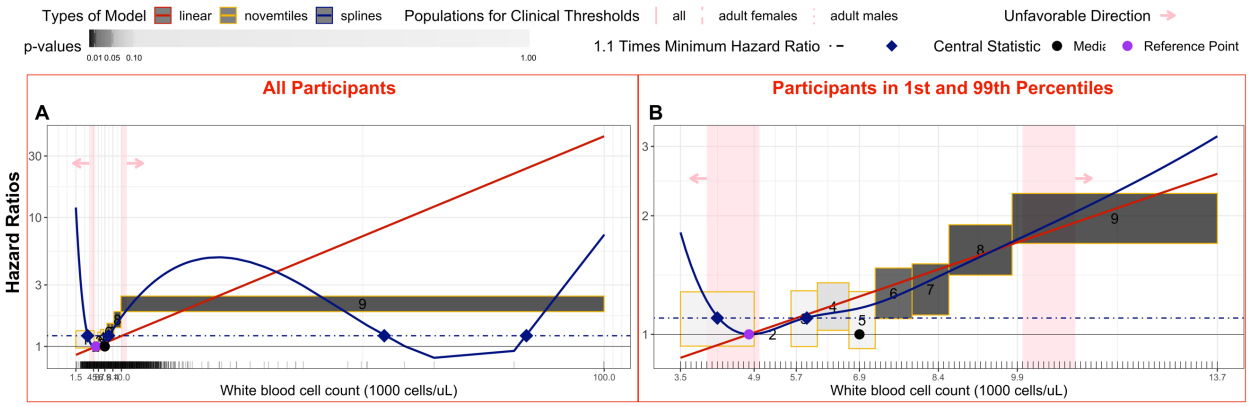

Figure A4.9. Stairway plots of hazard ratios across all models to describe the relative mortality risk for a biomarker of the immune system, White Blood Cell Counts.

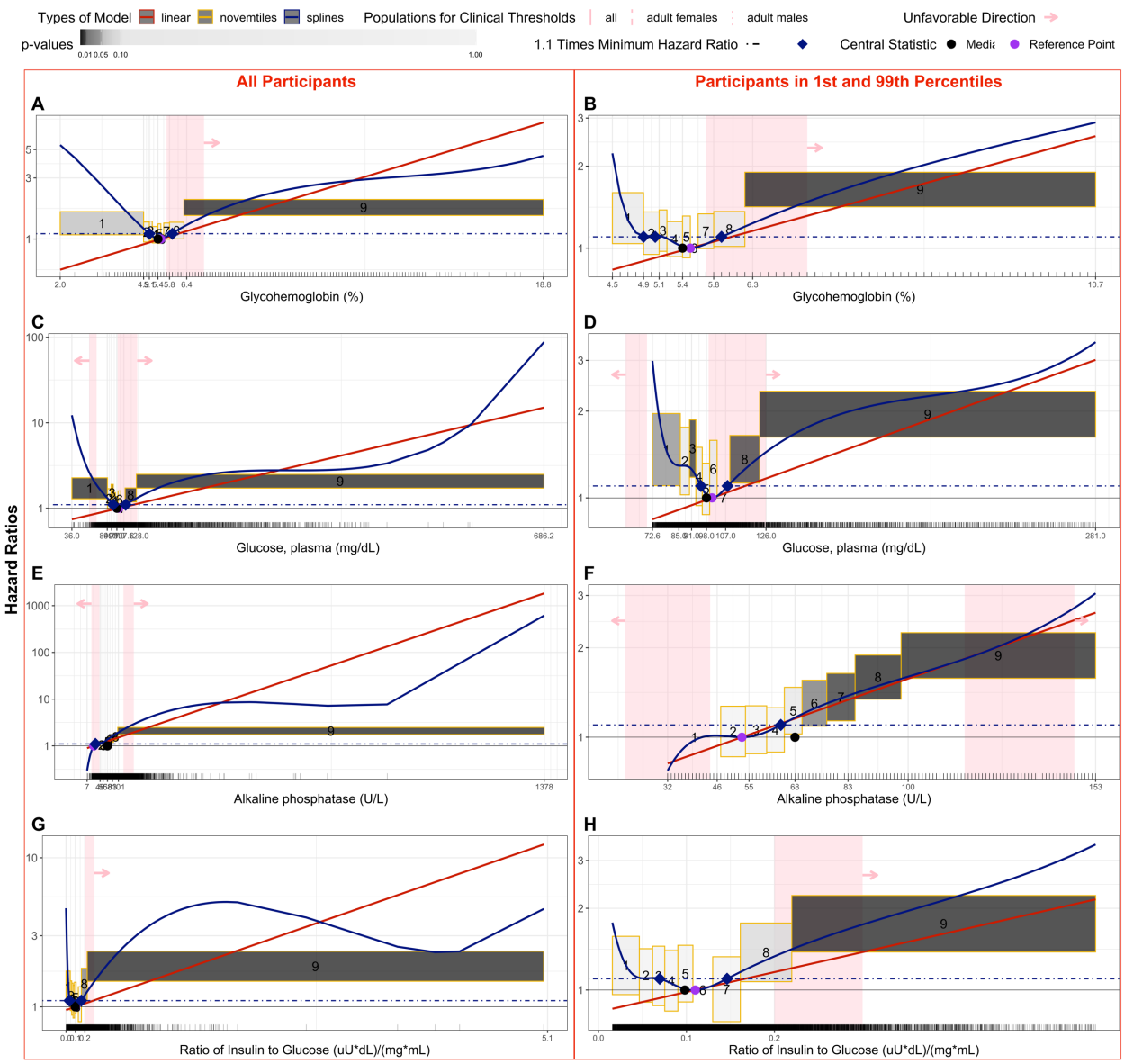

Figure A4.10. Stairway plots of hazard ratios across all models to describe the relative mortality risk for biomarkers of the metabolic system.

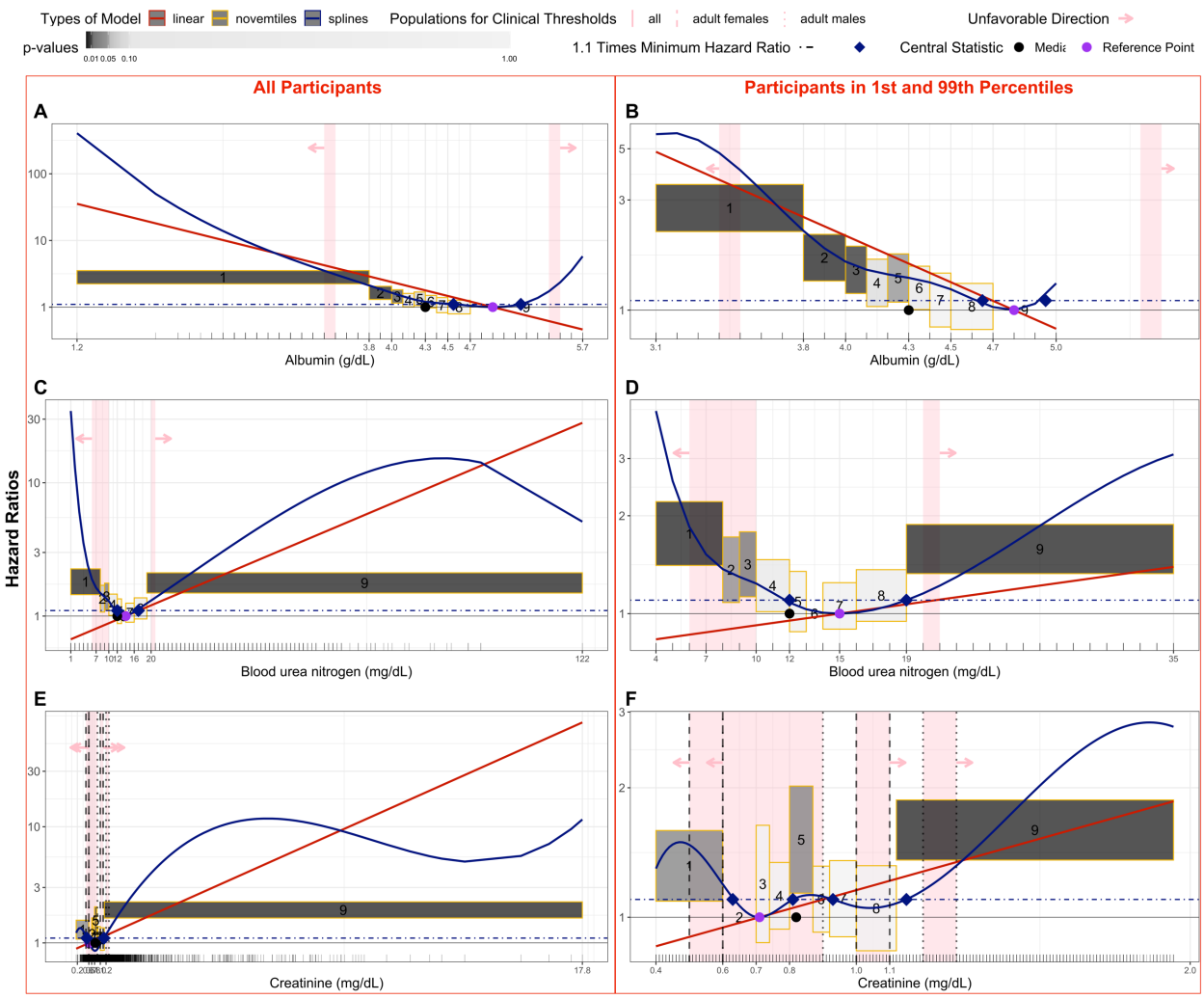

Figure A4.11. Stairway plots of hazard ratios across all models to describe the relative mortality risk for biomarkers of nephrology.

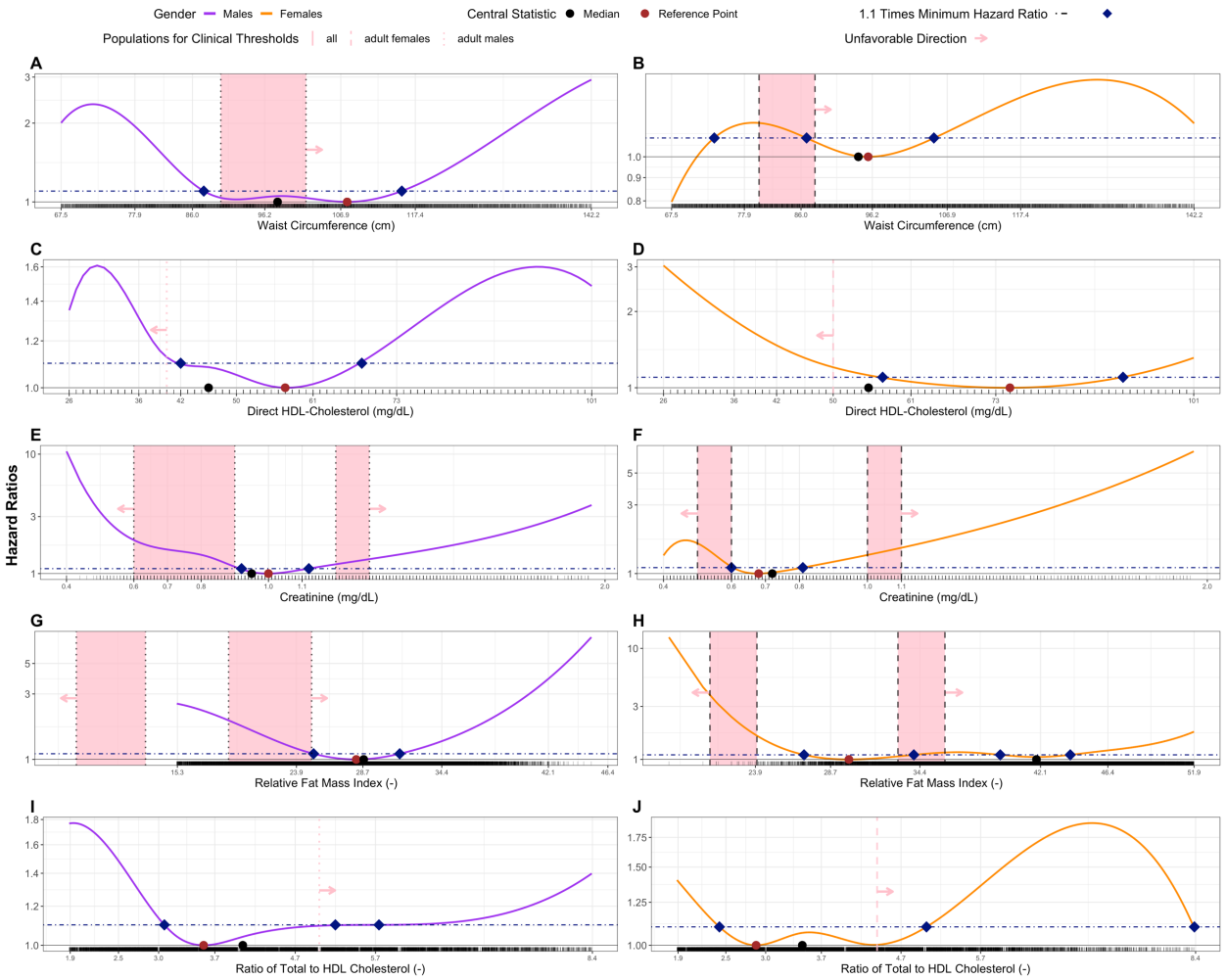

Figure A4.12. Sex-stratified non-linear associations between all-cause mortality and each physiological indicator with available sex-specific clinical thresholds. The non-linear associations were determined using a cubic spline regression model adjusted for age and race/ethnicity. Participants with measurements between the 1st and 99th percentiles of a physiological indicator are included. The purple and orange lines represent the relative mortality risk with respect to median of a physiological indicator for males and females, respectively. The black dot represents the median for a physiological indicator. The dashed navy line represents when the hazard ratio is 10% higher than the minimum hazard ratio. The navy diamonds indicate the concentration at which the hazard ratio shows a 10% increase from the minimum hazard ratio. The pink lines and rectangles represent the values of the clinical thresholds with the width of the rectangles representing the ranges of the threshold.

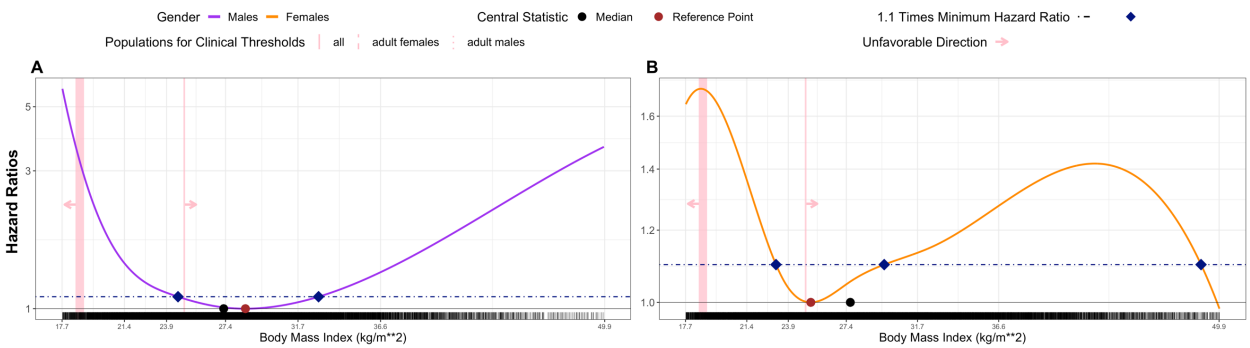

Figure A4.13. Sex-stratified non-linear associations between all-cause mortality and each physiological indicator of body composition, Body Mass Index.

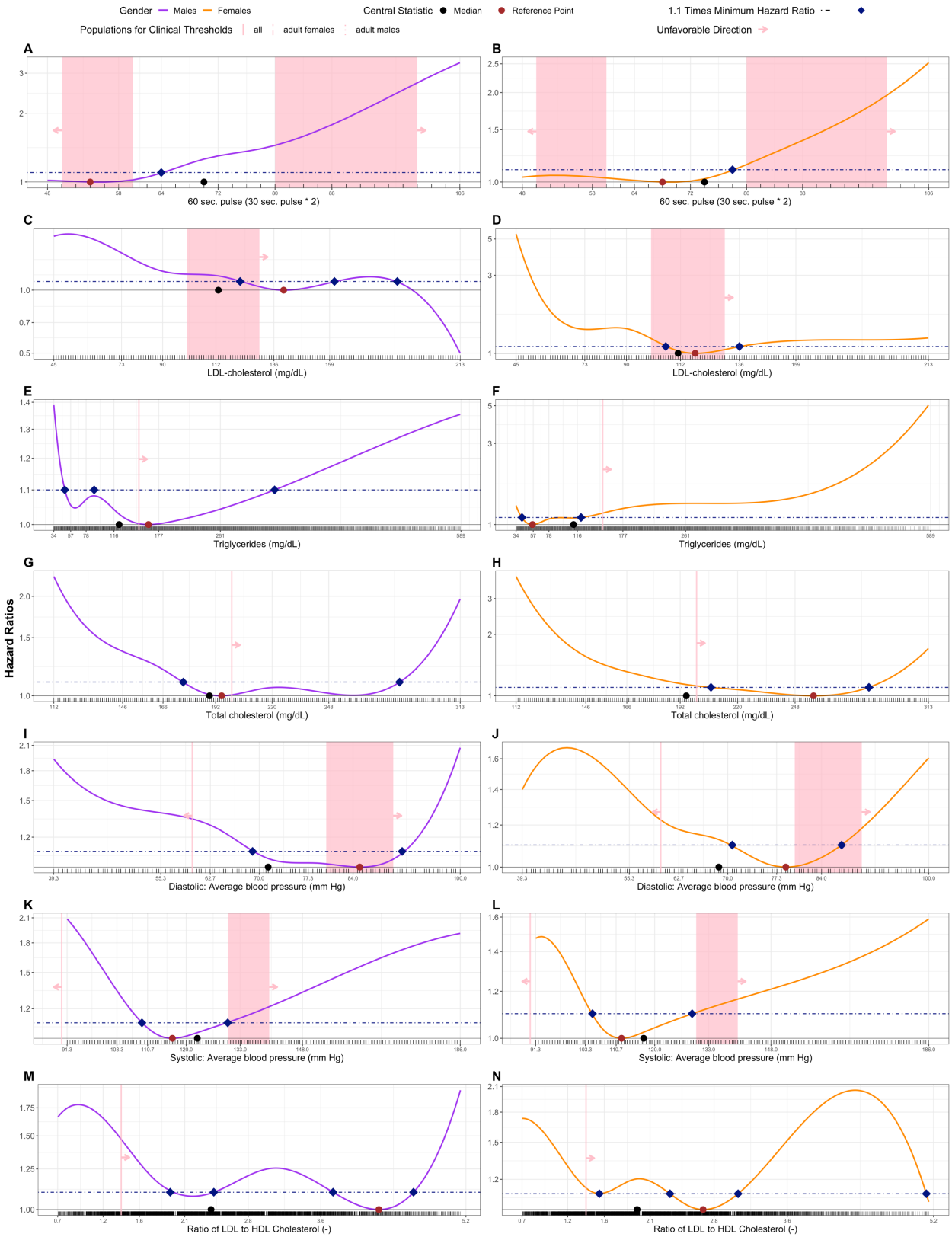

Figure A4.14. Sex-stratified non-linear associations between all-cause mortality and each physiological indicator of the cardiovascular system.

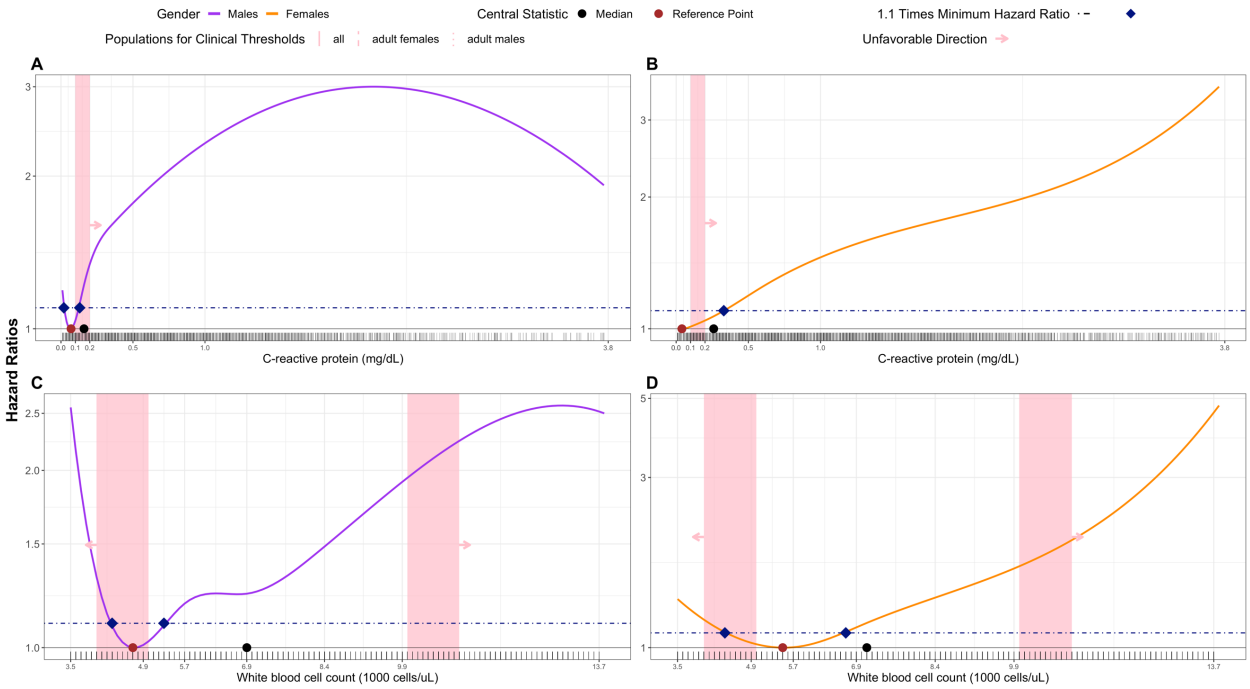

Figure A4.15. Sex-stratified non-linear associations between all-cause mortality and each physiological indicator of the immune system.

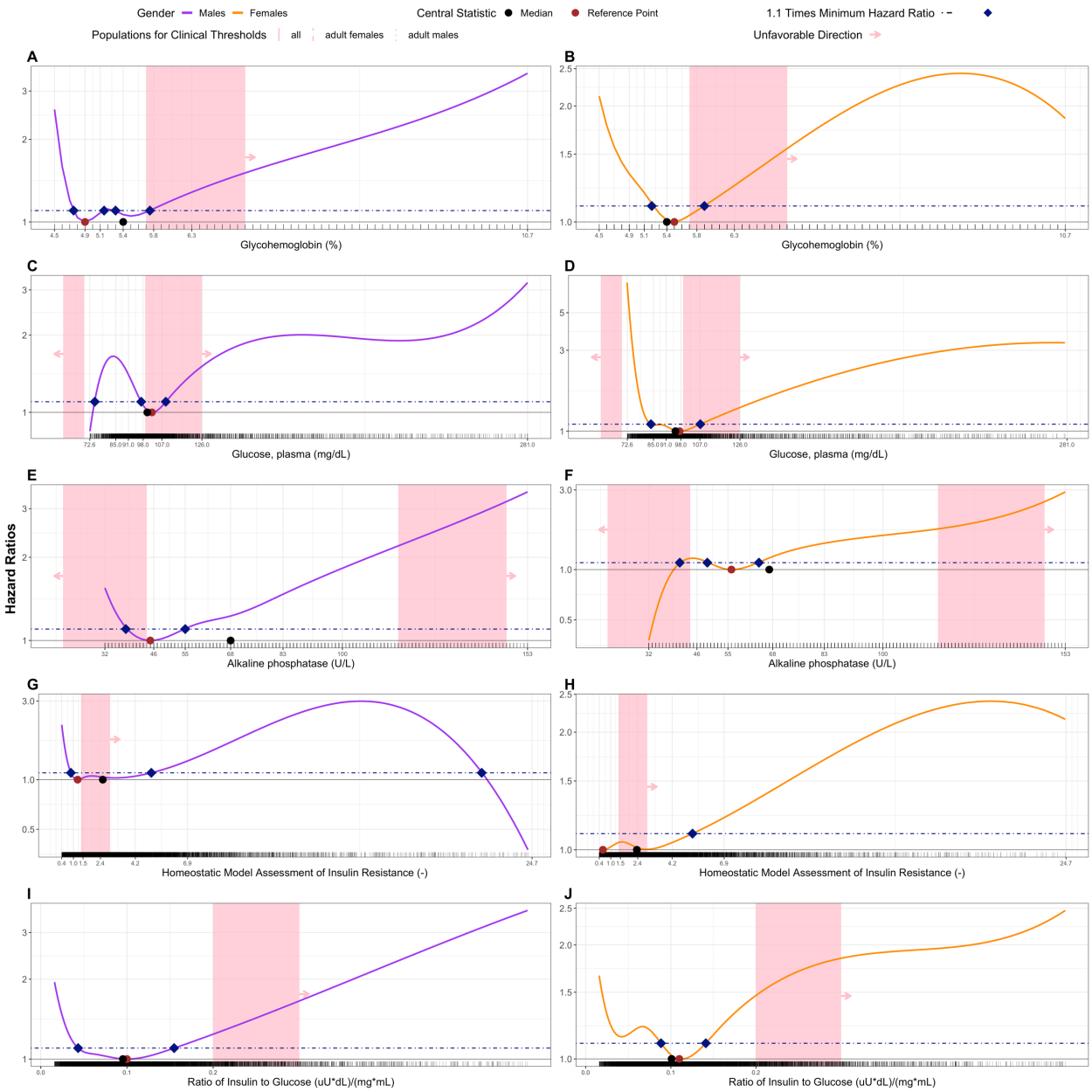

Figure A4.16. Sex-stratified non-linear associations between all-cause mortality and each physiological indicator of the metabolic system.

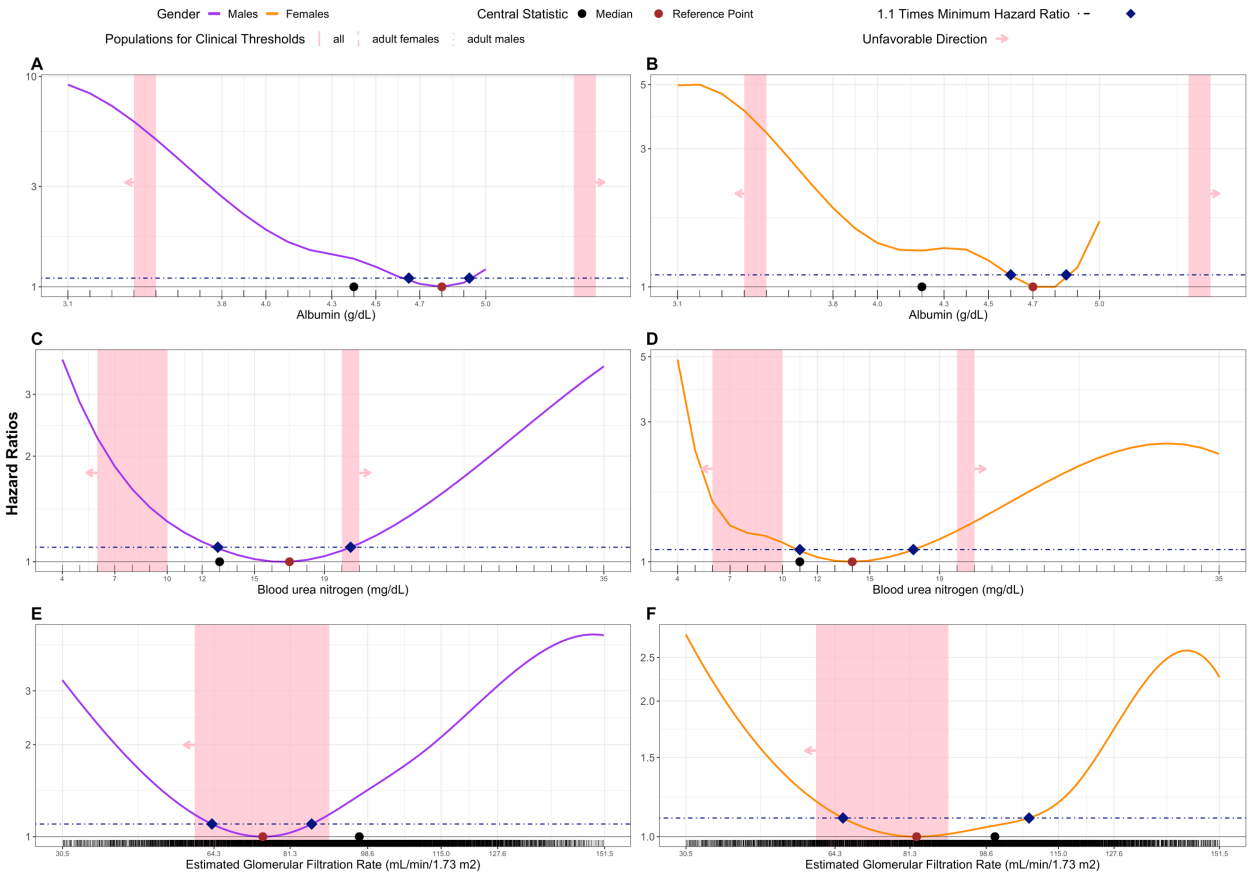

Figure A4.17. Sex-stratified non-linear associations between all-cause mortality and each physiological indicator of nephrology.

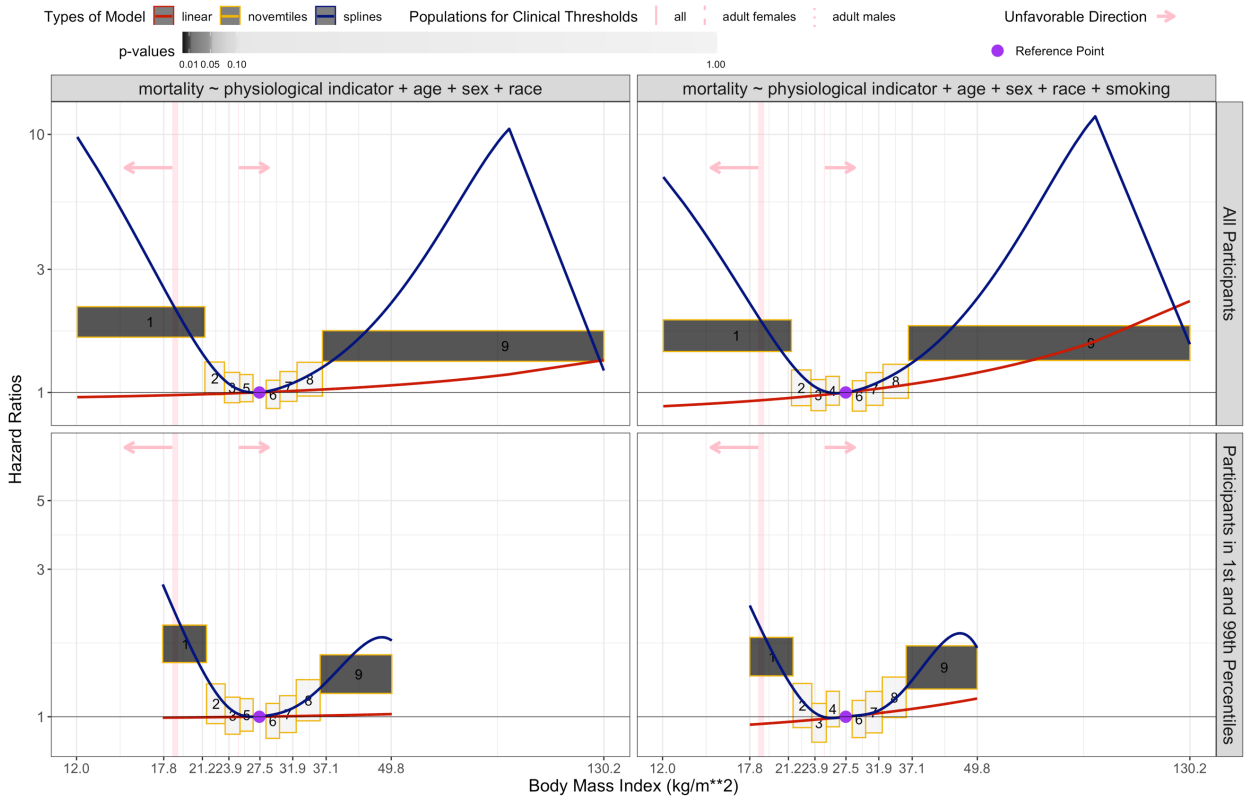

Figure A4.18. Stairway plot of hazard ratios describing the associations between all-cause mortality and BMI with and without adjusting for smoking. Smoking was defined using log-transformed blood cotinine levels.

Table A4.1. Number of participants by mortality status, gender, and race for the entire NHANES population and subpopulations with data for each physiological indicator.

Name	Alive	Deceased	Female	Male	Mexican Americans	Other Hispanics	Non-Hispanic White Americans	Non-Hispanic Black Americans	Other Race/ Multi-Racial
NHANES Population [a]			41647 (50.73)	40444 (49.27)	19161 (23.34)	6166 (7.51)	31216 (38.03)	19400 (23.63)	6148 (7.49)
NHANES Subpopulation [b]	39444 (87.6)	5588 (12.4)	23295 (51.7)	21737 (48.3)	8562 (19.0)	3286 (7.3)	20312 (45.1)	9665 (21.5)	3207 (7.2)
Body Mass Index (kg/m**2) [c]	38877 (88.26)	5170 (11.74)	22770 (51.69)	21277 (48.31)	8389 (19.05)	3232 (7.34)	19837 (45.04)	9437 (21.42)	3152 (7.16)
Standing Height (cm) [c]	39043 (88.16)	5245 (11.84)	22892 (51.69)	21396 (48.31)	8434 (19.04)	3247 (7.33)	19949 (45.04)	9495 (21.44)	3163 (7.14)
Subscapular Skinfold (mm) [c]	22422 (85.02)	3952 (14.98)	13305 (50.45)	13069 (49.55)	5681 (21.54)	1756 (6.66)	12970 (49.18)	4850 (18.39)	1117 (4.24)
Triceps Skinfold (mm) [c]	25357 (85.02)	4467 (14.98)	14886 (49.91)	14938 (50.09)	6521 (21.86)	1990 (6.67)	14255 (47.8)	5791 (19.42)	1267 (4.25)
Waist Circumference (cm) [c]	37695 (88.57)	4863 (11.43)	21914 (51.49)	20644 (48.51)	8182 (19.23)	3137 (7.37)	19198 (45.11)	9017 (21.19)	3024 (7.11)
Weight (kg) [c]	38933 (88.02)	5300 (11.98)	22860 (51.68)	21373 (48.32)	8415 (19.02)	3243 (7.33)	19954 (45.11)	9466 (21.4)	3155 (7.13)
Relative Fat Mass Index (-) [c]	37661 (88.77)	4764 (11.23)	21858 (51.52)	20567 (48.48)	8165 (19.25)	3128 (7.37)	19107 (45.04)	9004 (21.22)	3021 (7.12)
60 sec. pulse (30 sec. pulse * 2) [c]	37652 (87.67)	5293 (12.33)	22101 (51.46)	20844 (48.54)	8163 (19.01)	3114 (7.25)	19507 (45.42)	9152 (21.31)	3009 (7.01)
Direct HDL-Cholesterol (mg/dL) [c]	36892 (87.97)	5043 (12.03)	21624 (51.57)	20311 (48.43)	8059 (19.22)	3073 (7.33)	19205 (45.8)	8624 (20.57)	2974 (7.09)
LDL-cholesterol (mg/dL) [c]	37837 (87.69)	5310 (12.31)	22203 (51.46)	20944 (48.54)	8208 (19.02)	3142 (7.28)	19605 (45.44)	9182 (21.28)	3010 (6.98)
Triglycerides (mg/dL) [c]	37085 (87.99)	5060 (12.01)	21729 (51.56)	20416 (48.44)	8103 (19.23)	3103 (7.36)	19311 (45.82)	8653 (20.53)	2975 (7.06)
Total cholesterol (mg/dL) [c]	17411 (88.4)	2285 (11.6)	10224 (51.91)	9472 (48.09)	3762 (19.1)	1487 (7.55)	9048 (45.94)	3995 (20.28)	1404 (7.13)
Diastolic: Average blood pressure (mm Hg) [c]	36963 (87.97)	5055 (12.03)	21651 (51.53)	20367 (48.47)	8089 (19.25)	3092 (7.36)	19260 (45.84)	8615 (20.5)	2962 (7.05)
Systolic: Average blood pressure (mm Hg) [c]	37084 (87.99)	5063 (12.01)	21730 (51.56)	20417 (48.44)	8103 (19.23)	3103 (7.36)	19313 (45.82)	8653 (20.53)	2975 (7.06)

Ratio of LDL to HDL Cholesterol (-) [c]	37529 (88)	5118 (12)	21909 (51.37)	20738 (48.63)	8127 (19.06)	3121 (7.32)	19367 (45.41)	9037 (21.19)	2995 (7.02)
Ratio of Total to HDL Cholesterol (-) [c]	37660 (87.82)	5221 (12.18)	22037 (51.39)	20844 (48.61)	8170 (19.05)	3133 (7.31)	19483 (45.44)	9091 (21.2)	3004 (7.01)
C-reactive protein (mg/dL) [c]	26527 (84.83)	4742 (15.17)	16165 (51.7)	15104 (48.3)	6756 (21.61)	2042 (6.53)	14945 (47.79)	6220 (19.89)	1306 (4.18)
White blood cell count (1000 cells/uL) [c]	37317 (87.82)	5176 (12.18)	21994 (51.76)	20499 (48.24)	8127 (19.13)	3105 (7.31)	19418 (45.7)	8829 (20.78)	3014 (7.09)
Glycohemoglobin (%) [c]	37476 (87.85)	5181 (12.15)	22062 (51.72)	20595 (48.28)	8171 (19.16)	3129 (7.34)	19512 (45.74)	8825 (20.69)	3020 (7.08)
Glucose, plasma (mg/dL) [c]	18222 (87.79)	2535 (12.21)	10711 (51.6)	10046 (48.4)	3987 (19.21)	1562 (7.53)	9507 (45.8)	4241 (20.43)	1460 (7.03)
Alkaline phosphatase (U/L) [c]	32830 (89.15)	3997 (10.85)	18967 (51.5)	17860 (48.5)	6920 (18.79)	2874 (7.8)	16625 (45.14)	7621 (20.69)	2787 (7.57)
Homeostatic Model Assessment of Insulin Resistance (-) [c]	15241 (86.31)	2418 (13.69)	9073 (51.38)	8586 (48.62)	3574 (20.24)	1294 (7.33)	8179 (46.32)	3582 (20.28)	1030 (5.83)
Ratio of Insulin to Glucose (uU*dL)/(mg*mL) [c]	15241 (86.31)	2418 (13.69)	9073 (51.38)	8586 (48.62)	3574 (20.24)	1294 (7.33)	8179 (46.32)	3582 (20.28)	1030 (5.83)
Albumin (g/dL) [c]	36983 (87.97)	5058 (12.03)	21664 (51.53)	20377 (48.47)	8088 (19.24)	3095 (7.36)	19271 (45.84)	8623 (20.51)	2964 (7.05)
Blood urea nitrogen (mg/dL) [c]	36982 (87.97)	5056 (12.03)	21663 (51.53)	20375 (48.47)	8088 (19.24)	3095 (7.36)	19269 (45.84)	8623 (20.51)	2963 (7.05)
Creatinine (mg/dL) [c]	32832 (89.14)	3999 (10.86)	18968 (51.5)	17863 (48.5)	6920 (18.79)	2874 (7.8)	16626 (45.14)	7623 (20.7)	2788 (7.57)
Estimated Glomerular Filtration Rate (mL/min/1.73 m2) [c]	32832 (89.14)	3999 (10.86)	18968 (51.5)	17863 (48.5)	6920 (18.79)	2874 (7.8)	16626 (45.14)	7623 (20.7)	2788 (7.57)

[a] The NHANES population is defined as the sample with data available for age, gender, and race. Mortality data is available for participants who are 18 years or older.

[b] The NHANES subpopulation is defined as the sample with data available for mortality, age, gender, and race.

[c] Statistics for each physiological indicator is based on a subpopulation with data available for mortality status, age, gender, race/ethnicity, and the given indicator.

Table A4.2. Percentiles of age (years) for the entire NHANES population and subpopulations with data for each physiological indicator.

Name	Min	1st	5th	10th	Median	Mean	90th	95th	99th	Max
NHANES Population [a]	0	0	1	2	23	30.67	69	78	84	85
NHANES Subpopulation [b]	18	18	19	21	46	47.09	75	80	85	85
Body Mass Index (kg/m**2) [c]	18	18	19	21	46	46.88	75	80	85	85
Standing Height (cm) [c]	18	18	19	21	46	46.89	75	80	85	85
Subscapular Skinfold (mm) [c]	18	18	19	20	45	46.76	76	80	85	85
Triceps Skinfold (mm) [c]	18	18	19	20	45	46.85	76	80	85	85
Waist Circumference (cm) [c]	18	18	19	21	45	46.66	74	80	85	85
Weight (kg) [c]	18	18	19	21	46	46.98	75	80	85	85
Relative Fat Mass Index (-) [c]	18	18	19	21	45	46.57	74	80	85	85
60 sec. pulse (30 sec. pulse * 2) [c]	18	18	19	21	46	47.14	75	80	85	85
Direct HDL-Cholesterol (mg/dL) [c]	18	18	19	21	46	47.12	75	80	85	85
LDL-cholesterol (mg/dL) [c]	18	18	19	21	46	47.23	75	80	85	85
Triglycerides (mg/dL) [c]	18	18	19	21	46	47.11	75	80	85	85
Total cholesterol (mg/dL) [c]	18	18	19	21	46	47.12	75	80	85	85
Diastolic: Average blood pressure (mm Hg) [c]	18	18	19	21	46	47.00	75	80	85	85
Systolic: Average blood pressure (mm Hg) [c]	18	18	19	21	46	47.10	75	80	85	85
Ratio of LDL to HDL Cholesterol (-) [c]	18	18	19	21	46	47.23	75	80	85	85
Ratio of Total to HDL Cholesterol (-) [c]	18	18	19	21	46	47.12	75	80	85	85
C-reactive protein (mg/dL) [c]	18	18	19	21	46	47.09	76	80	85	85
White blood cell count (1000 cells/uL) [c]	18	18	19	21	46	47.19	75	80	85	85
Glycohemoglobin (%) [c]	18	18	19	21	46	47.19	75	80	85	85
Glucose, plasma (mg/dL) [c]	18	18	19	21	46	47.32	75	80	85	85
Alkaline phosphatase (U/L) [c]	18	18	19	21	46	47.27	75	80	85	85
Homeostatic Model Assessment of Insulin Resistance (-) [c]	18	18	19	21	46	47.17	75	80	85	85

Ratio of Insulin to Glucose (uU*dL)/(mg*mL) [c]	18	18	19	21	46	47.17	75	80	85	85
Albumin (g/dL) [c]	18	18	19	21	46	47.11	75	80	85	85
Blood urea nitrogen (mg/dL) [c]	18	18	19	21	46	47.11	75	80	85	85
Creatinine (mg/dL) [c]	18	18	19	21	46	47.27	75	80	85	85
Estimated Glomerular Filtration Rate (mL/min/1.73 m2) [c]	18	18	19	21	46	47.27	75	80	85	85

[a] The NHANES population is defined as the sample with data available for age, gender, and race. Mortality data is available for participants who are 18 years or older.

[b] The NHANES subpopulation is defined as the sample with data available for mortality, age, gender, and race.

[c] Statistics for each physiological indicator is based on a subpopulation with data available for mortality status, age, gender, race/ethnicity, and the given indicator.

Table A4.3. Percentiles of time to death (month) for a NHANES subpopulation and the subpopulations with data for each physiological indicator.

Name	Min	1st	5th	10th	Median	Mean	90th	95th	99th	Max
NHANES Subpopulation [a]	1	13	19	27	93	97.52	177	189	199	201
Body Mass Index (kg/m**2) [b]	1	13	19	27	93	97.52	177	189	199	201
Standing Height (cm) [b]	1	13	19	27	93	97.64	177	189	199	201
Subscapular Skinfold (mm) [b]	1	13	50	65	116	118.63	183	192	199	201
Triceps Skinfold (mm) [b]	1	13	49	65	116	118.86	183	192	199	201
Waist Circumference (cm) [b]	1	13	20	27.7	94	98.40	177	189	199	201
Weight (kg) [b]	1	13	19	27	92	97.41	177	189	199	201
Relative Fat Mass Index (-) [b]	1	13	20	28	94	98.49	178	189	199	201
60 sec. pulse (30 sec. pulse * 2) [b]	1	12	19	27	93	97.42	177	189	199	201
Direct HDL-Cholesterol (mg/dL) [b]	1	13	19	27	93	97.52	177	189	199	201
LDL-cholesterol (mg/dL) [b]	1	13	19	27	91	96.45	176	189	198	201
Triglycerides (mg/dL) [b]	1	13	19	27	93	97.60	177	189	199	201
Total cholesterol (mg/dL) [b]	1	13	19	27	93	97.52	177	189	199	201
Diastolic: Average blood pressure (mm Hg) [b]	1	13	19	27	92	97.40	177	189	199	201
Systolic: Average blood pressure (mm Hg) [b]	1	13	19	27	92	97.38	177	189	199	201
Ratio of LDL to HDL Cholesterol (-) [b]	1	13	19	27	91	96.45	176	189	198	201
Ratio of Total to HDL Cholesterol (-) [b]	1	13	19	27	93	97.52	177	189	199	201
C-reactive protein (mg/dL) [b]	1	13	49	65	117	118.82	183	192	199	201
White blood cell count (1000 cells/uL) [b]	1	13	19	27	93	97.45	177	189	199	201
Glycohemoglobin (%) [b]	1	13	19	27	93	97.47	177	189	199	201
Glucose, plasma (mg/dL) [b]	1	13	19	27	93	97.67	177	189	198	201
Alkaline phosphotase (U/L) [b]	1	13	19	25	83	89.85	168	190	199	201
Homeostatic Model Assessment of Insulin Resistance (-) [b]	1	13	39	46	104	109.08	181	191	199	201

Ratio of Insulin to Glucose (uU*dL)/(mg*mL) [b]	1	13	39	46	104	109.08	181	191	199	201
Albumin (g/dL) [b]	1	13	19	27	93	97.58	177	189	199	201
Blood urea nitrogen (mg/dL) [b]	1	13	19	27	93	97.58	177	189	199	201
Creatinine (mg/dL) [b]	1	13	19	25	83	89.85	168	190	199	201
Estimated Glomerular Filtration Rate (mL/min/1.73 m2) [b]	1	13	19	25	83	89.85	168	190	199	201

[a] The NHANES subpopulation is defined as the sample with data available for mortality, age, gender, and race.

[b] Statistics for each physiological indicator is based on a subpopulation with data available for mortality status, age, gender, race/ethnicity, and the given indicator.

Table A4.4. Definition, interpretation, and justification of prediction measures.

	Definition	Interpretation	Justification
Akaike Information Criterion (AIC)	Estimator of the relative amount of information lost by a model	A lower AIC implies higher quality of the model and minimum information lost.	The AIC is useful for comparing models but is not informative when interpreting the goodness of fit of a single isolated model.
Concordance Index	Proportion of concordant pairs divided by the total number of all possible pairs.	The Concordance Index represents the model ability to correctly rank the survival times based on the individual risk scores. A Concordance Index of 1 implies a perfect prediction, while 0.5 implies random predictions.	The Concordance Index is the standard measure for model assessment in survival analysis.
Nagelkerke R^2	Adjusted version of the Cox & Snell R2 that adjusted for the scale of the statistics to range from 0 to 1. The Cox & Snell R^2 reflect the improvement of the given model over the intercept model.	A Nagelkerke R^2 of 1 implies a perfect model.	The Nagelkerke R^2 better highlights the overfitting of the linear and spline models.

BIBLIOGRAPHY

1. Sexton, K., L. Needham, L. & L. Pirkle, J. Human Biomonitoring of Environmental Chemicals: Measuring chemicals in human tissues is the" gold standard" for assessing people's exposure to pollution. *Am. Sci.* **92**, 38–45 (2004).

2. Wild, C. P. The exposome: from concept to utility. *Int. J. Epidemiol.* **41**, 24–32 (2012).

3. DeBord, D. G. *et al.* Use of the "exposome" in the practice of epidemiology: a primer on-omic technologies. *Am. J. Epidemiol.* **184**, 302–314 (2016).

4. Kraft, P. & Hunter, D. J. Genetic risk prediction—are we there yet? *N. Engl. J. Med.* **360**, 1701–1703 (2009).

5. Rappaport, S. M. Implications of the exposome for exposure science. *J. Expo. Sci. Environ. Epidemiol.* **21**, 5–9 (2011).

6. Rappaport, S. M. Genetic factors are not the major causes of chronic diseases. *PLoS One* **11**, e0154387 (2016).

7. Wang, A. *et al.* The Pregnancy Chemisome in Relation to Birth Outcomes and Consumer Product Use: Suspect Screening of Industrial Chemicals. in *ISEE Conference Abstracts* vol. 2018 (2018).

8. Smyth, H. F. & Smyth Jr, H. F. Spray painting hazards as determined by the Pennsylvania and the National Safety Council surveys. *J. Ind. Hyg.* **10**, 163–214 (1928).

9. Roach, S. A. A more rational basis for air sampling programs. *Am. Ind. Hyg. Assoc. J.* **27**, 1–12 (1966).

10. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D. & Milzani, A. Biomarkers of oxidative damage in human disease. *Clin. Chem.* **52**, 601–623 (2006).

11. Dunn, W. B. *et al.* A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. *Int. J. Epidemiol.* **37**, i23–i30 (2008).

12. Rappaport, S. M. & Smith, M. T. Environment and disease risks. *Science* (2010) doi:10.1126/science.1192603.

13. Rappaport, S. M. Biomarkers intersect with the exposome. *Biomarkers* **17**, 483–489 (2012).

14. CDC. *Fourth National Report on Human Exposure to Environmental Chemicals*.

https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf (2009).

15. Metcalf, S. W. & Orloff, K. G. Biomarkers of exposure in community settings. *J. Toxicol. Environ. Heal. Part A* **67**, 715–726 (2004).

16. World Health Organization. *Biomarkers in Risk Assessment: Validity and Vlidation-Environmental Health Criteria 222*. (2001).

17. Sobus, J. R. *et al.* Uses of NHANES biomarker data for chemical risk assessment: trends, challenges, and opportunities. *Environ. Health Perspect.* **123**, 919–927 (2015).

18. NCHS. Survey Design Factors Course. *Continuous NHANES Web Tutorial* https://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/intro.htm (2013).

19. NCHS. National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/index.htm (2017).

20. Martin Sanchez, F., Gray, K., Bellazzi, R. & Lopez-Campos, G. Exposome informatics: considerations for the design of future biomedical research information systems. *J. Am. Med. Informatics Assoc.* **21**, 386–390 (2013).

21. Haines, J. L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science (80-.).* **308**, 419–421 (2005).

22. Patel, C. J., Bhattacharya, J. & Butte, A. J. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One* **5**, e10746 (2010).

23. Hardy, M. L. The toxicology of the three commercial polybrominated diphenyl oxide (ether) flame retardants. *Chemosphere* **46**, 757–777 (2002).

24. Kato, K. *et al.* Mono (2-ethyl-5-hydroxyhexyl) phthalate and mono-(2-ethyl-5-oxohexyl) phthalate as biomarkers for human exposure assessment to di-(2-ethylhexyl) phthalate. *Environ. Health Perspect.* **112**, 327–330 (2004).

25. Xu, X. *et al.* Association between exposure to alkylbenzenes and cardiovascular disease among National Health and Nutrition Examination Survey (NHANES) participants. *Int. J. Occup. Environ. Health* **15**, 385–391 (2009).

26. Caldwell, K. L., Hartel, J., Jarrett, J. & Jones, R. L. Inductively coupled plasma mass spectrometry to measure multiple toxic elements in urine in NHANES 1999-2000. *At. Spectrosc.* **26**, 1–7 (2005).

27. Lin, Y.-S., Rathod, D., Ho, W.-C. & Caffrey, J. J. Cadmium exposure is associated with elevated blood C-reactive protein and fibrinogen in the US population: the third national health and nutrition examination survey (NHANES III, 1988–1994). *Ann. Epidemiol.* **19**, 592–596 (2009).

28. Zota, A. R., Calafat, A. M. & Woodruff, T. J. Temporal Trends in Phthalate Exposures:

Findings from the National Health and Nutrition Examination Survey, 2001–2010. *Environ. Health Perspect.* **122**, 235–241 (2014).

29. James-Todd, T. M., Chiu, Y.-H. & Zota, A. R. Racial/ethnic disparities in environmental endocrine disrupting chemicals and women's reproductive health outcomes: epidemiological examples across the life course. *Curr. Epidemiol. reports* **3**, 161–180 (2016).

30. Birks, L. *et al.* Occupational exposure to endocrine-disrupting chemicals and birth weight and length of gestation: A European meta-analysis. *Environ. Health Perspect.* **124**, 1785–1793 (2016).

31. Cave, M. *et al.* Polychlorinated biphenyls, lead, and mercury are associated with liver disease in American adults: NHANES 2003–2004. *Environ. Health Perspect.* **118**, 1735–1742 (2010).

32. Sexton, K. *et al.* Children's exposure to volatile organic compounds as determined by longitudinal measurements in blood. *Environ. Health Perspect.* **113**, 342–349 (2005).

33. Borrell, L. N., Factor-Litvak, P., Wolff, M. S., Susser, E. & Matte, T. D. Effect of socioeconomic status on exposures to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE) among pregnant African-American women. *Arch. Environ. Heal. An Int. J.* **59**, 250–255 (2004).

34. Hauser, R., Williams, P., Altshul, L. & Calafat, A. M. Evidence of interaction between polychlorinated biphenyls and phthalates in relation to human sperm motility. *Environ. Health Perspect.* **113**, 425–430 (2005).

35. Koeppe, E. S., Ferguson, K. K., Colacino, J. A. & Meeker, J. D. Relationship between urinary triclosan and paraben concentrations and serum thyroid measures in NHANES 2007–2008. *Sci. Total Environ.* **445**, 299–305 (2013).

36. Awata, H., Linder, S., Mitchell, L. E. & Delclos, G. L. Biomarker levels of toxic metals among Asian populations in the United States: NHANES 2011–2012. *Environ. Health Perspect.* (2017) doi:10.1289/EHP27.

37. Brenner, S. A., Neu-Baker, N. M., Caglayan, C. & Zurbenko, I. G. Occupational exposure to airborne nanomaterials: an assessment of worker exposure to aerosolized metal oxide nanoparticles in semiconductor wastewater treatment. *J. Occup. Environ. Hyg.* **12**, 469–481 (2015).

38. Hong, Y.-C. *et al.* Community level exposure to chemicals and oxidative stress in adult population. *Toxicol. Lett.* **184**, 139–144 (2009).

39. Agarwal, S., Zaman, T., Murat Tuzcu, E. & Kapadia, S. R. Heavy metals and cardiovascular disease: results from the National Health and Nutrition Examination Survey (NHANES) 1999-2006. *Angiology* **62**, 422–429 (2011).

40. Streit, L. 8 Signs and Symptoms of Vitamin A Deficiency. *healthline* https://www.healthline.com/nutrition/vitamin-a-deficiency-symptoms#section4 (2018).

41. Den Hond, E. *et al.* Biomarkers of human exposure to personal care products: Results from the Flemish Environment and Health Study (FLEHS 2007–2011). *Sci. Total Environ.* **463**–**464**, 102–110 (2013).

42. Schwartz, G. G., Il'yasova, D. & Ivanova, A. Urinary cadmium, impaired fasting glucose, and diabetes in the NHANES III. *Diabetes Care* **26**, 468–470 (2003).

43. Wang, S.-L. *et al.* Infant exposure to polychlorinated dibenzo-p-dioxins, dibenzofurans and biphenyls (PCDD/Fs, PCBs)—correlation between prenatal and postnatal exposure. *Chemosphere* **54**, 1459–1473 (2004).

44. Bhanegaonkar, A. J. Exposure to Volatile Organic Compounds and Effect on Neurobehavioral Function. (2005).

45. Hoppin, J. A., Ulmer, R. & London, S. J. Phthalate exposure and pulmonary function. *Environ. Health Perspect.* **112**, 571–574 (2004).

46. Erlinger, T. P., Muntner, P. & Helzlsouer, K. J. WBC count and the risk of cancer mortality in a national sample of US adults: results from the Second National Health and Nutrition Examination Survey mortality study. *Cancer Epidemiol. Prev. Biomarkers* **13**, 1052–1056 (2004).

47. Mensah, G. A., Brown, D. W., Croft, J. B. & Greenlund, K. J. Major coronary risk factors and death from coronary heart disease: baseline and follow-up mortality data from the Second National Health and Nutrition Examination Survey (NHANES II). *Am. J. Prev. Med.* **29**, 68–74 (2005).

48. Brown, D. W. *et al.* Associations between white blood cell count and risk for cerebrovascular disease mortality: NHANES II Mortality Study, 1976–1992. *Ann. Epidemiol.* **14**, 425–430 (2004).

49. Gillum, R. F., Mussolino, M. E. & Madans, J. H. Counts of neutrophils, lymphocytes, and monocytes, cause-specific mortality and coronary heart disease: the NHANES-I epidemiologic follow-up study. *Ann. Epidemiol.* **15**, 266–271 (2005).

50. Allison, D. B., Zhu, S. K., Plankey, M., Faith, M. S. & Heo, M. Differential associations of body mass index and adiposity with all-cause mortality among men in the first and second National Health and Nutrition Examination Surveys (NHANES I and NHANES II) follow-up studies. *Int. J. Obes.* **26**, 410–416 (2002).

51. Garg, A. X., Clark, W. F., Haynes, R. B. & House, A. A. Moderate renal insufficiency and the risk of cardiovascular mortality: results from the NHANES I. *Kidney Int.* **61**, 1486–1494 (2002).

52. Shimokata, H. & Kuzuya, F. [Aging, basal metabolic rate, and nutrition]. *Nihon Ronen*

Igakkai Zasshi. **30**, 572–576 (1993).

53. Speakman, J. R. Body size, energy metabolism and lifespan. *J. Exp. Biol.* **208**, 1717–1730 (2005).

54. Renwick, A. G. Toxicokinetics in infants and children in relation to the ADI and TDI. *Food Addit. Contam.* **15**, 17–35 (1998).

55. Scheuplein, R., Charnley, G. & Dourson, M. Differential Sensitivity of Children and Adults to Chemical Toxicity: I. Biological Basis. *Regul. Toxicol. Pharmacol.* **35**, 429–447 (2002).

56. ATSDR, A. Z. Case Studies in Environmental Medicine. (1991).

57. Csaba, G. & Dobozy, O. The sensitivity of sugar receptors--analysis in adult animals of influences exerted at neonatal age. *Endokrinologie* **69**, 227–232 (1977).

58. Gauderman, W. J. *et al.* The effect of air pollution on lung development from 10 to 18 years of age. *N. Engl. J. Med.* **351**, 1057–1067 (2004).

59. Just, A. C. *et al.* Vinyl flooring in the home is associated with children's airborne butylbenzyl phthalate and urinary metabolite concentrations. *J. Expo. Sci. Environ. Epidemiol.* **25**, 574–579 (2015).

60. Tsou, M.-C. *et al.* Mouthing activity data for children aged 7 to 35 months old in Taiwan. *J. Expo. Sci. Environ. Epidemiol.* **25**, 388–398 (2015).

61. Xu, Y., Cohen Hubal, E. A. & Little, J. C. Predicting Residential Exposure to Phthalate Plasticizer Emitted from Vinyl Flooring: Sensitivity, Uncertainty, and Implications for Biomonitoring. *Environ. Health Perspect.* **118**, 253–258 (2010).

62. Kumar, A. & Pastore, P. Lead and cadmium in soft plastic toys. *Curr. Sci.* **93**, 818–822 (2007).

63. Needham, L. L., Barr, D. B. & Calafat, A. M. Characterizing children's exposures: beyond NHANES. *Neurotoxicology* **26**, 547–553 (2005).

64. Shea, K. M. Pediatric Exposure and Potential Toxicity of Phthalate Plasticizers. *Pediatrics* **111**, 1467 LP – 1474 (2003).

65. Sjödin, A. *et al.* Serum Concentrations of Polybrominated Diphenyl Ethers (PBDEs) and Polybrominated Biphenyl (PBB) in the United States Population: 2003–2004. *Environ. Sci. Technol.* **42**, 1377–1384 (2008).

66. Gill, U., Chu, I., Ryan, J. J. & Feeley, M. Polybrominated diphenyl ethers: human tissue levels and toxicology. in *Reviews of environmental contamination and toxicology* 55–97 (Springer, 2004).

67. Ballew, C. *et al.* Blood lead concentration and children's anthropometric dimensions in the

Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994. *J. Pediatr.* **134**, 623–630 (1999).

68. Canfield, R. L. *et al.* Intellectual impairment in children with blood lead concentrations below 10 μg per deciliter. *N. Engl. J. Med.* **348**, 1517–1526 (2003).

69. Lidsky, T. I. & Schneider, J. S. Lead neurotoxicity in children: basic mechanisms and clinical correlates. *Brain* **126**, 5–19 (2003).

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

71. Prüss-Ustün, A., Vickers, C., Haefliger, P. & Bertollini, R. Knowns and unknowns on burden of disease due to chemicals: a systematic review. *Environ. Heal.* **10**, 9 (2011).

72. World Health Organization. *The public health impact of chemicals: knowns and unknowns*. (2016).

73. Quinn, C. L. & Wania, F. Understanding Differences in the Body Burden–Age Relationships of Bioaccumulating Contaminants Based on Population Cross Sections versus Individuals. *Environ. Health Perspect.* **120**, 554–559 (2012).

74. Ritter, R. *et al.* Intrinsic human elimination half-lives of polychlorinated biphenyls derived from the temporal evolution of cross-sectional biomonitoring data from the United Kingdom. *Environ. Health Perspect.* **119**, 225–231 (2011).

75. Jolliet, O. *et al.* Influence of age on serum dioxin concentrations as a function of cogender half-life and historical peak food contamination. in *Proceedings of the Dioxin 2008 Conference* (2008).

76. Gomis, M. I., Vestergren, R., MacLeod, M., Mueller, J. F. & Cousins, I. T. Historical human exposure to perfluoroalkyl acids in the United States and Australia reconstructed from biomonitoring data using population-based pharmacokinetic modelling. *Environ. Int.* **108**, 92–102 (2017).

77. Wong, F., MacLeod, M., Mueller, J. F. & Cousins, I. T. Enhanced elimination of perfluorooctane sulfonic acid by menstruating women: evidence from population-based pharmacokinetic modeling. *Environ. Sci. Technol.* **48**, 8807–8814 (2014).

78. Williams, D. R., Mohammed, S. A. & Shields, A. E. Understanding and effectively addressing breast cancer in African American women: Unpacking the social context. *Cancer* vol. 122 2138–2149 (2016).

79. Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J. Am. Med. Assoc.* **295**, 2492–2502 (2006).

80. Stark, A. *et al.* African ancestry and higher prevalence of triple-negative breast cancer.

Cancer **116**, 4926–4932 (2010).

81. Braun, L. Race, Ethnicity, and Health: Can Genetics Explain Disparities? *Perspect. Biol. Med.* (2007) doi:10.1353/pbm.2002.0023.

82. Diez Roux, A. V. *Conceptual Approaches to the Study of Health Disparities*. *SSRN* (2012) doi:10.1146/annurev-publhealth-031811-124534.

83. Cooper, R. S., Kaufman, J. S. & Ward, R. Race and Genomics. *N. Engl. J. Med.* (2003) doi:10.1056/nejmsb022863.

84. Jing, L., Su, L. & Ring, B. Z. Ethnic background and genetic variation in the evaluation of cancer risk: A systematic review. *PLoS One* (2014) doi:10.1371/journal.pone.0097522.

85. Rodgers, K. M., Udesky, J. O., Rudel, R. A. & Brody, J. G. Environmental chemicals and breast cancer: an updated review of epidemiological literature informed by biological mechanisms. *Environ. Res.* **160**, 152–182 (2018).

86. Campeau, P. M., Foulkes, W. D. & Tischkowitz, M. D. Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Hum. Genet.* **124**, 31–42 (2008).

87. Martin, A.-M. & Weber, B. L. Genetic and hormonal risk factors in breast cancer. *J. Natl. Cancer Inst.* **92**, 1126–1135 (2000).

88. Hayes, R. B. Review of occupational epidemiology of chromium chemicals and respiratory cancer. *Sci. Total Environ.* **71**, 331–339 (1988).

89. Villeneuve, S. *et al.* Occupation and occupational exposure to endocrine disrupting chemicals in male breast cancer: a case–control study in Europe. *Occup. Environ. Med.* **67**, 837–844 (2010).

90. NIOSH, N. I. for O. S. and H. Workplace Safety & Health Topics. Skin Exposures and Effects.". (2019).

91. Assembly, U. N. G. & Tuncak, B. *Report of the Special Rapporteur on the implications for human rights of the environmentally sound management and disposal of hazardous substances and wastes*. (United Nations General Assembly, 2018).

92. Ngamwong, Y. *et al.* Additive synergism between asbestos and smoking in lung cancer risk: a systematic review and meta-analysis. *PLoS One* **10**, e0135798 (2015).

93. Shim, Y. K., Lewin, M. D., Ruiz, P., Eichner, J. E. & Mumtaz, M. M. Prevalence and associated demographic characteristics of exposure to multiple metals and their species in human populations: The United States NHANES, 2007–2012. *J. Toxicol. Environ. Heal. Part A* **80**, 502–512 (2017).

94. Sadetzki, S. *et al.* Occupational exposure to metals and risk of meningioma: a multinational case-control study. *J. Neurooncol.* **130**, 505–515 (2016).

95. Hengstler, J. G. *et al.* Occupational exposure to heavy metals: DNA damage induction and DNA repair inhibition prove co-exposures to cadmium, cobalt and lead as more dangerous than hitherto expected. *Carcinogenesis* **24**, 63–73 (2003).

96. Park, S. K., Tao, Y., Meeker, J. D., Harlow, S. D. & 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* **9**, e98632 (2014).

97. Kuang, H. *et al.* Co-exposure to polycyclic aromatic hydrocarbons, benzene and toluene may impair lung function by increasing oxidative damage and airway inflammation in asthmatic children. *Environ. Pollut.* 115220 (2020).

98. Shiue, I. Urinary heavy metals, phthalates and polyaromatic hydrocarbons independent of health events are associated with adult depression: USA NHANES, 2011–2012. *Environ. Sci. Pollut. Res.* **22**, 17095–17103 (2015).

99. Bao, W., Liu, B., Simonsen, D. W. & Lehmler, H.-J. Association between exposure to pyrethroid insecticides and risk of all-cause and cause-specific mortality in the general US adult population. *JAMA Intern. Med.* **180**, 367–374 (2020).

100. Michels, K. B., De Vivo, I., Calafat, A. M. & Binder, A. M. In utero exposure to endocrine-disrupting chemicals and telomere length at birth. *Environ. Res.* **182**, 109053 (2020).

101. Christensen, K. L. Y., Carrico, C. K., Sanyal, A. J. & Gennings, C. Multiple classes of environmental chemicals are associated with liver disease: NHANES 2003–2004. *Int. J. Hyg. Environ. Health* **216**, 703–709 (2013).

102. Ricardo, A. C. *et al.* Retinopathy and CKD as predictors of all-cause and cardiovascular mortality: National Health and Nutrition Examination Survey (NHANES) 1988-1994. *Am. J. kidney Dis.* **64**, 198–203 (2014).

103. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

104. Jain, R. B. Trends and concentrations of selected polycyclic aromatic hydrocarbons in general US population: Data from NHANES 2003–2008. *Cogent Environ. Sci.* **1**, 1031508 (2015).

105. Carwile, J. L. & Michels, K. B. Urinary bisphenol A and obesity: NHANES 2003–2006. *Environ. Res.* **111**, 825–830 (2011).

106. Park, S. K., Zhao, Z. & Mukherjee, B. Construction of environmental risk score beyond standard linear models using machine learning methods: application to metal mixtures, oxidative stress and cardiovascular disease in NHANES. *Environ. Heal.* **16**, 102 (2017).

107. Filipowicz, R. *et al.* Associations of serum skeletal alkaline phosphatase with elevated C-reactive protein and mortality. *Clin. J. Am. Soc. Nephrol.* **8**, 26–32 (2013).

108. Hojs Fabjan, T., Penko, M. & Hojs, R. Cystatin C, creatinine, estimated glomerular filtration, and long-term mortality in stroke patients. *Ren. Fail.* **36**, 81–86 (2014).

109. Smith, G. L. *et al.* Serum urea nitrogen, creatinine, and estimators of renal function: mortality in older patients with cardiovascular disease. *Arch. Intern. Med.* **166**, 1134–1142 (2006).

110. Mainous III, A. G., Everett, C. J., Liszka, H., King, D. E. & Egan, B. M. Prehypertension and mortality in a nationally representative cohort. *Am. J. Cardiol.* **94**, 1496–1500 (2004).

111. Aronson, D. *et al.* Serum blood urea nitrogen and long-term mortality in acute ST-elevation myocardial infarction. *Int. J. Cardiol.* **127**, 380–385 (2008).

112. Zaccardi, F. *et al.* Nonlinear association of BMI with all-cause and cardiovascular mortality in type 2 diabetes mellitus: a systematic review and meta-analysis of 414,587 participants in prospective studies. *Diabetologia* **60**, 240–248 (2017).

113. Shafi, T. *et al.* Comparing the association of GFR estimated by the CKD-EPI and MDRD study equations and mortality: the third national health and nutrition examination survey (NHANES III). *BMC Nephrol.* **13**, 42 (2012).

114. Wu, C.-K. *et al.* Cystatin C and long-term mortality among subjects with normal creatinine-based estimated glomerular filtration rates: NHANES III (Third National Health and Nutrition Examination Survey). *J. Am. Coll. Cardiol.* **56**, 1930–1936 (2010).

115. Fang, J. Pulse pressure: a predictor of cardiovascular mortality among young normotensive subjects. *Blood Press.* **9**, 260–266 (2000).

116. Oterdoom, L. H. *et al.* Urinary creatinine excretion, an indirect measure of muscle mass, is an independent predictor of cardiovascular disease and mortality in the general population. *Atherosclerosis* **207**, 534–540 (2009).

117. Song, X. *et al.* Cardiovascular and all-cause mortality in relation to various anthropometric measures of obesity in Europeans. *Nutr. Metab. Cardiovasc. Dis.* **25**, 295–304 (2015).

118. Hawkins, D. M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **44**, 1–12 (2004).

119. Nguyen, V. K., Colacino, J. A., Arnot, J. A., Kvasnicka, J. & Jolliet, O. Characterization of age-based trends to identify chemical biomarkers of higher levels in children. *Environ. Int.* **122**, (2019).

120. Nguyen, V. K. *et al.* A comprehensive analysis of racial disparities in chemical biomarker concentrations in United States women, 1999–2014. *Environ. Int.* **137**, 105496 (2020).

121. Xue, J., Liu, S. V., Zartarian, V. G., Geller, A. M. & Schultz, B. D. Analysis of NHANES measured blood PCBs i. The general US population and application of SHEDS model to identify key exposure factors. *J. Expo. Sci. Environ. Epidemiol.* (2014) doi:10.1038/jes.2013.91.

122. Calafat, A. M., Wong, L.-Y., Kuklenyik, Z., Reidy, J. A. & Needham, L. L. Polyfluoroalkyl Chemicals in the U.S. Population: Data from the National Health and Nutrition Examination Survey (NHANES) 2003–2004 and Comparisons with NHANES 1999–2000. *Environ. Health Perspect.* **115**, 1596–1602 (2007).

123. Kato, K., Wong, L.-Y., Jia, L. T., Kuklenyik, Z. & Calafat, A. M. Trends in Exposure to Polyfluoroalkyl Chemicals in the U.S. Population: 1999−2008. *Environ. Sci. Technol.* **45**, 8037–8045 (2011).

124. Trudel, D. *et al.* Estimating consumer exposure to PFOS and PFOA. *Risk Anal. An Int. J.* **28**, 251–269 (2008).

125. Mondal, D., Lopez-Espinosa, M.-J., Armstrong, B., Stein, C. R. & Fletcher, T. Relationships of perfluorooctanoate and perfluorooctane sulfonate serum concentrations between mother–child pairs in a population with perfluorooctanoate exposure from drinking water. *Environ. Health Perspect.* **120**, 752–757 (2012).

126. Gyllenhammar, I. *et al.* Perfluoroalkyl acids (PFAAs) in serum from 2–4-month-old infants: Influence of maternal serum concentration, gestational age, breast-feeding, and contaminated drinking water. *Environ. Sci. Technol.* **52**, 7101–7110 (2018).

127. Schecter, A. *et al.* Perfluorinated compounds, polychlorinated biphenyls, and organochlorine pesticide contamination in composite food samples from Dallas, Texas, USA. *Environ. Health Perspect.* **118**, (2010).

128. Richter, P. A., Bishop, E. E., Wang, J. & Swahn, M. H. Tobacco smoke exposure and levels of urinary metals in the US youth and adult population: the National Health and Nutrition Examination Survey (NHANES) 1999–2004. *Int. J. Environ. Res. Public Health* **6**, 1930–1946 (2009).

129. Silva, M. J., Jia, T., Samandar, E., Preau, J. L. J. & Calafat, A. M. Environmental exposure to the plasticizer 1,2-cyclohexane dicarboxylic acid, diisononyl ester (DINCH) in U.S. adults (2000-2012). *Environ. Res.* **126**, 159–163 (2013).

130. JL, P., DJ, B., EW, G. & al, et. The decline in blood lead levels in the united states: The national health and nutrition examination surveys (nhanes). *JAMA* **272**, 284–291 (1994).

131. CDC. Fourth National Report on Human Exposure to Environmental Chemicals. 1–519 (2009).

132. Hornung, R. W. & Reed, L. D. Estimation of Average Concentration in the Presence of Nondetectable Values. *Appl. Occup. Environ. Hyg.* **5**, 46–51 (1990).

133. Arnot, J. A., Brown, T. N. & Wania, F. Estimating Screening-Level Organic Chemical Half-Lives in Humans. *Environ. Sci. Technol.* **48**, 723–730 (2014).

134. OECD. *The report from the expert group on (quantitative) structure-activity relationships [(q) sars] on the principles for the validation of (q) sars.* (2004).

135. OECD. *Guidance document on the validation of (quantitative) structure-activity relationship [(q)sar] models.* (2014).

136. Fowler, T. A Brief History of Lead Regulation. *Science Progress Where Science, Technology, and Policy Meet* https://scienceprogress.org/2008/10/a-brief-history-of-lead-regulation/ (2008).

137. Newell, R. & Rogers, K. *The U.S. Experience With the Phasedown of Lead in Gasoline*. (2003).

138. FDA. Limiting Lead in Lipstick and Other Cosmetics. https://www.fda.gov/Cosmetics/ProductsIngredients/Products/ucm137224.htm#initial_survey (2018).

139. US EPA. *Basic Questions and Answers for the Drinking Water Strategy Contaminant Groups Effort*. https://www.epa.gov/dwstandardsregulations (2011).

140. US EPA. Drinking Water Contaminants – Standards and Regulations. https://www.epa.gov/dwstandardsregulations/use-lead-free-pipes-fittings-fixtures-solder-and-flux-drinking-water (2017).

141. NCHS. Continuous NHANES Web Tutorial Sample Design. (2010).

142. Stevens, R. G., Jones, D. Y., Micozzi, M. S. & Taylor, P. R. Body Iron Stores and the Risk of Cancer. *N. Engl. J. Med.* **319**, 1047–1052 (1988).

143. Korn, E. L. & Graubard, B. I. Epidemiologic studies utilizing surveys: accounting for the sampling design. *Am. J. Public Health* **81**, 1166–1173 (1991).

144. Royston, P. Lowess Smoothing. *Stata Tech. Bull.* **1**, (1992).

145. United States Census Bureau. QuickFacts UNITED STATES. https://www.census.gov/quickfacts/fact/table/US/PST045217 (2016).

146. US EPA. *Perfluoroalkyl Sulfonates; Significant New Use Rule*. (2003).

147. US EPA. *Standards of Performance for New Stationary Sources and Emission Guidelines for Existing Sources, Municipal Waste Combustors*. (2007).

148. 3M Company. *Voluntary use and exposure information profile for perfluorooctanoic acid and salts*. (2000).

149. Kärrman, A. *et al.* Exposure of perfluorinated chemicals through lactation: levels of matched human milk and serum and a temporal trend, 1996–2004, in Sweden. *Environ. Health Perspect.* **115**, 226–230 (2007).

150. Mogensen, U. B., Grandjean, P., Nielsen, F., Weihe, P. & Budtz-Jørgensen, E. Breastfeeding as an exposure pathway for perfluorinated alkylates. *Environ. Sci. Technol.*

49, 10466–10473 (2015).

151. Thomsen, C. *et al.* Changes in concentrations of perfluorinated compounds, polybrominated diphenyl ethers, and polychlorinated biphenyls in Norwegian breast-milk during twelve months of lactation. *Environ. Sci. Technol.* **44**, 9550–9556 (2010).

152. Guney, M. & Zagury, G. J. Contamination by ten harmful elements in toys and children's jewelry bought on the North American market. *Environ. Sci. Technol.* **47**, 5921–5930 (2013).

153. Dixon, S. L. *et al.* Exposure of US children to residential dust lead, 1999–2004: II. The contribution of lead-contaminated dust to children's blood lead levels. *Environ. Health Perspect.* **117**, 468–474 (2009).

154. Lanphear, B. P. *et al.* The contribution of lead-contaminated house dust and residential soil to children's blood lead levels. A pooled analysis of 12 epidemiologic studies. *Environ. Res. YORK-* **79**, 51–68 (1998).

155. Bhattacharyya, M. H. Bioavailability of orally administered cadmium and lead to the mother, fetus, and neonate during pregnancy and lactation: an overview. *Sci. Total Environ.* **28**, 327–342 (1983).

156. Silbergeld, E. K. Lead in bone: implications for toxicology during pregnancy and lactation. *Environ. Health Perspect.* **91**, 63–70 (1991).

157. ATSDR. *TOXICOLOGICAL PROFILE FOR TUNGSTEN*. https://www.atsdr.cdc.gov/toxprofiles/tp186.pdf (2005).

158. Kampmann, C. *et al.* Biodegradation of tungsten embolisation coils used in children. *Pediatr. Radiol.* **32**, 839–843 (2002).

159. Healthy Building Network. Pharos Project. https://www.pharosproject.net/ (2017).

160. Hileman, B. California Bans Phthalates In Toys For Children. *Chemical & Engineering News* (2007).

161. Royce, S. E. & Needleman, H. L. Case studies in environmental medicine: lead toxicity. *ATSDR. Atlanta US Dep. Heal. Hum. Serv.* (1992).

162. Api, A. M. Toxicological profile of diethyl phthalate: a vehicle for fragrance and cosmetic ingredients. *Food Chem. Toxicol.* **39**, 97–108 (2001).

163. Calafat, A. M., Ye, X., Wong, L.-Y., Bishop, A. M. & Needham, L. L. Urinary Concentrations of Four Parabens in the U.S. Population: NHANES 2005–2006. *Environ. Health Perspect.* **118**, 679–685 (2010).

164. Freedman, R. J. Reflections on Beauty as It Relates to Health in Adolescent Females. *Women Health* **9**, 29–45 (1984).

165. Gentina, E., Palan, K. M. & Fosse-Gomez, M.-H. The practice of using makeup: A consumption ritual of adolescent girls. *J. Consum. Behav.* **11**, 115–123 (2012).

166. Zota, A. R. & Shamasunder, B. The environmental injustice of beauty: framing chemical exposures from beauty products as a health disparities concern. *Am. J. Obstet. Gynecol.* **217**, 418.e1-418.e6 (2017).

167. Culhane, J. F. & Goldenberg, R. L. Racial Disparities in Preterm Birth. *Seminars in Perinatology* (2011) doi:10.1053/j.semperi.2011.02.020.

168. Marcinkevage, J. A. *et al.* Race/Ethnicity Disparities in Dysglycemia Among U.S. Women of Childbearing Age Found Mainly in the Nonoverweight/Nonobese. *Diabetes Care* **36**, 3033 LP – 3039 (2013).

169. Cowie, C. C. *et al.* Full accounting of diabetes and pre-diabetes in the U.S. population in 1988-1994 and 2005-2006. *Diabetes Care* (2009) doi:10.2337/dc08-1296.

170. Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V. & Sun, P. Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA Oncol.* (2015) doi:10.1001/jamaoncol.2015.2510.

171. DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A. & Jemal, A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA. Cancer J. Clin.* (2017) doi:10.3322/caac.21412.

172. Mayeda, E. R., Glymour, M. M., Quesenberry, C. P. & Whitmer, R. A. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimer's Dement.* (2016) doi:10.1016/j.jalz.2015.12.007.

173. Thorpe, L. E. *et al.* Trends and Racial/Ethnic Disparities in Gestational Diabetes Among Pregnant Women in New York City, 1990–2001. *Am. J. Public Health* **95**, 1536–1539 (2005).

174. Tanaka, M. *et al.* Racial Disparity in Hypertensive Disorders of Pregnancy in New York State: A 10-Year Longitudinal Population-Based Study. *Am. J. Public Health* **97**, 163–170 (2007).

175. Kaufman, J. S., Dolman, L., Rushani, D. & Cooper, R. S. The contribution of genomic research to explaining racial disparities in cardiovascular disease: A systematic review. *Am. J. Epidemiol.* (2015) doi:10.1093/aje/kwu319.

176. Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* (2011) doi:10.1038/nature10405.

177. Lim, S. S. *et al.* A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* (2012) doi:10.1016/S0140-6736(12)61766-8.

178. Zota, A. R. & Shamasunder, B. The environmental injustice of beauty: framing chemical exposures from beauty products as a health disparities concern. *Am. J. Obstet. Gynecol.* **217**, 418.e1-418.e6 (2017).

179. Ruiz, D., Becerra, M., Jagai, J. S., Ard, K. & Sargis, R. M. Disparities in environmental exposures to endocrine-disrupting chemicals and diabetes risk in vulnerable populations. *Diabetes Care* (2018) doi:10.2337/dc16-2765.

180. Wang, A., Padula, A., Sirota, M. & Woodruff, T. J. Environmental influences on reproductive health: the importance of chemical exposures. *Fertility and Sterility* (2016) doi:10.1016/j.fertnstert.2016.07.1076.

181. Juarez, P. D. & Matthews-Juarez, P. Applying an Exposome-Wide (ExWAS) Approach to Cancer Research. *Front. Oncol.* (2018) doi:10.3389/fonc.2018.00313.

182. Hoover, E. *et al.* Indigenous peoples of North America: Environmental exposures and reproductive justice. *Environmental Health Perspectives* (2012) doi:10.1289/ehp.1205422.

183. NCHS. Module 3: Weighting. https://wwwn.cdc.gov/nchs/nhanes/tutorials/Module3.aspx (2018).

184. American Society of Hematology. Anemia. https://www.hematology.org/Patients/Anemia/ (2020).

185. Nguyen, V. K., Colacino, J. A., Arnot, J. A., Kvasnicka, J. & Jolliet, O. Characterization of age-based trends to identify chemical biomarkers of higher levels in children. *Environ. Int.* (2019) doi:10.1016/j.envint.2018.10.042.

186. Brulle, R. J. & Pellow, D. N. ENVIRONMENTAL JUSTICE: Human Health and Environmental Inequalities. *Annu. Rev. Public Health* (2005) doi:10.1146/annurev.publhealth.27.021405.102124.

187. Landrigan, P. J., Rauh, V. A. & Galvez, M. P. Environmental justice and the health of children. *Mt. Sinai J. Med.* (2010) doi:10.1002/msj.20173.

188. Belova, A., Greco, S. L., Riederer, A. M., Olsho, L. E. W. W. & Corrales, M. A. A method to screen U.S. environmental biomonitoring data for race/ethnicity and income-related disparity. *Environ. Heal. A Glob. Access Sci. Source* **12**, 114 (2013).

189. Nelson, J. W., Scammell, M. K., Hatch, E. E. & Webster, T. F. Social disparities in exposures to bisphenol A and polyfluoroalkyl chemicals: A cross-sectional study within NHANES 2003-2006. *Environ. Heal. A Glob. Access Sci. Source* (2012) doi:10.1186/1476-069X-11-10.

190. Hu, X. C. *et al.* Detection of Poly- and Perfluoroalkyl Substances (PFASs) in U.S. Drinking Water Linked to Industrial Sites, Military Fire Training Areas, and Wastewater Treatment Plants. *Environ. Sci. Technol. Lett.* (2016) doi:10.1021/acs.estlett.6b00260.

191. Laitinen, J. A., Koponen, J., Koikkalainen, J. & Kiviranta, H. Firefighters' exposure to perfluoroalkyl acids and 2-butoxyethanol present in firefighting foams. *Toxicol. Lett.* (2014) doi:10.1016/j.toxlet.2014.09.007.

192. Chen, D. *et al.* Bisphenol Analogues Other Than BPA: Environmental Occurrence, Human Exposure, and Toxicity - A Review. *Environmental Science and Technology* (2016) doi:10.1021/acs.est.5b05387.

193. Helm, J. S., Nishioka, M., Brody, J. G., Rudel, R. A. & Dodson, R. E. Measurement of endocrine disrupting and asthma-associated chemicals in hair products used by Black women. *Environ. Res.* (2018) doi:10.1016/j.envres.2018.03.030.

194. López-Carrillo, L. *et al.* Exposure to Phthalates and Breast Cancer risk in Northern Mexico. *Environ. Health Perspect.* (2010) doi:10.1289/ehp.0901091.

195. Gonzalez, T. L., Rae, J. M., Colacino, J. A. & Richardson, R. J. Homology models of mouse and rat estrogen receptor-α ligand-binding domain created by in silico mutagenesis of a human template: Molecular docking with 17β-estradiol, diethylstilbestrol, and paraben analogs. *Comput. Toxicol.* (2019) doi:10.1016/j.comtox.2018.11.003.

196. Gonzalez, T. L. *et al.* Metabolites of n-butylparaben and iso-butylparaben exhibit estrogenic properties in MCF-7 and T47D human breast cancer cell lines. *Toxicol. Sci.* (2018) doi:10.1093/toxsci/kfy063.

197. Okubo, T., Yokoyama, Y., Kano, K. & Kano, I. ER-dependent estrogenic activity of parabens assessed by proliferation of human breast cancer MCF-7 cells and expression of ERα and PR. *Food Chem. Toxicol.* (2001) doi:10.1016/S0278-6915(01)00073-4.

198. Pan, S. *et al.* Parabens and Human Epidermal Growth Factor Receptor Ligand Cross-Talk in Breast Cancer Cells. *Environ. Health Perspect.* **124**, 563–569 (2016).

199. Gopalakrishnan, K. *et al.* Changes in Mammary Histology and Transcriptome Profiles by Low-Dose Exposure to Environmental Phenols at Critical Windows of Development(). *Environ. Res.* **152**, 233–243 (2017).

200. Stiel, L., Adkins-Jackson, P. B., Clark, P., Mitchell, E. & Montgomery, S. A review of hair product use on breast cancer risk in African American women. *Cancer Med.* (2016) doi:10.1002/cam4.613.

201. Singh, G. K., Kogan, M. D. & Dee, D. L. Nativity/immigrant status, race/ethnicity, and socioeconomic determinants of breastfeeding initiation and duration in the United States, 2003. *Pediatrics* **119**, S38–S46 (2007).

202. Hines, E. P. *et al.* Concentrations of environmental phenols and parabens in milk, urine and serum of lactating North Carolina women. *Reprod. Toxicol.* **54**, 120–128 (2015).

203. Ye, X., Wong, L.-Y., Zhou, X. & Calafat, A. M. Urinary concentrations of 2,4-dichlorophenol and 2,5-dichlorophenol in the U.S. population (National Health and

Nutrition Examination Survey, 2003-2010): trends and predictors. *Environ. Health Perspect.* **122**, 351–355 (2014).

204. Wei, Y. & Zhu, J. Para-Dichlorobenzene Exposure Is Associated with Thyroid Dysfunction in US Adolescents. *J. Pediatr.* (2016) doi:10.1016/j.jpeds.2016.06.085.

205. Hsiao, P. K., Lin, Y. C., Shih, T. S. & Chiung, Y. M. Effects of occupational exposure to 1,4-dichlorobenzene on hematologic, kidney, and liver functions. *Int. Arch. Occup. Environ. Health* (2009) doi:10.1007/s00420-009-0398-5.

206. Takahashi, O., Ohashi, N., Nakae, D. & Ogata, A. Parenteral paradichlorobenzene exposure reduces sperm production, alters sperm morphology and exhibits an androgenic effect in rats and mice. *Food Chem. Toxicol.* (2011) doi:10.1016/j.fct.2010.09.029.

207. Hegazy, A. A., Zaher, M. M., Abd el-hafez, M. A., Morsy, A. A. & Saleh, R. A. Relation between anemia and blood levels of lead, copper, zinc and iron among children. *BMC Res. Notes* **3**, 133 (2010).

208. Kwong, W. T., Friello, P. & Semba, R. D. Interactions between iron deficiency and lead poisoning: epidemiology and pathogenesis. *Sci. Total Environ.* **330**, 21–37 (2004).

209. Jacobs, D. E. *et al.* Lead and other heavy metals in dust fall from single-family housing demolition. *Public Health Rep.* (2013) doi:10.1177/003335491312800605.

210. Suvarapu, L. N. & Baek, S. O. Determination of heavy metals in the ambient atmosphere: A review. *Toxicology and Industrial Health* (2016) doi:10.1177/0748233716654827.

211. Pieper, K. J., Tang, M. & Edwards, M. A. Flint Water Crisis Caused by Interrupted Corrosion Control: Investigating 'ground Zero' Home. *Environ. Sci. Technol.* (2017) doi:10.1021/acs.est.6b04034.

212. Pumarega, J., Gasull, M., Lee, D. H., López, T. & Porta, M. Number of persistent organic pollutants detected at high concentrations in blood samples of the United States population. *PLoS One* (2016) doi:10.1371/journal.pone.0160432.

213. Alcala, C. S. & Phillips, L. J. PCB concentrations in women based on breastfeeding history: NHANES 2001–2004. *Environ. Res.* **154**, 35–41 (2017).

214. Wolff, M. S., Toniolo, P. G., Lee, E. W., Rivera, M. & Dubin, N. Blood levels of organochlorine residues and risk of breast cancer. *J. Natl. Cancer Inst.* (1993) doi:10.1093/jnci/85.8.648.

215. Cohn, B. A. *et al.* DDT exposure in utero and breast cancer. *J. Clin. Endocrinol. Metab.* (2015) doi:10.1210/jc.2015-1841.

216. Hansen, S. *et al.* Prenatal exposure to persistent organic pollutants and offspring allergic sensitization and lung function at 20 years of age. *Clin. Exp. Allergy* (2016) doi:10.1111/cea.12631.

217. Dewailly, É. *et al.* Susceptibility to infections and immune status in Inuit infants exposed to organochlorines. *Environ. Health Perspect.* (2000) doi:10.1289/ehp.00108205.

218. Sagiv, S. K. *et al.* Prenatal organochlorine exposure and behaviors associated with attention deficit hyperactivity disorder in school-aged children. *Am. J. Epidemiol.* (2010) doi:10.1093/aje/kwp427.

219. Mendez, M. A. *et al.* Prenatal organochlorine compound exposure, rapid weight gain, and overweight in infancy. *Environ. Health Perspect.* (2011) doi:10.1289/ehp.1002169.

220. Park, S. L. *et al.* Genetic determinants of CYP2A6 activity across racial/ethnic groups with different risks of lung cancer and effect on their smoking intensity. *Carcinogenesis* **37**, 269–279 (2016).

221. James, K. A. *et al.* A case-cohort study examining lifetime exposure to inorganic arsenic in drinking water and diabetes mellitus. *Environ. Res.* **123**, 33–38 (2013).

222. Wang, S.-L. *et al.* Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. *Environ. Int.* **33**, 805–811 (2007).

223. Balakrishnan, P. *et al.* Ethnic, geographic, and genetic differences in arsenic metabolism at low arsenic exposure: a preliminary analysis in the multi-ethnic study of atherosclerosis (mesa). *Int. J. Environ. Res. Public Health* **15**, 1179 (2018).

224. Newman, L. A. *et al.* Hereditary susceptibility for triple negative breast cancer associated with Western Sub-Saharan African ancestry: Results from an International Surgical Breast Cancer Collaborative. *Ann. Surg.* **270**, 484–492 (2019).

225. Jemal, A. & Fedewa, S. A. Is the prevalence of ER-negative breast cancer in the US higher among Africa-born than US-born black women? *Breast Cancer Res. Treat.* **135**, 867–873 (2012).

226. Stark, A. *et al.* African ancestry and higher prevalence of triple-negative breast cancer. *Cancer* **116**, 4926–4932 (2010).

227. Tournamille, C., Le Van Kim, C., Gane, P., Cartron, J.-P. & Colin, Y. Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. *Hum. Genet.* **95**, 407–410 (1995).

228. Hansell, C. A. H., Hurson, C. E. & Nibbs, R. J. B. DARC and D6: silent partners in chemokine regulation? *Immunol. Cell Biol.* **89**, 197–206 (2011).

229. Akalin, E. & Neylan, J. F. The influence of Duffy blood group on renal allograft outcome in African Americans. *Transplantation* **75**, 1496–1500 (2003).

230. Wattigney, W. A., Irvin-Barnwell, E., Pavuk, M. & Ragin-Wilson, A. Regional Variation in Human Exposure to Persistent Organic Pollutants in the United States, NHANES. *J. Environ. Public Health* (2015) doi:10.1155/2015/571839.

231. Barr, D. B. *et al.* Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. *Environ. Health Perspect.* **113**, 192–200 (2005).

232. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* (2019) doi:10.1016/j.cell.2019.02.048.

233. Stanaway, J. D. *et al.* Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Stu. *Lancet* **392**, 1923–1994 (2018).

234. Kijko, G., Jolliet, O. & Margni, M. Occupational Health Impacts Due to Exposure to Organic Chemicals over an Entire Product Life Cycle. *Environ. Sci. Technol.* **50**, 13105–13114 (2016).

235. Mater, G., Paris, C. & Lavoué, J. Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA. *Am. J. Ind. Med.* **59**, 379–391 (2016).

236. CDC. Metals - Urine (L06HM_C). (2007).

237. CDC. Polyfluoroalkyl Chemicals (L24PFC_C). https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/L24PFC_C.htm (2007).

238. Soley-Bori, M. Dealing with missing data: Key assumptions and methods for applied analysis. *Bost. Univ.* **4**, 1–19 (2013).

239. Duan, W. *et al.* Levels of a mixture of heavy metals in blood and urine and all-cause, cardiovascular disease and cancer mortality: A population-based cohort study. *Environ. Pollut.* 114630 (2020).

240. Sanders, A. P. *et al.* Combined exposure to lead, cadmium, mercury, and arsenic and kidney health in adolescents age 12–19 in NHANES 2009–2014. *Environ. Int.* **131**, 104993 (2019).

241. Bosson-Rieutort, D., Sarazin, P., Bicout, D. J., Ho, V. & Lavoué, J. Occupational Co-exposures to Multiple Chemical Agents from Workplace Measurements by the US Occupational Safety and Health Administration. *Ann. Work Expo. Heal.* **64**, 402–415 (2020).

242. Kapraun, D. F., Wambaugh, J. F., Ring, C. L., Tornero-Velez, R. & Setzer, R. W. A method for identifying prevalent chemical combinations in the US population. *Environ. Health Perspect.* **125**, 87017 (2017).

243. Kinnunen, M.-L., Feldt, T., Kinnunen, U., Kaprio, J. & Pulkkinen, L. Association between long-term job strain and metabolic syndrome factor across sex and occupation. *J. Individ. Differ.* **27**, 151–161 (2006).

244. Winkleby, M. A., Jatulis, D. E., Frank, E. & Fortmann, S. P. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am. J. Public Health* **82**, 816–820 (1992).

245. Engström, G., Hedblad, B., Rosvall, M., Janzon, L. & Lindgärde, F. Occupation, marital status, and low-grade inflammation: mutual confounding or independent cardiovascular risk factors? *Arterioscler. Thromb. Vasc. Biol.* **26**, 643–648 (2006).

246. Harford, T. C. & Brooks, S. D. Cirrhosis mortality and occupation. *J. Stud. Alcohol* **53**, 463–468 (1992).

247. Morris, C. E. *et al.* Effect of a simulated tactical occupation task on physiological strain index, stress and inflammation. *Int. J. Occup. Saf. Ergon.* (2018).

248. U.S. Department of Labor. Fact Sheet #17I: Blue-Collar Workers and the Part 541 Exemptions Under the Fair Labor Standards Act (FLSA). (2019).

249. NCHS. Metals - urine (lab06hm). (2005).

250. NCHS. Metals - urine (l06hm_b). (2005).

251. Everitt, B. S., Landau, S., Leese, M. & Stahl, D. Hierarchical clustering. *Clust. Anal.* **5**, 71–110 (2011).

252. Freedman, D. A. *Statistical models: theory and practice*. (cambridge university press, 2009).

253. Székely, G. J., Rizzo, M. L. & Bakirov, N. K. Measuring and testing dependence by correlation of distances. *Ann. Stat.* **35**, 2769–2794 (2007).

254. Saraçli, S., Doğan, N. & Doğan, İ. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J. Inequalities Appl.* **2013**, 203 (2013).

255. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

256. McComb, R. B., Bowers Jr, G. N. & Posen, S. *Alkaline phosphatase*. (Springer Science & Business Media, 2013).

257. Maton, A. *Human biology and health*. (Prentice Hall, 1997).

258. Thompson, D., Pepys, M. B. & Wood, S. P. The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* **7**, 169–177 (1999).

259. Lebovitz, H. E. Insulin resistance: definition and consequences. *Exp. Clin. Endocrinol. diabetes* **109**, S135–S148 (2001).

260. Levey, A. S. *et al.* National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ann. Intern. Med.* **139**, 137–147

(2003).

261. Frisancho, A. R. Triceps skin fold and upper arm muscle size norms for assessment of nutritional status. *Am. J. Clin. Nutr.* **27**, 1052–1058 (1974).

262. Eaton–Evans, J. NUTRITIONAL ASSESSMENT| Anthropometry. (2005).

263. CDC. LEAD: Jobs That May Have Lead Exposure. https://www.cdc.gov/niosh/topics/lead/jobs.html (2018).

264. Miao, H. & Ji, J. S. Trends of Blood Cadmium Concentration Among Workers and Non-Workers in the United States (NHANES 2003 to 2012). *J. Occup. Environ. Med.* **61**, e503–e509 (2019).

265. Raymond, C. V. J. Estimating the lung deposition of particulate polycyclic aromatic hydrocarbons associated with multimodal urban aerosols. *Inhal. Toxicol.* **10**, 183–204 (1998).

266. VanRooij, J. G. M., De Roos, J. H. C., Bodelier-Bade, M. M. & Jongeneelen, F. J. Absorption of polycyclic aromatic hydrocarbons through human skin: differences between anatomical sites and individuals. *J. Toxicol. Environ. Heal. Part A Curr. Issues* **38**, 355–368 (1993).

267. VanRooij, J. G., Bodelier-Bade, M. M. & Jongeneelen, F. J. Estimation of individual dermal and respiratory uptake of polycyclic aromatic hydrocarbons in 12 coke oven workers. *Occup. Environ. Med.* **50**, 623–632 (1993).

268. Toxic-Free Future. PFAS: A Nonstick Nightmare. https://toxicfreefuture.org/key-issues/chemicals-of-concern/pfas-nonstick-nightmare/ (2020).

269. Minnesota Department of Health. Volatile Organic Compounds in Your Home. https://www.health.state.mn.us/communities/environment/air/toxins/voc.htm (2020).

270. Baker, R. R. & Bishop, L. J. The pyrolysis of tobacco ingredients. *J. Anal. Appl. Pyrolysis* **71**, 223–311 (2004).

271. Fowles, J. & Dybing, E. Application of toxicological risk assessment principles to the chemical constituents of cigarette smoke. *Tob. Control* **12**, 424–430 (2003).

272. Vu, A. T. *et al.* Polycyclic aromatic hydrocarbons in the mainstream smoke of popular US cigarettes. *Chem. Res. Toxicol.* **28**, 1616–1626 (2015).

273. CDC. Arsenic. https://www.cdc.gov/niosh/topics/arsenic/default.html (2019).

274. CDC. Mercury. (2019).

275. Trzcinka-Ochocka, M., Gazewski, A. & Brodzka, R. Exposure to mercury vapors in dental workers in Poland. *Int. J. Occup. Med. Environ. Health* **20**, 147–154 (2007).

276. Shimshack, J. P., Ward, M. B. & Beatty, T. K. M. Mercury advisories: information, education, and fish consumption. *J. Environ. Econ. Manage.* **53**, 158–179 (2007).

277. Tyrrell, J., Melzer, D., Henley, W., Galloway, T. S. & Osborne, N. J. Associations between socioeconomic status and environmental toxicant concentrations in adults in the USA: NHANES 2001–2010. *Environ. Int.* **59**, 328–335 (2013).

278. Program, N. T. Toxicology and carcinogenesis studies of benzophenone (CAS No. 119-61-9) in F344/N rats and B6C3F1 mice (feed studies). *Natl. Toxicol. Progr. Tech. Rep. Ser* **533**, 1–264 (2006).

279. Benzophenone, H. Hazardous substance database. (2013).

280. CDC. Benzophenone-3 (BP-3) Factsheet. (2017).

281. Kim, J.-H., Lee, H. S., Park, H.-M. & Lee, Y.-J. Serum alkaline phosphatase level is positively associated with metabolic syndrome: A nationwide population-based study. *Clin. Chim. Acta* **500**, 189–194 (2020).

282. Nakata, A. Psychosocial job stress and immunity: a systematic review. in *Psychoneuroimmunology* 39–75 (Springer, 2012).

283. Nakata, A., Irie, M. & Takahashi, M. A single-item global job satisfaction measure is associated with quantitative blood immune indices in white-collar employees. *Ind. Health* **51**, 193–201 (2013).

284. Eubank, R. L. *Spline smoothing and nonparametric regression*. vol. 90 (M. Dekker New York, 1988).

285. Sánchez-Chaparro, M.-A. *et al.* Estimating the glomerular filtration rate in the Spanish working population: chronic kidney disease prevalence and its association with risk factors. *J. Hypertens.* **32**, 1970–1978 (2014).

286. Gu, J. K. *et al.* Prevalence of obesity by occupation among US workers: the National Health Interview Survey 2004–2011. *J. Occup. Environ. Med. Coll. Occup. Environ. Med.* **56**, 516 (2014).

287. Gupta, N., Heiden, M., Mathiassen, S. E. & Holtermann, A. Prediction of objectively measured physical activity and sedentariness among blue-collar workers using survey questionnaires. *Scand. J. Work. Environ. Health* 237–245 (2016).

288. Vandelanotte, C. *et al.* Associations between occupational indicators and total, work-based and leisure-time sitting: a cross-sectional study. *BMC Public Health* **13**, 1110 (2013).

289. Venkatasamy, V. V., Pericherla, S., Manthuruthil, S., Mishra, S. & Hanno, R. Effect of physical activity on insulin resistance, inflammation and oxidative stress in diabetes mellitus. *J. Clin. diagnostic Res. JCDR* **7**, 1764 (2013).

290. Kundu, S., Sengupta, S., Chatterjee, S., Mitra, S. & Bhattacharyya, A. Cadmium induces lung inflammation independent of lung cell proliferation: a molecular approach. *J. Inflamm.* **6**, 19 (2009).

291. Manca, D., Ricard, A. C., Van Tra, H. & Chevalier, G. Relation between lipid peroxidation and inflammation in the pulmonary toxicity of cadmium. *Arch. Toxicol.* **68**, 364–369 (1994).

292. Blum, J. L. *et al.* Short-term inhalation of cadmium oxide nanoparticles alters pulmonary dynamics associated with lung injury, inflammation, and repair in a mouse model. *Inhal. Toxicol.* **26**, 48–58 (2014).

293. Colacino, J. A., Arthur, A. E., Ferguson, K. K. & Rozek, L. S. Dietary antioxidant and anti-inflammatory intake modifies the effect of cadmium exposure on markers of systemic inflammation and oxidative stress. *Environ. Res.* **131**, 6–12 (2014).

294. Knoflach, M. *et al.* Non-toxic cadmium concentrations induce vascular inflammation and promote atherosclerosis. *Circ. J.* **75**, 2491–2495 (2011).

295. Metryka, E. *et al.* Lead (Pb) exposure enhances expression of factors associated with inflammation. *Int. J. Mol. Sci.* **19**, 1813 (2018).

296. Boskabaddy, M. H. & Farkhondeh, T. Inhaled lead exposure affects tracheal responsiveness and lung inflammation in guinea pigs during sensitization. *Biol. Trace Elem. Res.* **154**, 363–371 (2013).

297. Ghareeb, D. A., Hussien, H. M., Khalil, A. A., El-Saadani, M. A. & Ali, A. N. Toxic effects of lead exposure on the brain of rats: Involvement of oxidative stress, inflammation, acetylcholinesterase, and the beneficial role of flaxseed extract. *Toxicol. Environ. Chem.* **92**, 187–195 (2010).

298. Songdej, N., Winters, P. C., McCabe Jr, M. J. & van Wijngaarden, E. A population-based assessment of blood lead levels in relation to inflammation. *Environ. Res.* **110**, 272–277 (2010).

299. Schober, W. *et al.* Environmental polycyclic aromatic hydrocarbons (PAHs) enhance allergic inflammation by acting on human basophils. *Inhal. Toxicol.* **19**, 151–156 (2007).

300. Zhang, H. *et al.* Association of internal exposure to polycyclic aromatic hydrocarbons with inflammation and oxidative stress in prediabetic and healthy individuals. *Chemosphere* 126748 (2020).

301. Shen, M. *et al.* Declining pulmonary function in populations with long-term exposure to polycyclic aromatic hydrocarbons-enriched PM2. 5. *Environ. Sci. Technol.* **52**, 6610–6616 (2018).

302. Brucker, N. *et al.* Biomarkers of occupational exposure to air pollution, inflammation and oxidative damage in taxi drivers. *Sci. Total Environ.* **463**, 884–893 (2013).

303. Gordon, T., Sheppard, D., McDonald, D. M., Distefano, S. & Scypinski, L. Airway hyperresponsiveness and inflammation induced by toluene diisocyanate in guinea pigs. *Am. Rev. Respir. Dis.* **132**, 1106–1112 (1985).

304. Saetta, M. *et al.* Airway mucosal inflammation in occupational asthma induced by toluene diisocyanate. *Am. Rev. Respir. Dis.* **145**, 160–168 (1992).

305. Jiménez-Garza, O. *et al.* Promoter methylation status in genes related with inflammation, nitrosative stress and xenobiotic metabolism in low-level benzene exposure: Searching for biomarkers of oncogenesis. *Food Chem. Toxicol.* **109**, 669–676 (2017).

306. Triggiani, M., Loffredo, S., Granata, F., I Staiano, R. & Marone, G. Modulation of mast cell and basophil functions by benzene metabolites. *Curr. Pharm. Des.* **17**, 3830–3835 (2011).

307. García-Niño, W. R. & Pedraza-Chaverrí, J. Protective effect of curcumin against heavy metals-induced liver damage. *Food Chem. Toxicol.* **69**, 182–201 (2014).

308. Eybl, V., Kotyzová, D. & Bludovská, M. The effect of curcumin on cadmium-induced oxidative damage and trace elements level in the liver of rats and mice. *Toxicol. Lett.* **151**, 79–85 (2004).

309. Salihovic, S. *et al.* Changes in markers of liver function in relation to changes in perfluoroalkyl substances-A longitudinal study. *Environ. Int.* **117**, 196–203 (2018).

310. DeWitt, J. *Toxicological effects of perfluoroalkyl and polyfluoroalkyl substances*. (Humana Press, 2015).

311. Liu, J., Drane, W., Liu, X. & Wu, T. Examination of the relationships between environmental exposures to volatile organic compounds and biochemical liver tests: application of canonical correlation analysis. *Environ. Res.* **109**, 193–199 (2009).

312. Van den Velde, S., Nevens, F., van Steenberghe, D. & Quirynen, M. GC–MS analysis of breath odor compounds in liver patients. *J. Chromatogr. B* **875**, 344–348 (2008).

313. Dadamio, J. *et al.* Breath biomarkers of liver cirrhosis. *J. Chromatogr. B* **905**, 17–22 (2012).

314. Probert, C. S. J. *et al.* Volatile organic compounds as diagnostic biomarkers in gastrointestinal and liver diseases. *J. Gastrointest. Liver Dis.* **18**, (2009).

315. Khalid, T., Richardson, P. & Probert, C. S. The liver breath! Breath volatile organic compounds for the diagnosis of liver disease. *Clin. Gastroenterol. Hepatol.* **12**, 524–526 (2014).

316. Wang, Q. & Xue, Y. Characterization of solid tumors induced by polycyclic aromatic hydrocarbons in mice. *Med. Sci. Monit. Basic Res.* **21**, 81 (2015).

317. Abdel-Shafy, H. I. & Mansour, M. S. M. A review on polycyclic aromatic hydrocarbons: source, environmental impact, effect on human health and remediation. *Egypt. J. Pet.* **25**,

107–123 (2016).

318. Milošević, N. *et al.* Could phthalates exposure contribute to the development of metabolic syndrome and liver disease in humans? *Environ. Sci. Pollut. Res.* **27**, 772–784 (2020).

319. Praveena, S. M. *et al.* Recent updates on phthalate exposure and human health: a special focus on liver toxicity and stem cell regeneration. *Environ. Sci. Pollut. Res.* **25**, 11333–11342 (2018).

320. Hwangbo, Y. *et al.* Blood cadmium and estimated glomerular filtration rate in Korean adults. *Environ. Health Perspect.* **119**, 1800–1805 (2011).

321. Shelley, R. *et al.* Associations of multiple metals with kidney outcomes in lead workers. *Occup. Environ. Med.* **69**, 727–735 (2012).

322. Goyer, R. A., Weinberg, C. R., Victery, W. M. & Miller, C. R. Lead induced nephrotoxicity: kidney calcium as an indicator of tubular injury. in *Nephrotoxicity* 11–20 (Springer, 1989).

323. Khalil-Manesh, F. *et al.* Experimental model of lead nephropathy. I. Continuous high-dose lead administration. *Kidney Int.* **41**, 1192–1203 (1992).

324. Blake, B. E., Pinney, S. M., Hines, E. P., Fenton, S. E. & Ferguson, K. K. Associations between longitudinal serum perfluoroalkyl substance (PFAS) levels and measures of thyroid hormone, kidney function, and body mass index in the Fernald Community Cohort. *Environ. Pollut.* **242**, 894–904 (2018).

325. Conway, B., Innes, K. E. & Long, D. Perfluoroalkyl substances and beta cell deficient diabetes. *J. Diabetes Complications* **30**, 993–998 (2016).

326. Conway, B. N., Badders, A. N., Costacou, T., Arthur, J. M. & Innes, K. E. Perfluoroalkyl substances and kidney function in chronic kidney disease, anemia, and diabetes. *Diabetes, Metab. Syndr. Obes. targets Ther.* **11**, 707 (2018).

327. Palatini, P. Glomerular hyperfiltration: a marker of early renal damage in pre-diabetes and pre-hypertension. (2012).

328. Shastri, S. & Sarnak, M. J. Chronic kidney disease: high eGFR and mortality: high true GFR or a marker of frailty? *Nat. Rev. Nephrol.* **7**, 680 (2011).

329. Chang, T.-Y. *et al.* Exposure to volatile organic compounds and kidney dysfunction in thin film transistor liquid crystal display (TFT-LCD) workers. *J. Hazard. Mater.* **178**, 934–940 (2010).

330. Hüppe, T. *et al.* Volatile organic compounds in patients with acute kidney injury and changes during dialysis. *Crit. Care Med.* **47**, 239–246 (2019).

331. Karami, S. *et al.* Renal cancer risk and occupational exposure to polycyclic aromatic hydrocarbons and plastics. *J. Occup. Environ. Med. Coll. Occup. Environ. Med.* **53**, 218

(2011).

332. Betteridge, D. J. What is oxidative stress? *Metab. Exp.* **49**, 3–8 (2000).

333. Farzan, S. F., Chen, Y., Trachtman, H. & Trasande, L. Urinary polycyclic aromatic hydrocarbons and measures of oxidative stress, inflammation and renal function in adolescents: NHANES 2003–2008. *Environ. Res.* **144**, 149–157 (2016).

334. Maeda, I. *et al.* Cigarette smoking and the association with glomerular hyperfiltration and proteinuria in healthy middle-aged men. *Clin. J. Am. Soc. Nephrol.* **6**, 2462–2469 (2011).

335. Antoniou, M. *et al.* Teratogenic effects of glyphosate-based herbicides: divergence of regulatory decisions from scientific evidence. *J Env. Anal Toxicol S* **4**, 525–2161 (2012).

336. Malits, J. *et al.* Renal Function and exposure to Bisphenol A and phthalates in children with Chronic Kidney Disease. *Environ. Res.* **167**, 575–582 (2018).

337. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and prediction*. (Springer Science & Business Media, 2009).

338. Cory-Slechta, D. A., Virgolini, M. B., Thiruchelvam, M., Weston, D. D. & Bauter, M. R. Maternal stress modulates the effects of developmental lead exposure. *Environ. Health Perspect.* **112**, 717–730 (2004).

339. McEwen, B. S. Allostasis and allostatic load: implications for neuropsychopharmacology. *Neuropsychopharmacology* **22**, 108–124 (2000).

340. Seeman, T. E., Singer, B. H., Rowe, J. W., Horwitz, R. I. & McEwen, B. S. Price of adaptation--allostatic load and its health consequences. MacArthur studies of successful aging. *Arch. Intern. Med.* **157**, 2259–2268 (1997).

341. Geronimus, A. T., Hicken, M., Keene, D. & Bound, J. "Weathering" and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States. *Am. J. Public Health* **96**, 826–833 (2006).

342. Horowitz, G. L., Altaie, S. & Boyd, J. C. *Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline*. (CLSI, 2010).

343. Rappoport, N. *et al.* Creating ethnicity-specific reference intervals for lab tests from EHR data. *bioRxiv* 213892 (2017).

344. Manrai, A. K., Patel, C. J. & Ioannidis, J. P. A. In the era of precision medicine and big data, who is normal? *Jama* **319**, 1981–1982 (2018).

345. Bangert, S. K. & Marshall, W. J. *Clinical biochemistry: metabolic and clinical aspects*. (Churchill Livingstone, 1995).

346. Zhu, S. *et al.* Waist circumference and obesity-associated risk factors among whites in the

third National Health and Nutrition Examination Survey: clinical action thresholds. *Am. J. Clin. Nutr.* **76**, 743 (2002).

347. Glassock, R. J. Con: Thresholds to define chronic kidney disease should not be age dependent. *Nephrol. Dial. Transplant. Off. Publ. Eur. Dial. Transpl. Assoc. Ren. Assoc.* **29**, 774 (2014).

348. World Health Organization. Physical status: use and interpretation of anthropometry; report of a WHO Expert Committee. Geneva: World Health Organization, 1995. 452. *WHO Tech. Rep. Ser.*

349. International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care* **32**, 1327–1334 (2009).

350. Consortium, C. K. D. P. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. *Lancet* **375**, 2073–2081 (2010).

351. Ruggiero, C. *et al.* White blood cell count and mortality in the Baltimore Longitudinal Study of Aging. *J. Am. Coll. Cardiol.* **49**, 1841–1850 (2007).

352. James, A. L. *et al.* Associations between white blood cell count, lung function, respiratory illness and mortality: the Busselton Health Study. *Eur. Respir. J.* **13**, 1115–1119 (1999).

353. Mazidi, M., Mikhailidis, D. P. & Banach, M. Associations between risk of overall mortality, cause-specific mortality and level of inflammatory factors with extremely low and high high-density lipoprotein cholesterol levels among American adults. *Int. J. Cardiol.* **276**, 242–247 (2019).

354. NCHS. NDI Mortality Data. *Centers for Disease Control and Prevention* https://www.cdc.gov/nchs/data-linkage/mortality.htm (2020).

355. Breslow, N. E. Analysis of survival data under the proportional hazards model. *Int. Stat. Rev. Int. Stat.* 45–57 (1975).

356. Durrleman, S. & Simon, R. Flexible regression models with cubic splines. *Stat. Med.* **8**, 551–561 (1989).

357. Picard, R. R. & Cook, R. D. Cross-validation of regression models. *J. Am. Stat. Assoc.* **79**, 575–583 (1984).

358. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).

359. Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *Jama* **247**, 2543–2546 (1982).

360. Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination.

Biometrika **78**, 691–692 (1991).

361. Tibshirani, R. J. & Efron, B. An introduction to the bootstrap. *Monogr. Stat. Appl. Probab.* **57**, 1–436 (1993).

362. Levey, A. S. *et al.* Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann. Intern. Med.* **145**, 247–254 (2006).

363. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).

364. Corti, M.-C. *et al.* HDL cholesterol predicts coronary heart disease mortality in older persons. *Jama* **274**, 539–544 (1995).

365. Li, Z.-H. *et al.* High-density lipoprotein cholesterol and all-cause and cause-specific mortality among the elderly. *J. Clin. Endocrinol. Metab.* **104**, 3370–3378 (2019).

366. Colpo, A. LDL Cholesterol:" Bad" Cholesterol or Bad Science? *J. Am. Physicians Surg.* **10**, 83 (2005).

367. Grundy, S. M. *et al.* 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J. Am. Coll. Cardiol.* **73**, e285–e350 (2019).

368. Olsson, A. G. *et al.* Can LDL cholesterol be too low? Possible risks of extremely low levels. *J. Intern. Med.* **281**, 534–553 (2017).

369. Kilpatrick, R. D. *et al.* Association between serum lipids and survival in hemodialysis patients and impact of race. *J. Am. Soc. Nephrol.* **18**, 293–303 (2007).

370. Vyroubal, P. *et al.* Hypocholesterolemia in clinically serious conditions--review. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **152**, 181–189 (2008).

371. Habib, A. N., Baird, B. C., Leypoldt, J. K., Cheung, A. K. & Goldfarb-Rumyantzev, A. S. The association of lipid levels with mortality in patients on chronic peritoneal dialysis. *Nephrol. Dial. Transplant.* **21**, 2881–2892 (2006).

372. Jacobs, D. *et al.* Report of the Conference on Low Blood Cholesterol: mortality associations. *Circulation* **86**, 1046–1060 (1992).

373. Chmielewski, M. *et al.* Low cholesterol in dialysis patients—causal factor for mortality or an effect of confounding? *Nephrol. Dial. Transplant.* **26**, 3325–3331 (2011).

374. NCHS. *Table 21. Selected health conditions and risk factors, by age: United States, selected years 1988–1994 through 2015–2016*. (2018).

375. Le Grange, D., Swanson, S. A., Crow, S. J. & Merikangas, K. R. Eating disorder not otherwise specified presentation in the US population. *Int. J. Eat. Disord.* **45**, 711–718 (2012).

376. Hudson, J. I., Hiripi, E., Pope Jr, H. G. & Kessler, R. C. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biol. Psychiatry* **61**, 348–358 (2007).

377. Flegal, K. M., Graubard, B. I., Williamson, D. F. & Gail, M. H. Excess deaths associated with underweight, overweight, and obesity. *Jama* **293**, 1861–1867 (2005).

378. Di Angelantonio, E. *et al.* Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet* **388**, 776–786 (2016).

379. Smith, G. D. *et al.* The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study. *Bmj* **339**, b5043 (2009).

380. Kritchevsky, S. B. *et al.* Intentional weight loss and all-cause mortality: a meta-analysis of randomized clinical trials. *PLoS One* **10**, (2015).

381. Bhaskaran, K., dos-Santos-Silva, I., Leon, D. A., Douglas, I. J. & Smeeth, L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3· 6 million adults in the UK. *Lancet Diabetes Endocrinol.* **6**, 944–953 (2018).

382. Ball, J. *et al.* Sex Differences in the Impact of Body Mass Index on the Risk of Future Atrial Fibrillation: Insights From the Longitudinal Population-Based Tromsø Study. *J. Am. Heart Assoc.* **7**, e008414 (2018).

383. Karastergiou, K., Smith, S. R., Greenberg, A. S. & Fried, S. K. Sex differences in human adipose tissues–the biology of pear shape. *Biol. Sex Differ.* **3**, 13 (2012).

384. Okada, R. *et al.* The number of metabolic syndrome components is a good risk indicator for both early-and late-stage kidney damage. *Nutr. Metab. Cardiovasc. Dis.* **24**, 277–285 (2014).

385. Perkins, R. M. *et al.* GFR decline and mortality risk among patients with chronic kidney disease. *Clin. J. Am. Soc. Nephrol.* **6**, 1879–1886 (2011).

386. Mottl, A. K. *et al.* Linkage analysis of glomerular filtration rate in American Indians. *Kidney Int.* **74**, 1185–1191 (2008).

387. Udler, M. S. *et al.* Effect of Genetic African Ancestry on eGFR and Kidney Disease. *J. Am. Soc. Nephrol.* **26**, 1682–1692 (2015).

388. McGrath, C., Palmarella, G., Solomon, S. & Dupuis, R. Precision prevention and public health. (2017).

389. BCRF. BCRF's Precision Prevention Initiative Announces $5 Million in Grants. *BCRF News* (2019).

390. Wilson, P. W. F. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–1847 (1998).

391. Kamath, P. S. *et al.* A model to predict survival in patients with end-stage liver disease. *Hepatology* **33**, 464–470 (2001).

392. Ridker, P. M., Buring, J. E., Rifai, N. & Cook, N. R. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *Jama* **297**, 611–619 (2007).

393. Nunez, T. C. *et al.* Early prediction of massive transfusion in trauma: simple as ABC (assessment of blood consumption)? *J. Trauma Acute Care Surg.* **66**, 346–352 (2009).

394. Fanaroff, A. C. *et al.* Risk Score to Predict Need for Intensive Care in Initially Hemodynamically Stable Adults With Non–ST-Segment–Elevation Myocardial Infarction. *J. Am. Heart Assoc.* **7**, e008894 (2018).

395. Scherzer, R. *et al.* A chronic kidney disease risk score to determine tenofovir safety in a prospective cohort of HIV-positive male veterans. *AIDS* **28**, 1289 (2014).

396. Ranucci, M., Pistuddi, V., Scolletta, S., De Vincentiis, C. & Menicanti, L. The ACEF II Risk Score for cardiac surgery: updated but still parsimonious. *Eur. Heart J.* **39**, 2183–2189 (2018).

397. D'agostino, R. B. *et al.* General cardiovascular risk profile for use in primary care. *Circulation* **117**, 743–753 (2008).

398. Mehran, R. *et al.* A simple risk score for prediction of contrast-induced nephropathy after percutaneous coronary intervention: development and initial validation. *J. Am. Coll. Cardiol.* **44**, 1393–1399 (2004).

399. Johnston, S. C. *et al.* Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet* **369**, 283–292 (2007).

400. Dix, D. J. *et al.* The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* **95**, 5–12 (2007).

401. Davis, A. P. *et al.* The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.* **47**, D948–D954 (2019).

402. Martin, J. A., Hamilton, B. E. & Osterman, M. J. K. Births in the United States, 2018. NCHS Data Brief, no 346. *Hyattsville, MD Natl. Cent. Heal. Stat.* (2019).

403. March of Dimes. March of dimes 2019 report card shines spotlight on the u.S. Maternal and infant health crisis. https://www.marchofdimes.org/news/march-of-dimes-2019-report-

card-shines-spotlight-on-the-u-s-maternal-and-infant-health-crisis.aspx (2019).

404. US Consumer Product Safety Commission. Phthalates. https://www.cpsc.gov/Business--Manufacturing/Business-Education/Business-Guidance/Phthalates-Information (2015).

405. Schettler, T. E. D. Human exposure to phthalates via consumer products. *Int. J. Androl.* **29**, 134–139 (2006).

406. Kefeni, K. K., Okonkwo, J. O., Olukunle, O. I. & Botha, B. M. Brominated flame retardants: sources, distribution, exposure pathways, and toxicity. *Environ. Rev.* **19**, 238–253 (2011).

407. Suzuki, G., Kida, A., Sakai, S. & Takigami, H. Existence state of bromine as an indicator of the source of brominated flame retardants in indoor dust. *Environ. Sci. Technol.* **43**, 1437–1442 (2009).

408. Lepow, M. L. *et al.* Investigations into sources of lead in the environment of urban children. *Environ. Res.* **10**, 415–426 (1975).

409. Hee, S. S. Q. *et al.* Evolution of efficient methods to sample lead sources, such as house dust and hand dust, in the homes of children. *Environ. Res.* **38**, 77–95 (1985).

410. Jaga, K. & Dharmani, C. Sources of exposure to and public health implications of organophosphate pesticides. *Rev. Panam. salud pública* **14**, 171–185 (2003).

411. Colt, J. S., Cyr, M. J., Zahm, S. H., Tobias, G. S. & Hartge, P. Inferring past pesticide exposures: a matrix of individual active ingredients in home and garden pesticides used in past decades. *Environ. Health Perspect.* **115**, 248–254 (2007).

412. Kumar, V. B. & Mohanta, D. Formation of nanoscale tungsten oxide structures and colouration characteristics. *Bull. Mater. Sci.* **34**, 435–442 (2011).

413. Buschmann, J. B., Davies, K. R. & Oetken, J. E. Tungsten halogen lamp. (2005).

414. Latch, D. E. *et al.* Aqueous photochemistry of triclosan: Formation of 2, 4-dichlorophenol, 2, 8-dichlorodibenzo-p-dioxin, and oligomerization products. *Environ. Toxicol. Chem. An Int. J.* **24**, 517–525 (2005).

415. Singer, H., Müller, S., Tixier, C. & Pillonel, L. Triclosan: occurrence and fate of a widely used biocide in the aquatic environment: field measurements in wastewater treatment plants, surface waters, and lake sediments. *Environ. Sci. Technol.* **36**, 4998–5004 (2002).

416. Saijo, Y. *et al.* Symptoms in relation to chemicals and dampness in newly built dwellings. *Int. Arch. Occup. Environ. Health* **77**, 461–470 (2004).

417. Dodson, R. E. *et al.* Measured and modeled personal exposures to and risks from volatile organic compounds. *Environ. Sci. Technol.* **41**, 8498–8505 (2007).

418. Lear, L. *Rachel Carson: witness for nature*. (Macmillan, 1998).

419. Dunlap, T. *DDT, silent spring, and the rise of environmentalism: Classic texts*. (University of Washington Press, 2017).

420. Vijgen, J. *et al.* Hexachlorocyclohexane (HCH) as new Stockholm Convention POPs—a global perspective on the management of Lindane and its waste isomers. *Environ. Sci. Pollut. Res.* **18**, 152–162 (2011).

421. WANG, Y. & CAI, Y. J. G. Research processes of persistent organic pollutants (POPs) newly listed and candidate POPs in Stockholm Convention. *Sci. Sin. Chim.* **40**, 99–123 (2010).

422. Cabaleiro, N., De La Calle, I., Bendicho, C. & Lavilla, I. An overview of sample preparation for the determination of parabens in cosmetics. *TrAC Trends Anal. Chem.* **57**, 34–46 (2014).

423. El Hussein, S., Muret, P., Berard, M., Makki, S. & Humbert, P. Assessment of principal parabens used in cosmetics after their passage through human epidermis–dermis layers (ex-vivo study). *Exp. Dermatol.* **16**, 830–836 (2007).

424. Buckley, J. P. *et al.* Consumer product exposures associated with urinary phthalate levels in pregnant women. *J. Expo. Sci. Environ. Epidemiol.* **22**, 468–475 (2012).

425. Benowitz, N. L. Biomarkers of environmental tobacco smoke exposure. *Environ. Health Perspect.* **107**, 349–355 (1999).

426. Scialla, M. It could take centuries for EPA to test all the unregulated chemicals under a new landmark bill. *PBS News Hour June* **22**, (2016).

427. Miller, G. W. *The Exposome: A New Paradigm for the Environment and Health*. (Academic Press, 2020).