Assessing the Impact of Metal and PFAS Exposure on Chronic Kidney Disease
An NHANES Data Analysis

by

Nuo Wei

BS Applied Mathematics, University at Buffalo, 2022

Submitted to the Graduate Faculty of the

Department of Biostatistics

School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Nuo Wei

It was defended on

April 26, 2024

and approved by

Thesis Advisor: Jeanine M Buchanich, PhD, Associate Professor,
Biostatistics School of Public Health, University of Pittsburgh

Jiebiao Wang, PhD, Assistant Professor
Biostatistics School of Public Health, University of Pittsburgh

Alison P Sanders, PhD, Assistant Professor
Environmental Occupational health School of Public Health, University of Pittsburgh

Assessing the Impact of Metal and PFAS Exposure on Chronic Kidney Disease: An NHANES Data Analysis

Nuo Wei, MS

University of Pittsburgh, 2024

Introduction: Chronic exposure to nephrotoxic substances is a public health concern, as these agents can significantly impair renal function, leading to chronic kidney disease (CKD), a condition characterized by a gradual loss of kidney function over time. The goal of this study was to explore whether exposure to certain metals (lead, cadmium, arsenic) and per- and polyfluoroalkyl substances (PFAS) was associated with CKD.

Methods: To analyze the association between single and mixed metal and/or PFAS exposures with CKD (n=983) in a nationally representative U.S. population, we utilized data from the National Health and Nutrition Examination Survey (NHANES) collected between 2011 and 2016 in which CKD was defined by estimated glomerular rate (eGFR) <60 mL/min/1.73m2. Single contaminant survey-weighted models were applied to enhance generalizability to the broader U.S. population. To assess mixture effects, we applied Weighted Quantile Sum (WQS) to assess the contributions of metals and PFAS jointly.

Result: In individual metal models, urinary lead and cadmium were inversely associated with the presence of CKD. In contrast, arsenic, PFOA, and PFOS did not show a significant association with CKD odds. Lead and cadmium were linked to a reduced likelihood of CKD. Age and hypertension were significant covariates, with increased age and the presence of hypertension correlating with higher odds of CKD. The WQS model found that a combined mixture of

contaminants was associated with increased eGFR. WQS analysis further identified cadmium and lead as top contributors to the mixture's association with higher continuous eGFR.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1 Background

Chronic Kidney Disease (CKD) represents a significant public health concern, affecting 10% of the of the global population. (Kovesdy, 2022) CKD is more common among older adults and people with high blood pressure and represents a substantial burden on low- and middle-income countries. CKD is of the leading causes of death worldwide and is one of the few non-communicable diseases for which related deaths have increased over past 20 years. (Kovesdy, 2022) The pathophysiology of CKD involves the gradual deterioration of kidney function, leading to the inability of the kidneys to effectively filter waste and excess fluids from the body. Current international guidelines define CKD as a serious health condition that mostly progresses without obvious symptoms in the early stages. (Mayo Foundation for Medical Education and Research, 2023) It is characterized by a decrease in kidney function, indicating by estimated glomerular filtration rate (eGFR) lower than 60 mL/min/1.73 m^2 or by markers of kidney damage persisting for at least three months. (Altamura et al., 2023) The disease's global surge is largely due to increased prevalence of traditional risk factors associated with development of work, such as diabetes, hypertension and obesity (Jin et al., 2018).

A prior study by Jin et al concluded that decreased kidney function was associated with reduced levels of metals in urine. (Jin et al., 2018) This association is not limited to kidney function levels below the normal range but is observed across the entire range of eGFR. This suggests that as kidney function declines, urinary excretion of metals also decreases. Furthermore, abnormal

levels of metals in urine may vary depending on an individual's eGFR value and age, meaning these factors need to be considered when assessing the health effects of urinary metal levels.

Many studies have examined the association of metal and PFAS and CKD or eGFR. Urinary levels of lead were nonlinearly and positively related to eGFR based on examining continuous relationship between urinary excretion rates of metals and eGFR in restricted cubic spline regression model (Jin et al., 2018). When using urinary creatinine to adjust urine concentration, urinary cadmium was positively associated with serum creatinine eGFR, and adjusted cadmiums had significant or borderline significant positive associations with eGFR (Weaver et al., 2014). In a dose-response relationship, total arsenic levels were significantly associated with CKD, particularly among participants with total arsenic levels greater than 20.74 and creatinine 11.78 microg/g or less (Hsueh et al., 2009). A strong relationship between each serum PFAS and eGFR in CKD in populations with anemia, but in the anemic population in the absence of CKD, there was a relatively weak relationship between PFAS and eGFR. This relationship held regardless of the presence or absence of diabetes, although it appeared to be stronger in patients with diabetes (Conway et al., 2018).

2.0 Methods

2.1 Data and Weighted Survey Design

2.1.1 Data Source

This study utilized data derived from the 2011-2016 National Health and Nutrition Examination Survey (NHANES). This survey encompassed a series of interviews that gather comprehensive information about demographic and socioeconomic factors, dietary habits, and health-related issues. Furthermore, it included a variety of medical, dental, and physiological evaluations, along with laboratory tests, all of which are carried out by qualified professionals. The primary objective of NHANES was to ascertain and comprehend the incidence of major diseases and the risk factors associated with them within the non-institutionalized American population, which is vital in shaping public health policies and strategies ("Centers for Disease Control and Prevention," 2024).

2.1.2 Weighted Survey Design

The NHANES consisted of a series of interviews, allowing us to apply weights to represent the results across the entire population. In NHANES, there were three sampling schemas: sampling weight, primary sampling unit (PSU), and strata. In order to represent the United States population, each sample participant was given a sample weight.

2.1.2.1 Sampling Weights

In the NHANES, sample weights for each participant were developed through a three-step process to ensure the survey results were representative of the U.S. population. Initially, a base weight was calculated to account for the unequal probabilities of selection, addressing the issue of certain demographic groups being over-sampled. Subsequently, adjustments were made for non-response, ensuring that the sample accurately reflected the entire population, despite variations in response rates across different groups. Finally, post-stratification adjustments were applied to align the survey estimates with demographic information of the U.S. population as provided by the Census Bureau, thereby correcting any discrepancies and ensuring the survey's representativeness. (“Centers for Disease Control and Prevention,” 2024).

$$Base\ weight = \frac{1}{final\ probability}$$

Where

$$Final\ probability = (Pr(PSU\ is\ selected) \times Pr(segment\ of\ the\ PSU\ is\ selected) \times Pr(household\ is\ selected) \times Pr(individual\ is\ selected))$$

2.1.2.2 Primary Sampling Unit (PSU)

PSUs, typically single counties, or groups of contiguous counties, were chosen based on Probability Proportional to Size (PPS), favoring areas with larger populations and higher concentrations of specific demographic subgroups targeted for oversampling. Certain PSUs with a large Measure of Size (MOS) were automatically included in the sample. The PSUs were then

divided into smaller segments, like city blocks, where again PPS was utilized for selection. Within each segment, dwelling units were listed, and households were randomly chosen, with a greater selection probability assigned to areas rich in oversampled demographics. Finally, from these households, individuals were randomly selected based on specific age-sex-race/ethnicity criteria, aiming for an average of two sampled individuals per household, to participate in NHANES, thus creating a stratified, multi-stage approach to accurately reflect the diverse U.S. population. (“Centers for Disease Control and Prevention,” 2024).

2.1.2.3 Strata

Stratification involved dividing the survey population into smaller, more homogenous groups in order to improve the precision of estimated weight based on specific characteristics such as age, sex, race, ethnicity, or other relevant demographic factors. Create and provide Masked Variance Units (MVUs) in the demographic data files for each survey cycle. These MVUs generated variance estimates that were very close to those estimated using the actual design variables and were applied to all analyses of the publicly released data. The variable name for the pseudo-stratum of the masked variance unit was "sdmvstra," and the variable name for the pseudo-PSU (Primary Sampling Unit) of the masked variance unit was "sdmvpsu." (“Centers for Disease Control and Prevention,” 2024).

2.1.3 Limit of Detection

The most common strategy for dealing with values below the limit of detection is imputation. This involves replacing censored values with zero, some fraction of the detection limit, and the detection limit itself. In NHANES, if the variable name ends in “LC” it means that this

variable was below the limit of detection, in this case, values "0" means the result was at or above the limit, and "1" means the result was below the limit. Equating the value of concentrations below the detection limit (LOD) to the square root of the detection limit divided by two has little effect on the geometric mean estimate. It was not necessary for calculating geometric means if the proportion of results below the detection limit was greater than 40%. Due to improvements in analytical methods, LOD values sometimes changed over time even for the same chemical ("Centers for Disease Control and Prevention," 2022). Table 1 below shows that values of LOD for metals did change a lot between 2011 and 2013, especially for urine arsenic and cadmium.

Table 1 LODs for Five Contaminants of Interest

	2011-2012	2013-2014	2015-2016
Lead, urine (ug/L)	0.08	0.03	0.03
Cadmium, urine (ug/L)	0.056	0.036	0.036
Arsenic, urine (ug/L)	1.25	0.26	0.26
n-perfluorooctanoic acid (n-PFOA) (ng/mL)	0.1	0.1	0.1
n-perfluorooctane sulfonic acid (n-PFOS) (ng/mL)	0.2	0.1	0.1

2.1.4 Outcome Variables

eGFR can be calculated using a formula that accounts for various factors such as creatinine levels, a waste product that results from normal muscle wear and tear, and/or cystatin C, a protein that slows down the breakdown of other protein cells. Additionally, the formula includes factors for age and sex. In this research, the eGFR was calculated using the CKD-EPI equation without race variables (Delgado et al., 2022). A lower eGFR may indicate kidneys are not filtering blood

efficiently and is an important indicator for diagnosing CKD which may be diagnosed clinically when eGFR measures below 60 mL/min/1.73m2 (GFR categories G3a-G5) or markers of kidney damage are present at least 3 months apart (Inker et al., 2014).

$$eGFR = 142 * min(\frac{Scr}{\kappa}, 1)^{\alpha} * max(\frac{Scr}{\kappa}, 1)^{-1.2} * 0.994^{age} * 1.012(if\ female)$$

The formula includes Scr, serum creatinine, and k, which was 0.7 for females and 0.9 for males, as well as α, which was -0.241 for females and -0.302 for males. The formula also utilized min to indicate the minimum of Scr/k or 1 and max to indicate the maximum of Scr/k or 1 (Delgado et al., 2022).

The study analyzed the eGFR data by calculating serum creatinine levels obtained from the NHANES standard Biochemistry Profile files three survey cycles: 2011-2012, 2013-2014, and 2015-2016. NHANES employs rigorous quality control (QC) procedures for serum creatinine analysis, which include the use of blind QC specimens, assaying controls at different times during the day, running BioRad Liquid Unassayed Multiqual Controls at both the beginning and end of sample analysis, and ensuring that results are within ±2 standard deviations, with further guidance provided by a Quality Control Flow Chart ("Centers for Disease Control and Prevention," 2017).

2.1.5 Explanatory Variables

The exposure characteristics considered included concentrations of micrograms per liter (ug/L) for urine lead, cadmium, and arsenic, and nanograms per milliliter (ng/mL) for serum PFOA and PFOS. This method measured multiple metals in urine samples through mass spectrometry, where liquid specimens are prepared by simple dilution, then introduced into an inductively coupled plasma (ICP) ionization source, reduced to small droplets via a nebulizer, and

subsequently passed through a series of regions including a focusing area, a dynamic reaction cell (DRC), a quadrupole mass filter, and finally to the detector, allowing for rapid selective counting of individual isotopes ("Centers for Disease Control and Prevention," 2018). The method measured multiple PFAS in serum sample through High-performance liquid chromatography (HPLC) separates analytes from other serum components, followed by detection and quantification through negative-ion TurboIonSpray ionization, a variant of electrospray ionization, and tandem mass spectrometry, enabling rapid detection of PFAS in human serum with detection limits in the low parts per billion (ppb or ng/mL) range ("Centers for Disease Control and Prevention," 2018). Table 2 below listed different variables, their descriptions, the NHANES file in which they could be found, and the years the data were collected. Each variable was a specific substance that had been measured in the urine or serum of NHANES participants, such as lead, cadmium, arsenic, PFOA, and PFOS, and the table provided the units of measurement for each substance.

Table 2 Table of Explanatory Variables

Variable	Description	NHANES file	Year
Lead (numeric)	Lead, urine (ug/L)	Metals-Urine (UM_H)	2011-2016 Laboratory
Cadmium (numeric)	Cadmium, urine (ug/L)	Metals-Urine (UM_H)	2011-2016 Laboratory
Arsenic (numeric)	Urinary arsenic, total (ug/L)	Arsenics - Total & Speciated - Urine (UAS_G)	2011-2016 Laboratory
PFOA (numeric)	n-perfluorooctanoic acid (n-PFOA) (ng/mL)	Perfluoroalkyl and Polyfluoroalkyl (PFAS_I)	2011-2016 Laboratory
PFOS (numeric)	n-perfluorooctane sulfonic acid (n-	Perfluoroalkyl and Polyfluoroalkyl	2011-2016 Laboratory

	PFOS) (ng/mL)	(PFAS_I)	

2.1.6 Covariates

The selection of potential covariates was informed by a study by Nan et al. (Nan et al., 2023), which highlighted the significance of understanding the relationship between metal exposure and kidney function while controlling for various variables. This led to the inclusion of a diverse array of covariates: age as a quantitative measure including those aged 18 and above, self-reported gender (categorized as male or female), and BMI, which was calculated using the formula: weight (lb) / [height (in)] ^2 * 703("Centers for Disease Control and Prevention," 2024). Self-reported smoking status was assessed through participants' responses to a questionnaire, categorizing them as everyday smokers, occasional smokers, or non-smokers. Self-reported diabetes status was determined by a doctor's diagnosis, which utilized a combination of a doctor's diagnosis, fasting plasma glucose levels ≥ 126 mg/dL, glycohemoglobin levels ≥ 6.5%, or the use of anti-diabetic medication. (American Diabetes, 2010) Self-reported hypertension status was based on a doctor's diagnosis which was identified by a systolic blood pressure (SBP) of 130 mmHg or higher, or a diastolic blood pressure (DBP) of 80 mmHg or above. (Whelton et al., 2018), and self-reported alcohol consumption was determined by asking participants whether they had consumed at least 12 alcoholic beverages of any type in the past year. Given that NHANES provided two types of urine measurements—one corrected for urine creatinine and the other not—urine creatinine was included as a covariate to account for this variable. This adjustment refined the analysis by controlling urine creatinine levels, which may affect the concentrations of other

substances measured in the urine. Below, Table 3 lists the different covariates, their descriptions, the NHANES file in which they were found, and the years the data were collected.

Table 3 Table of Covariates

Variable	Description	NHANES file	Year
Covariates			
Age(numeric)	Age in years of the participant at the time of screening. Individuals 80 and over are topcoded at 80 years of age.	Demographic Variables & Sample Weights (DEMO_G)	2011-2016 Demographics
Gender (Categorical)	Gender of the participant.	Demographic Variables & Sample Weights (DEMO_G)	2011-2016 Demographics
BMI (numeric)	Body Measures Component Status Code	Body Measures (BMX_J)	2011-2016 Examination
nicotine exposure (Smoking) (Categorical)	Do you now smoke cigarettes?	Smoking - Cigarette Use (SMQ_I)	2011-2016 Questionnaire
Diabetes (Categorical)	have you/has ever been told by a doctor or health professional that has diabetes or sugar diabetes?	Diabetes (P_DIQ)	2011-2016 Questionnaire
Hypertension (Categorical)	ever been told by a doctor or other health professional that had hypertension, also called high blood pressure?	Blood Pressure & Cholesterol (BPQ)	2011-2016 Questionnaire

Alcohol use (Numeric)	In any one year, (have you/has SP) had at least 12 drinks of any type of alcoholic beverage? By a drink, I mean a 12 oz. beer, a 5 oz. glass of wine, or a one and a half ounces of liquor.	Alcohol Use (ALQ_G)	2011-2016 Questionnaire
Creatinine (Numeric)	Creatinine, urine (mg/dL)	Metals-Urine (UM_H)	2011-2016 Laboratory

2.2 Logistic Regression

2.2.1 Weighted Logistic Regression

We used a weighted survey design to investigate the relationship between the health condition and individuals metal or PFAS exposure, because the outcomes were two binary variables valued at 0 or 1. These models adjusted for the probability of exposure to different metals and their potential impact on kidney function and hypertension status, with covariates including age, gender, BMI, smoking status, diabetes, hypertension, alcohol use, and urine creatinine. To assess model fit, we used R-squared and Adjusted R-squared values to determine how well our linear regression models explained the variance in the outcome variables. A higher R-squared value indicates that the model explains a greater portion of the variability.

Suppose that the United Stated population$U = \{1,2, \dots N\}$ was divided into $h = 1,2, \dots N$ strata which were county segments, each stratum was further divided into $j = 1,2, \dots n_h$ primary sample units (PSU) which were each Household NHANES considered in survey. Each of PSU

was constituted by $i = 1,2, \ldots n_{hj}$ secondary sample units (SSU) which was individual who participated in study. It was also assumed that observations consisted of n'_{hj} SSU chosen from n'_h PSU in the stratum h. $n = \sum_{h=1}^{H} \sum_{j=1}^{n'_k} \sum_{i=1}^{n'_{hj}} n_{hji}$ was the total number of the observed data, and each sample had an corresponding sampling weight determined by inverse of $w_{hjik} = \frac{1}{\pi_{hjik}}$ for the $hjik$-th unit.

$$Logit\{\mathrm{p}(\mathrm{Y}_{hjik} = 1|\mathrm{x}_{hjik})\} = \ln\{\frac{P(Y_{hjik} = 1|x_{hjik})}{(1 - P(Y_{hjik} = 1|x_{hjik})}\} = x'_{hjik}\beta$$

In this equation,Y_{hjik} was the probability of the outcome being 1 given the covariates x_{hjik}, β was the regression coefficients, and x'_{hjik} was the transpose of the covariate's matrix for the $hjik$-th unit (Cassy et al., 2016).

2.3 Weighted Quantile Sum Regression

In our study, Weighted Quantile Sum (WQS) regression was utilized to construct a weighted index in a supervised manner to assess the overall effect of environmental exposure, as well as the contribution of each component of the mixture to this overall effect. The WQS model estimated a body burden index to identify 'bad actors' in a set of highly correlated environmental chemicals (Carrico et al., 2015). This index was then incorporated into a regression model along with appropriate covariates to evaluate its association with a dependent variable or outcome.

$$g(\mu) = \beta_0 + \beta_1(\sum_{i=1}^{c} w_i q_i) + Z'\phi$$

$w_i$was the unknown weight associated with each component of the mixture, which were estimated by bootstrap samples, $q_i$was the value for letting the values of the c components be scored into quantiles such as (1st, 2nd , 3rd quartile) for $i = 1$ to c. $\sum_{i=1}^{c} w_i q_i$ represents the index that weights and sums the components included in the metal mixture, $\beta_0$is the intercept, β_1 is the coefficient associated to the WQS index, $Z'\phi$ are the vector of covariates and parameters. A training dataset which was used in the ensemble step to make sure weighted indices and a validation dataset were required in WQS. $\sum_{i=1}^{c} w_i = 1 \; and \; 0 \leq w_i \leq 1$ were constraints for estimating weights.

When the weights of each ensemble step sample were estimated, the WQS index was $WQS = \sum_{j=1}^{c} \overline{w_j} q_j$, where $\overline{w_j}$ was the weights found to be associated with positive or negative correlations in the aggregation step $\overline{w_j} = \frac{1}{B} \sum_{b=1}^{B} w_{j(b)} f(\widehat{\beta_{1(b)}})$ where $f(\widehat{\beta_{1(b)}})$ was a the predetermined "signal function" of the estimated slope parameter associated with the WQS in the bootstrap sample (constrained sum of 1). Utilizing of validation data and modeling to determine the significance of WQS index (Renzetti et al., 2023).

$$g(\mu) = \beta_0 + \beta_1 WQS + Z'\phi$$

In this study, the WQS permitted the evaluation of the mixture effect of multiple toxicants simultaneously—such as lead, cadmium, arsenic, PFOA, and PFOS—represented as a weighted index, which collectively contributed to the risk of CKD. The outcome variable for our analysis was the numerical values of eGFR. Covariates included a range of personal health factors, such as age, gender, BMI, smoking status, hypertension, diabetes status, and urine creatinine. Models were run both including and excluding people with hypertension.

3.0 Result

3.1 Study Population

This study included participants aged 18 and older who had available exposure measurements for contaminants of interest: urine lead, cadmium, or arsenic, as well as serum PFOA or PFOS. Individuals with missing data for the exposure of interest, serum creatinine, and covariates were excluded from the dataset (Figure 1). We applied R code to merge multiple datasets from NHANES relevant to our analysis variables, using the `left_join` function from the `dplyr` package, and matched them with `SEQN` identifiers. As described in Figure1, among 9971 participants from NHANES 2015-2016, we included 5992 participants that were over the age of 18. Additionally, we excluded 607 participants with missing values on serum creatinine, 3157 participants with missing values on metal exposure, and 1999 participants with missing values on PFAS from the study. After excluding 24 missing values on covariates, 199 participants are considered in 2015-2016. Among the 10175 participants from NHANES 2013-2014, we included 6113 individuals aged 18 and above. We then excluded 489 participants due to missing serum creatinine values, 3229 participants due to missing data on metal exposure, and 2146 participants due to missing data on PFAS. Additionally, 27 participants were excluded because of missing values on covariates. Consequently, 222 participants were considered for the study in 2013-2014. From a total of 9756 participants in 2011-2012, we selected 5864 participants who were over the age of 18. However, exclusions were made as follows: 707 for missing serum creatinine levels, 3047 for incomplete urinary metal exposure data, and 577 for absent serum PFAS information. An additional 977 were excluded due to missing covariate data, leaving 556 participants eligible for

the study period of 2011-2012. Three years of datasets were combined using the `full_join` function, based on common variable identifiers. As a result, the study cohort amounted to 983 participants in the primary study.

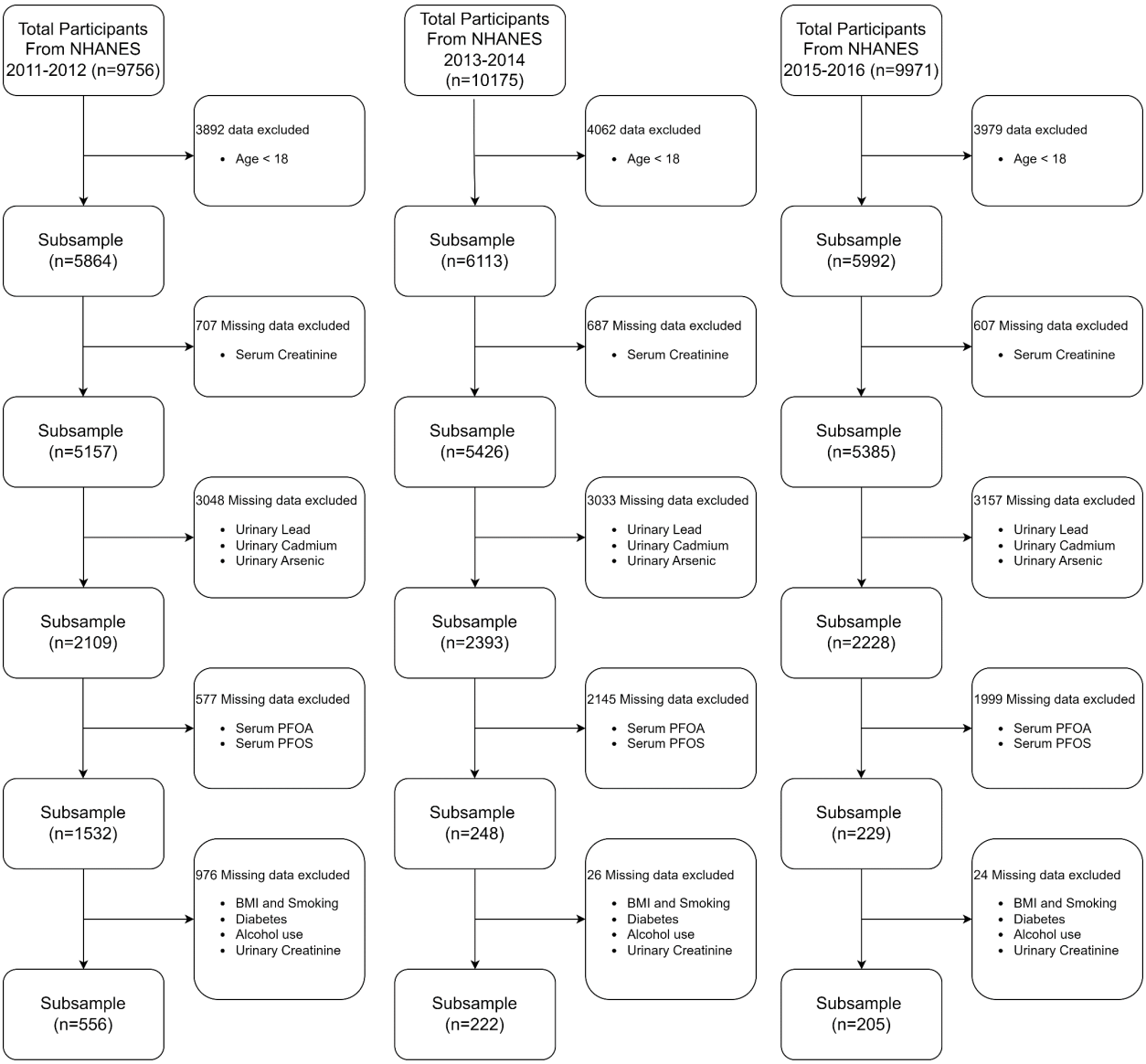

Figure 1 Flow Diagram of Outcome

3.2 Summary Statistics

Tables 4 and 5 provide summaries of continuous and categorical variables, respectively, including descriptions of data before and after weighting. Table 4 presents the mean and standard deviations for continuous variables, while Table 5 details the proportions of categorical variables.

Analysis with the primary outcome of CKD included 983 participants, the weighted mean and standard deviation for eGFR were 97.91 and 19.57 mL/min/1.73m2, respectively. Furthermore, 68 participants were identified as having CKD, representing 6.9% of the primary outcome.

Table 4 Summary Statistics for Continuous Variables in Outcomes

Variable	Unweighted Mean ±Standard Deviation	Weighted Mean ±Standard Deviation
Outcome		
eGFR, (mL/min/1.73m2)	95.49 ± 21.65	97.91 ±19.57
Exposure		
Lead, urine (ug/L)	0.72± 1.06	0.70 ±1.13
Cadmium, urine (ug/L)	0.54 ± 0.67	0.51 ± 0.73
Arsenic, urine (ng/mL)	15.88 ± 33.42	12.71 ± 24.4
PFOA, serum (ng/mL)	2.42 ± 3.10	2.66 ± 4.55
PFOS, serum (ng/mL)	8.47 ± 12.3	7.39 ± 8.57
Covariates		
Age, year	49.49 ± 16.67	48.26 ± 15.58
BMI, kg/m^2	28.85 ± 6.86	28.75 ± 6.71
Creatinine, urine (mg/dL)	125.97 ± 84.18	117.97 ± 81.39

Table 5 Summary Statistics for Cateogrical Variables in Outcomes

	Details	Unweighted Frequency	Unweighted Percent	Weighted Percent
Outcome				
CKD	No CKD CKD	915 68	93.1 6.9	95.8 4.2
Covariates				
Gender	Male Female	571 412	58.1 41.9	53.9 46.1
nicotine exposure (Smoking)	Not at all Some day Every day	311 48 624	31.6 4.9 63.5	32.6 5.5 61.8
Diabetes	No Yes	846 137	86.1 13.9	89.1 10.9
Hypertension	No Yes	588 395	59.8 40.1	63.2 36.8
Alcohol use	No Yes	138 845	14.0 85.98	10.4 89.6

3.3 Weighted Binary Logistic Models

3.3.1 Lead

As shown in Table 6, we conducted binary logistic regression to explore the associations between CKD and urinary lead levels, while adjusting for personal health factors as covariates. The model with age included lead exposure, age, gender, BMI, urinary creatinine, diabetes status, hypertension status, and alcohol consumption as predictors. The odds ratio of 0.27 for lead meant that each unit increase in urine lead was associated with a 73% decrease in the odds of having

CKD, and this association was statistically significant with a p-value of 0.02. The statistically significant predictors in this model were age and hypertension. Age had an odds ratio of 1.13, indicating that for each additional year of age, the odds of having CKD increased by 13%. Hypertension also appeared as a significant covariate with an odds ratio of 2.56, suggesting that the odds of having CKD were 2.56 times higher for individuals with hypertension compared to those without hypertension.

In the model without age, smoking status, diabetes, and hypertension were significant predictors. Diabetes had an odds ratio of 2.8, indicating that the odds of having CKD were 2.8 times higher for individuals with diabetes compared to those without diabetes. Hypertension also appeared as a significant covariate with an odds ratio of 6.36, suggesting that the odds of having CKD were 6.36 times higher for individuals with hypertension compared to those without hypertension.

Table 6 Weighted Binary Logistic Regression to Predict CKD for Lead Exposure

Variable	Details	With Age		Without Age	
		Odd ratio [95% CI]	P-value	Odd ratio [95% CI]	P-value
Exposure					
Lead(numeric)	Lead, urine (ug/L)	0.27 [0.09, 0.79]	0.02 *	0.68 [0.36, 1.30]	0.26
Covariates					
Age(numeric)	Age	1.13 [1.08, 1.18]	1.1e^-0.5 *	----	----
Gender (Categorical) Reference: men	Women	1.63 [0.69, 3.85]	0.27	1.47 [0.69, 3.22]	0.34
BMI (numeric)	Body Measures Component Status (kg/m**2)	0.99 [0.94, 1.05]	0.78	0.97 [0.92, 1.02]	0.21
Creatinine (numeric)	Creatinine, urine (mg/dL)	1.01 [1.01, 1.01]	1.37e^-0.5 *	1.00 [1.00, 1.01]	0.03 *
Smoking status Reference: Not at all	Some day	2.63 [0.71, 9.80]	0.16	0.64 [0.16, 2.49]	0.52
	Every day	1.21 [0.49, 2.98]	0.69	0.41 [0.22, 0.74]	0.01 *
Diabetes (Categorical) Reference: No	Yes	0.40 [0.16, 1.00]	0.06	2.80 [1.18, 6.68]	0.03 *

Hypertension (Categorical) Reference: No	Yes	2.56 [1.27, 5.16]	0.01 *	6.36 [3.43, 11.79]	9.58e-07 *
Alcohol use (Categorical) Reference: No	Yes	0.83 [0.31, 2.24]	0.72	0.59 [0.25, 1.44]	0.26

* Indicates P-value <0.05.

3.3.2 Cadmium

In Table 7, we conducted a binary logistic regression analysis to explore the associations between CKD and urinary cadmium levels, while adjusting for personal health factors as covariates. The model included cadmium exposure, age, gender, BMI, urinary creatinine, diabetes status, hypertension status, and alcohol consumption as predictors. An odds ratio of 0.40 for cadmium meant that each unit increase in urine cadmium was associated with a 60% decrease in the odds of having CKD, and this association was statistically significant with a p-value of 0.003. It's important to note that the statistically significant predictors in this model were age and hypertension. Age had an odds ratio of 1.13, indicating that for each additional year of age, the odds of having CKD increased by 13%. Hypertension also appeared as a significant covariate with an odds ratio of 2.87, suggesting that the odds of having CKD were 2.87 times higher for individuals with hypertension compared to those without hypertension.

For the model without age, smoking status, diabetes, and hypertension were significant predictors. Diabetes had an odds ratio of 2.82, indicating that the odds of having CKD were 2.82 times higher for individuals with diabetes compared to those without diabetes. Hypertension also appeared as a significant covariate with an odds ratio of 6.46, suggesting that the odds of having CKD were 6.46 times higher for individuals with hypertension compared to those without hypertension.

Table 7 Weighted Binary Logistic Regression to Predict CKD for Cadmium Exposure

Variable	Details	With Age		Without Age	
		Odd ratio [95% CI]	P-value	Odd ratio [95% CI]	P-value
Exposure					
Cadmium (numeric)	Cadmium, urine (ug/L)	0.40 [0.23, 0.70]	0.0027 *	0.87[0.69,1.10]	0.26
Covariates					
Age(numeric)	Age	1.13 [1.07, 1.18]	2.33e^-0.5 *	---	---
Gender (Categorical) Reference: men	Women	2.31 [0.92, 5.80]	0.08	1.56 [0.70, 3.45]	0.28
BMI (numeric)	Body Measures Component Status (kg/m**2)	0.99 [0.93, 1.05]	0.77	0.97 [0.92, 1.03]	0.29
Creatinine (numeric)	Creatinine, urine (mg/dL)	1.01 [1.00, 1.01]	0.0002 *	1.00 [1.00, 1.01]	0.07
Smoking status Reference: Not at all	Some day	3.03 [0.92, 9.99]	0.08	0.66 [0.17, 2.54]	0.54
	Every day	1.41 [0.57, 3.49]	0.46	0.41 [0.23, 0.75]	0.01 *
Diabetes (Categorical) Reference: No	Yes	2.55 [0.99, 6.57]	0.06	2.82 [1.18, 6.76]	0.03 *

Hypertension (Categorical) Reference: No	Yes	2.87 [1.47, 5.60]	0.004 *	6.46 [3.41, 12.25]	1.51e-06 *
Alcohol use (Categorical) Reference: No	Yes	0.93 [0.35, 2.48]	0.89	0.61 [0.25, 1.47]	0.28

* Indicates P-value <0.05.

3.3.3 Arsenic

In Table 8, we conducted a binary logistic regression to explore the associations between CKD and urinary arsenic levels, while adjusting for personal health factors as covariates. The model included arsenic exposure, age, gender, BMI, urinary creatinine, diabetes status, hypertension status, and alcohol consumption as predictors. However, this association was not statistically significant, with a p-value of 0.227. The statistically significant predictors in this model were age and hypertension. Age had an odds ratio of 1.12, indicating that for each additional year of age, the odds of having CKD increased by 12%. Hypertension also appeared as a significant covariate with an odds ratio of 2.62, suggesting that the odds of having CKD were 2.62 times higher for individuals with hypertension compared to those without hypertension.

In the model without age, smoking status, diabetes, and hypertension were significant predictors. Diabetes had an odds ratio of 2.77, indicating that the odds of having CKD were 2.77 times higher for individuals with diabetes compared to those without diabetes. Hypertension also appeared as a significant covariate with an odds ratio of 6.25, suggesting that the odds of having CKD were 6.25 times higher for individuals with hypertension compared to those without hypertension.

Table 8 Weighted Binary Logistic Regression to Predict CKD for Arsenic Exposure

Variable	Details	With Age		Without Age	
		Odd ratio [95% CI]	P-value	Odd ratio [95% CI]	P-value
Exposure					
Arsenic(numeric)	Urinary arsenic, total (ug/L)	0.99[0.97,1.01]	0.23	1.00 [0.99, 1.01]	0.58
Covariates					
Age(numeric)	Age	1.12 [1.06, 1.17]	8.35e^-0.5 *	---	---
Gender (Categorical) Reference: men	Women	1.62 [0.65, 4.05]	0.31	1.50 [0.67, 3.33]	0.33
BMI (numeric)	Body Measures Component Status (kg/m**2)	1.01 [0.96, 1.06]	0.65	0.97 [0.92, 1.03]	0.33
Creatinine (numeric)	Creatinine, urine (mg/dL)	1.01 [1.00, 1.01]	0.005 *	1.00 [1.00, 1.01]	0.14
Smoking status Reference: Not at all	Some day	2.52 [0.76, 8.40]	0.14	0.65 [0.17, 2.53]	0.54
	Every day	1.07 [0.44, 2.59]	0.89	0.40 [0.22, 0.72]	0.004 *
Diabetes (Categorical) Reference: No	Yes	2.45 [0.99, 6.05]	0.06	2.77 [1.16, 6.66]	0.03 *
Hypertension	Yes	2.62 [1.33, 5.17]	0.0085 *	6.25 [3.34, 11.69]	1.43e-06 *

(Categorical) Reference: No					
Alcohol use (Categorical) Reference: No	Yes	0.85 [0.32, 2.25]	0.75	0.60 [0.25, 1.46]	0.27

* Indicates P-value <0.05.

3.3.4 PFOA

In Table 9, we conducted a binary logistic regression analysis to explore the associations between CKD and serum PFOA levels, while adjusting for personal health factors as covariates. The model included PFOA exposure, age, gender, BMI, urinary creatinine, diabetes status, hypertension status, and alcohol consumption as predictors. PFOA was not statistically significant in predicting the presence of CKD, with a p-value of 0.47. The significant covariate in this model was age, with an odds ratio of 1.12 and a p-value of 0.0001, implying that for each additional year of age, the odds of having CKD increased by 12%. The only other significant predictor in this model was hypertension, with an odds ratio of 2.7, suggesting that the odds of having CKD were 2.7 times higher for individuals with hypertension compared to those without.

In the model without age, smoking status, diabetes, and hypertension were significant predictors. Diabetes had an odds ratio of 2.77, indicating that the odds of having CKD were 2.77 times higher for individuals with diabetes compared to those without diabetes. Hypertension appeared as a significant covariate with an odds ratio of 6.24, suggesting that the odds of having CKD were 6.24 times higher for individuals with hypertension compared to those without hypertension.

Table 9 Weighted Binary Logistic Regression to Predict CKD for PFOA Exposure

Variable	Details	With Age		Without Age	
		Odd ratio [95% CI]	P-value	Odd ratio [95% CI]	P-value
Exposure					
PFOA (numeric)	n-perfluorooctanoic acid (n-PFOA) (ng/mL)	0.93 [0.78, 1.11]	0.46	0.98 [0.87, 1.10]	0.69
Covariates					
Age(numeric)	Age	1.12 [1.06, 1.17]	0.0001 *	---	---
Gender (Categorical) Reference: men	Women	1.63 [0.65, 4.12]	0.31	1.48 [0.65, 3.37]	0.35
BMI (numeric)	Body Measures Component Status (kg/m**2)	1.01 [0.96, 1.07]	0.67	0.97 [0.92, 1.03]	0.34
Creatinine (numeric)	Creatinine, urine (mg/dL)	1.00 [1.00, 1.01]	0.015 *	1.00 [1.00, 1.01]	0.16
Smoking status Reference: Not at all	Some day	2.32 [0.63, 8.50]	0.21	0.65 [0.17, 2.50]	0.53
	Every day	1.02 [0.42, 2.52]	0.96	0.39 [0.22, 0.70]	0.003 *
Diabetes (Categorical) Reference: No	Yes	2.40 [1.01, 5.73]	0.06	2.77 [1.17, 6.58]	0.03 *

Hypertension (Categorical) Reference: No	Yes	2.70 [1.38, 5.25]	0.006 *	6.24 [3.36, 11.60]	1.2e-06 *
Alcohol use (Categorical) Reference: No	Yes	0.79 [0.31, 2.03]	0.63	0.59 [0.25, 1.42]	0.25

3.3.5 PFOS

In Table 10, we conducted a binary logistic regression analysis to explore the associations between CKD and serum PFOS levels, while adjusting for personal health factors as covariates. The model included PFOS exposure, age, gender, BMI, urinary creatinine, diabetes status, hypertension status, and alcohol consumption as predictors. However, PFOS was not statistically significant in predicting the presence of CKD, with a p-value of 0.71. The significant covariate in this model was age, with an odds ratio of 1.11 and a p-value of 0.0001, implying that for each additional year of age, the odds of having CKD increased by 11%. The other significant predictor in this model was hypertension, which had an odds ratio of 2.68, suggesting that the odds of having CKD were 2.68 times higher for individuals with hypertension compared to those without.

In the model without age, PFOS, smoking status, diabetes, and hypertension were significant predictors. An odds ratio of 1.02 for PFOS meant that each unit increase in PFOS was associated with a 2% increase in the odds of having CKD, and this association was statistically significant with a p-value of 0.04. Diabetes had an odds ratio of 2.91, indicating that the odds of having CKD were 2.91 times higher for individuals with diabetes compared to those without diabetes. Hypertension appeared as a significant covariate with an odds ratio of 6, suggesting that the odds of having CKD were 6 times higher for individuals with hypertension compared to those without. Other covariates like gender, BMI, smoking status, and alcohol use were included in the model, but none of them showed statistical significance in this context.

Table 10 Weighted Binary Logistic Regression to Predict CKD for PFOS Exposure

Variable	Details	With Age		Without Age	
		Odd ratio [95% CI]	P-value	Odd ratio [95% CI]	P-value
Exposure					
PFOS (numeric)	n-perfluorooctane sulfonic acid (n-PFOS) (ng/mL)	1.00 [0.98, 1.03]	0.71	1.02 [1.00, 1.04]	0.04 *
Covariates					
Age(numeric)	Age	1.11 [1.06, 1.17]	0.0001 *	---	---
Gender (Categorical) Reference: men	Women	1.70 [0.65, 4.44]	0.29	1.67 [0.71, 3.91]	0.25
BMI (numeric)	Body Measures Component Status (kg/m**2)	1.01 [0.96, 1.07]	0.62	0.98 [0.93, 1.03]	0.40
Creatinine (numeric)	Creatinine, urine (mg/dL)	1.00 [1.00, 1.01]	0.011 *	1.00 [1.00, 1.01]	0.15
Smoking status Reference: Not at all	Some day	2.53 [0.70, 9.14]	0.16	0.64 [0.16, 2.62]	0.54
	Every day	1.12 [0.47, 2.64]	0.80	0.45 [0.26, 0.78]	0.01 *
Diabetes (Categorical) Reference: No	Yes	2.51 [1.04, 6.04]	0.05	2.91 [1.20, 7.06]	0.02 *

Hypertension (Categorical) Reference: No	Yes	2.68 [1.37, 5.27]	0.0068 *	6.00 [3.25, 11.10]	1.53e-06 *
Alcohol use (Categorical) Reference: No	Yes	0.84 [0.33, 2.12]	0.71	0.63 [0.27, 1.45]	0.29

* Indicates P-value <0.05.

3.4 Weighted Quantile Sum Regression

The WQS index with hypertension had an estimate of 2.28, indicating a positive relationship between contaminants and eGFR with higher exposure to certain contaminants being associated with increased eGFR. As the weighted exposure to these contaminants increased, the outcome variable eGFR also increased. The p-value associated with the WQS index was less than 0.05, indicating that the association between the combined exposures and the outcome was statistically significant. It was also important to note that while the overall WQS index was significantly associated with the outcome, individual exposures might have significant effects that were not captured by this model. Other covariates in the model, such as age, urinary creatinine, and hypertension status, showed significant associations with the outcome.

In the WQS regression model without hypertension (n = 588), the index had an estimate of 1.72, indicating a positive relationship between contaminants and eGFR with higher exposure to certain contaminants being associated with increased eGFR. The covariates age and urinary creatinine were statistically significant. Age and eGFR are negatively correlated, meaning that as age increases, eGFR decreases. Tables 11 provide summaries of WQS and covariates, respectively, including descriptions of data with and without hypertension.

Table 11 Summary of Weighted Quantile Sum Regression

Variable	Details	With Hypertension		Without Hypertension	
		Estimate	P-value	Estimate	P-value
WQS	WQS index	2.28	1.20e-08 *	1.72	0.002 *
Covariates					
Age(numeric)	Age	-0.96	2e-16 *	-0.82	2e-16 *
Gender (Categorical) Reference: men	Women	-2.18	0.11	0.41	0.79
BMI (numeric)	body Measures Component Status (kg/m**2)	0.09	0.34	0.09	0.44
Creatinine (numeric)	Creatinine, urine (mg/dL)	-0.09	5.31e-16*	-0.06	2.77e-06 *
Diabetes (Categorical) Reference: No	Yes	1.27	0.51	-2.31	0.44
Smoking status Reference: No at all	Some day	-5.40	0.19	-5.49	0.08
	Not at all	-1.71	0.24	-0.41	0.81

Hypertension (Categorical) Reference: No	Yes	4.51	0.00158 *	---	--
Alcohol use (Categorical) Reference: No	Yes	0.93	0.62	-0.67	0.76

* Indicates P-value <0.05.

According to Table 12 and Figure 2, cadmium had the highest mean weight at 0.5, suggesting it had the strongest association with the outcome among the contaminants considered. Lead followed closely with a weight of 0.4 in the model with hypertension.

According to Table 12 and Figure 3, cadmium had the highest mean weight at 0.373, again suggesting it had the strongest association with the outcome among the contaminants considered. Lead closely followed with a weight of 0.277 in the model without hypertension.

These two indicators had the greatest impact on the outcome for both models. Arsenic, PFOA, and PFOS, on the other hand, had relatively low mean weights, indicating that they had negligible associations with the outcome in this model. The accompanying bar plot visually represented these weights, with substantial bars for cadmium and lead reflecting their higher weights, while arsenic, PFOA, and PFOS had almost imperceptible bars, indicating their minimal weights.

Table 12 Summary of WQS Index

Mix name	Mean weight with Hypertension	Mean weight without Hypertension
Cadmium	0.508	0.373
Lead	0.406	0.277
Arsenic	0.048	0.141
PFOA	0.020	0.132
PFOS	0.018	0.078

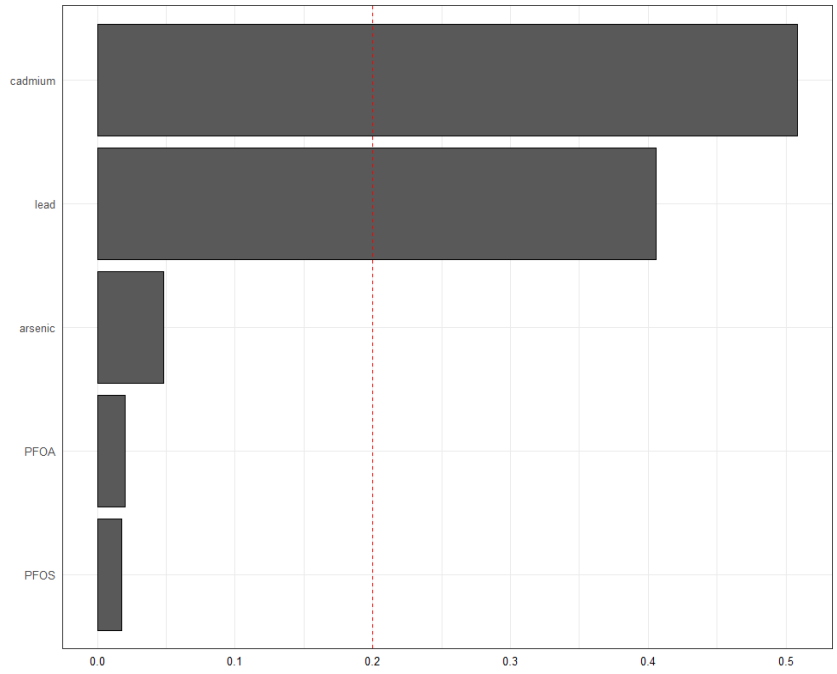

Figure 2 Bar Plot of WQS Index with Hypertension

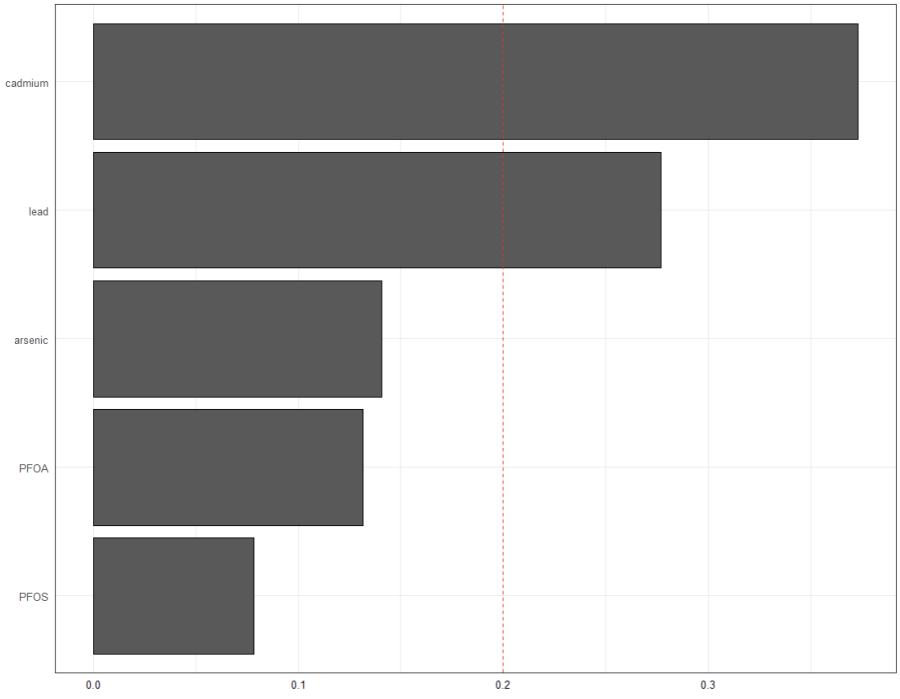

Figure 3 Bar Plot of WQS Index without Hypertension

Figure 4 suggested that there was a positive linear association between the WQS index and the outcome variable eGFR in the presented data set, indicating that as "wqs" increased, the "y_adj" variable also increased on average. The mixture of metals contributed to the increase in eGFR. This finding was consistent with previous information showing that the WQS index was a significant predictor of the outcome in the regression model.

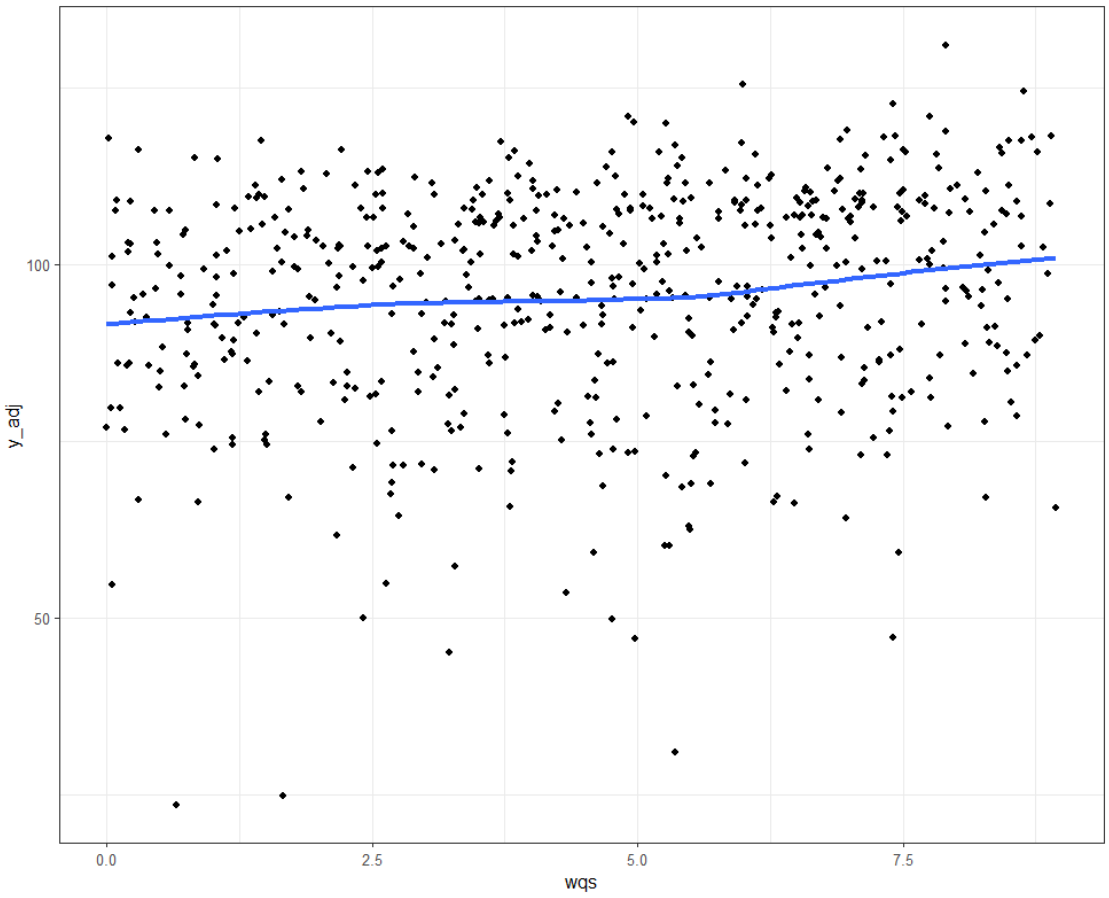

Figure 4 Scatter Plot of WQS Index with Hypertension

Figure 5, a scatter plot, displayed the relationship between metal and PFAS concentrations and numeric eGFR. The trend line in the figure showed a slight rise followed by a fall, suggesting that in the lower index concentration range, eGFR might slightly increase, but in the higher index concentration range, eGFR decreased instead. This trend indicated that high concentrations of metals could have a negative impact on kidney function.

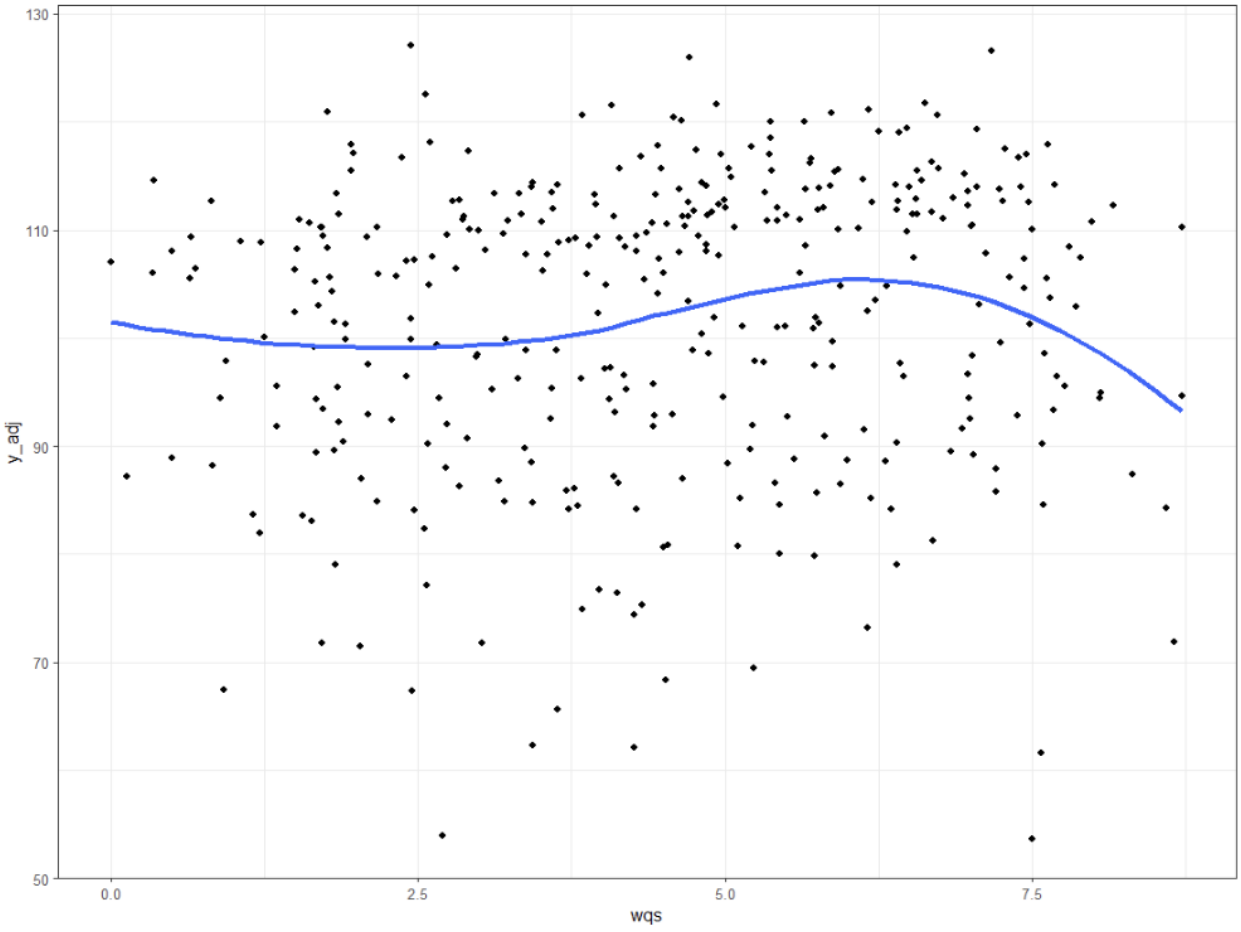

Figure 5 Scatter Plot of WQS Index without Hypertension

4.0 Discussion

The goal of this study was to identify relationships between metals and presence of CKD with characteristics of individual health status such as age, gender, BMI, smoking, diabetes, hypertension, and alcohol use under NHANES weighted survey design.

In summary, the logistic regression with age suggests that there were two statistically significant inverse relationships between urine lead and cadmium and the presence of CKD with 0.27 (95%CI: 0.09, 0.79) odds ratio for lead and 0.4 (95%CI: 0.23, 0.70) odds ratio for cadmium. Consistent with previous results, urinary levels of lead and cadmium increased with eGFR (Jin et al., 2018; Weaver et al., 2014). From weighted quantile sum, it shows that cadmium and lead were the two most contributing to numeric eGFR with mean weight 0.51 and 0.41 which were consistent with previous studies. However, arsenic, PFOA, and PFOS were not associated with the odds of presence of CKD. In contrast to the existing literature that has reported linear increases in urinary levels of arsenic and other metals with eGFR (Jin et al., 2018) and PFAS are inversely associated with kidney function in CKD (Conway et al., 2018). For covariates, we observed that age and presence of hypertension were two most significant characteristics among all the regressions models, what we could confirm were that as age increases, the odds ratios of CKD would also increase, similarly, compared with individual without hypertension, the odds ratios of CKD within individual with hypertension would also increase.

The logistic regression analysis without age suggests that PFOS has a statistically significant positive association with CKD, with an odds ratio of 1.02 (95% CI: 1.00, 1.04). This aligns with previous research indicating that each PFAS was inversely associated with kidney function in CKD (Conway et al., 2018). However, lead, cadmium, arsenic, and PFOA were not

associated with increased odds of CKD. Notably, removing age from the model resulted in significant changes to the odds ratio for lead and cadmium, potentially due to collinearity or because age may have a stronger confounding effect. Furthermore, smoking status, diabetes, and hypertension remain significant predictors in models without age. Hypertension's effect was similar to models with age, and the odds ratio for CKD among individuals with diabetes was higher compared to those without diabetes.

In WQS models without hypertension, lead and cadmium contribute less to numeric eGFR compared to WQS models with hypertension, with mean weights of 0.373 and 0.277 respectively. However, cadmium and lead are still the two most significant contributors to numeric eGFR. The trend in the scatter plot suggests that high concentrations of metals may have a negative impact on kidney function.

eGFR was assessed by serum creatinine, which may not be an ideal marker since it is typically used in conjunction with cystatin C, but this was not measured in NHANES. Data on covariates such as smoking, diabetes, hypertension, and alcohol use were collected from self-reported NHANES questionnaire data, which may lead to recall bias. This bias could be minimized by replacing self-reports with objective data, such as using medical records to verify the accuracy of reports. The global prevalence of chronic kidney disease (CKD) is estimated at 10% (Kovesdy, 2022), while the prevalence in the United States is 15% (Kidney Disease, 2023). However, in our data, the prevalence is only 4.2%. This discrepancy might be because the survey participants were relatively healthy and more capable and willing to take the survey.

Another critical limitation of this study is the inadequate sample size for explanatory variables. Sample size plays an important role in ensuring the statistical power to detect significant associations between variables. In this context, a larger sample size might have provided the

statistical power necessary to detect subtler associations, especially for arsenic, PFOA, and PFOS. Additionally, a small sample size could affect the generalizability of the findings to the broader population, as it may not represent the diversity within the whole population. In future research, it is possible to obtain more robust sample sizes without removing different datasets from production. For example, if we are studying the relationship between metals and CKD, there is no need to remove PFAS from the original dataset. This approach could lead to a more reliable sample size.

Previous studies (Sanders et al., 2019) have shown that metal levels in urine are inversely related to eGFR. However, this relationship no longer emerges when metal levels are assessed in blood. To verify this phenomenon in future analysis, two additional regression analyses could be conducted, using blood lead and cadmium as predictor variables, respectively, to confirm their association with eGFR. Based on our findings, there is a notable relationship between metals such as lead and cadmium and the lower odds of CKD. This relationship varies depending on factors such as age, hypertension, and diabetes, indicating that these individual characteristics may influence the impact of metal exposure on kidney function. In summary, the relationship between metal exposure and kidney function is intricate and may be influenced by a variety of covariates. Future studies could explore these associations in larger sample sizes to provide more statistical power and improve the generalizability of results. Additionally, considering other markers for eGFR could help improve the accuracy of the results. By conducting additional regression analyses with blood levels of metals, future studies can verify the observed relationships between urine metals and CKD, offering deeper insights into the biological mechanisms.

Appendix A Code in R

```{r}
library(dplyr)
library(ggplot2)
library(stringr)
library(stringi)
library(openxlsx)
library(tidyverse)
library(haven)
library(emmeans)
library(car)
library(survey)
library(table1)
library(gWQS)
library(jtools)

```

```{r}

Creatinine.serum15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/BIOPRO_I.XPT")

metal.urine15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/UMS_I.XPT")
arsenic.urine15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/UTASS_I.XPT")
PFAS15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/PFAS_I.XPT")

DEMO15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/DEMO_I.XPT")

bmi15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/BMX_I.XPT")
nicotine15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/SMQ_I.XPT")
diabete15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/DIQ_I.XPT")
Hypertension15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/BPQ_I.XPT")
```

```
alcohol15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/ALQ_I.XPT")
pressure15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/BPX_I.XPT")

creatin.urine15 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2015data/ALB_CR_I.XPT")
```

```{r}
nhanesAnalysis15 <- DEMO15 %>%
  select("SEQN","RIAGENDR","RIDAGEYR","SDMVSTRA","SDMVPSU") %>%
  rename(age = RIDAGEYR, gender = RIAGENDR, strata= SDMVSTRA, psu = SDMVPSU) %>%
  filter(age>=18) %>%
  mutate(gender = factor(gender, labels=c("Men", "Women")))


Creatinine15 <- Creatinine.serum15 %>% select(SEQN, LBXSCR) %>% rename(Creatinine.serum = LBXSCR)

lead_cad15 <- metal.urine15 %>% select (SEQN,URXUPB,URXUCD) %>% rename(lead = URXUPB, cadmium = URXUCD)
arsenic15 <- arsenic.urine15 %>% select(SEQN, URXUAS) %>% rename(arsenic = URXUAS)
pfas15 <- PFAS15 %>% select (SEQN,WTSB2YR,LBXNFOA,LBXNFOS) %>% rename(subsampleweight2 = WTSB2YR,PFOA = LBXNFOA, PFOS = LBXNFOS)


bmi.c15 <- bmi15 %>% select(SEQN, BMXBMI) %>% rename(BMI = BMXBMI)

smoking.c15 <- nicotine15 %>% select(SEQN, SMQ040) %>% rename(smoking_status = SMQ040)%>%
  filter(smoking_status == 1 | smoking_status == 2 | smoking_status == 3) %>%
  mutate(smoking_status = factor(smoking_status, labels=c("Every day", "Some days", "Not at all")))


diabete.c15 <- diabete15 %>% select(SEQN, DIQ010) %>% rename(have_diabetes = DIQ010)%>%
```

```
  filter(have_diabetes == 1 | have_diabetes == 2) %>%
  mutate(have_diabetes = factor(have_diabetes, labels=c("Yes", "No")))


hypertension.c15 <- Hypertension15 %>% select(SEQN, BPQ020) %>%
rename(have_hypertnesion = BPQ020)%>%
  filter(have_hypertnesion == 1 | have_hypertnesion == 2) %>%
  mutate(have_hypertnesion = factor(have_hypertnesion, labels=c("Yes", "No")))

alcohol.c15 <- alcohol15 %>% select(SEQN, ALQ101) %>% rename(have_drink = ALQ101)%>%
  filter(have_drink == 1 | have_drink == 2) %>%
  mutate(have_drink = factor(have_drink, labels=c("Yes", "No")))


creatin15 <- creatin.urine15 %>% select(SEQN, URXUCR) %>% rename(Creatinine.urine =
URXUCR)



```
```
```{r}
# primary outcome
one15 <- nhanesAnalysis15 %>%
   left_join(Creatinine15, by = "SEQN")   %>% drop_na()

#one

one15 <- one15 %>%
   left_join(lead_cad15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
   left_join(arsenic15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
   left_join(pfas15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
   left_join(bmi.c15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
   left_join(smoking.c15, by = "SEQN")   %>% drop_na()
```
```

```
one15 <- one15 %>%
  left_join(diabete.c15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
  left_join(hypertension.c15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
  left_join(alcohol.c15, by = "SEQN")   %>% drop_na()

one15 <- one15 %>%
  left_join(creatin15, by = "SEQN")   %>% drop_na()


one15 #205
```

```{r}


Creatinine.serum13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/BIOPRO_H.XPT")

metal.urine13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/UMS_H.XPT")
arsenic.urine13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/UTASS_H.XPT")
PFAS13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/SSPFAS_H.XPT")

DEMO13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/DEMO_H.XPT")

bmi13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/BMX_H.XPT")
nicotine13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/SMQ_H.XPT")
diabete13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/DIQ_H.XPT")
Hypertension13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/BPQ_H.XPT")
alcohol13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/ALQ_H.XPT")
pressure13 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2013data/BPX_H.XPT")

```

```{r}
nhanesAnalysis13 <- DEMO13 %>%
  select("SEQN","RIAGENDR","RIDAGEYR","SDMVSTRA","SDMVPSU") %>%
  rename(age = RIDAGEYR, gender = RIAGENDR, strata= SDMVSTRA, psu = SDMVPSU) %>%
  filter(age>=18) %>%
  mutate(gender = factor(gender, labels=c("Men", "Women")))



Creatinine13 <- Creatinine.serum13 %>% select(SEQN, LBXSCR) %>% rename(Creatinine.serum = LBXSCR)

lead_cad13 <- metal.urine13 %>% select (SEQN,URXUPB,URXUCD,URXUCR) %>% rename(lead = URXUPB, cadmium = URXUCD,Creatinine.urine = URXUCR)
arsenic13 <- arsenic.urine13 %>% select(SEQN, URXUAS) %>% rename(arsenic = URXUAS)
pfas13 <- PFAS13 %>% select (SEQN,WTSSBH2Y,SSNPFOA,SSNPFOS) %>% rename(subsampleweight2=WTSSBH2Y,PFOA = SSNPFOA, PFOS = SSNPFOS)


bmi.c13 <- bmi13 %>% select(SEQN, BMXBMI) %>% rename(BMI = BMXBMI)
smoking.c13 <- nicotine13 %>% select(SEQN, SMQ040) %>% rename(smoking_status = SMQ040)%>%
  filter(smoking_status == 1 | smoking_status == 2 | smoking_status == 3) %>%
  mutate(smoking_status = factor(smoking_status, labels=c("Every day", "Some days", "Not at all")))


diabete.c13 <- diabete13 %>% select(SEQN, DIQ010) %>% rename(have_diabetes = DIQ010)%>%filter(have_diabetes == 1 | have_diabetes == 2) %>%
  mutate(have_diabetes = factor(have_diabetes, labels=c("Yes", "No")))


hypertension.c13 <- Hypertension13 %>% select(SEQN, BPQ020) %>% rename(have_hypertnesion = BPQ020)%>%filter(have_hypertnesion == 1 | have_hypertnesion == 2) %>%
  mutate(have_hypertnesion = factor(have_hypertnesion, labels=c("Yes", "No")))

alcohol.c13 <- alcohol13 %>% select(SEQN, ALQ101) %>% rename(have_drink = ALQ101)%>%filter(have_drink == 1 | have_drink == 2) %>%
  mutate(have_drink = factor(have_drink, labels=c("Yes", "No")))
```

```{r}
# primary outcome
one13 <- nhanesAnalysis13 %>%
```

```
  left_join(Creatinine13, by = "SEQN")   %>% drop_na()

#one

one13 <- one13 %>%
  left_join(lead_cad13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(arsenic13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(pfas13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(bmi.c13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(smoking.c13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(diabete.c13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(hypertension.c13, by = "SEQN")   %>% drop_na()

one13 <- one13 %>%
  left_join(alcohol.c13, by = "SEQN")   %>% drop_na()


one13 #222
```

```{r}


Creatinine.serum11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/BIOPRO_G.XPT")

metal.urine11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/UHMS_G.XPT")
arsenic.urine11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/UASS_G.XPT")
PFAS11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/PFC_G.XPT")
```

```
DEMO11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/DEMO_G.XPT")

bmi11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/BMX_G.XPT")
nicotine11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/SMQ_G.XPT")
diabete11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/DIQ_G.XPT")
Hypertension11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/BPQ_G.XPT")
alcohol11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/ALQ_G.XPT")
pressure11 <- read_xpt("E:/University of Pittsburgh/Master Second Year/BIOST 2022 Thesis/2011data/BPX_G.XPT")

```

```{r}
nhanesAnalysis11 <- DEMO11 %>%
  select("SEQN","RIAGENDR","RIDAGEYR","SDMVSTRA","SDMVPSU") %>%
  rename(age = RIDAGEYR, gender = RIAGENDR, strata= SDMVSTRA, psu = SDMVPSU) %>%
  filter(age>=18) %>%
  mutate(gender = factor(gender, labels=c("Men", "Women")))



Creatinine11 <- Creatinine.serum11 %>% select(SEQN, LBXSCR) %>% rename(Creatinine.serum = LBXSCR)

lead_cad11 <- metal.urine11 %>% select (SEQN,URXUPB,URXUCD,URXUCR) %>% rename(lead = URXUPB, cadmium = URXUCD,Creatinine.urine=URXUCR)
arsenic11 <- arsenic.urine11 %>% select(SEQN, URXUAS) %>% rename(arsenic = URXUAS)
pfas11 <- PFAS11 %>% select (SEQN,WTSA2YR,LBXPFOA,LBXPFOS) %>% rename(subsampleweight2=WTSA2YR,PFOA = LBXPFOA, PFOS = LBXPFOS)


bmi.c11 <- bmi11 %>% select(SEQN, BMXBMI) %>% rename(BMI = BMXBMI)
smoking.c11 <- nicotine11 %>% select(SEQN, SMQ040) %>% rename(smoking_status = SMQ040)%>%
  filter(smoking_status == 1 | smoking_status == 2 | smoking_status == 3) %>%
  mutate(smoking_status = factor(smoking_status, labels=c("Every day", "Some days", "Not at all")))
```

```
diabete.c11 <- diabete11 %>% select(SEQN, DIQ010) %>% rename(have_diabetes = DIQ010)%>%filter(have_diabetes == 1 | have_diabetes == 2) %>%
  mutate(have_diabetes = factor(have_diabetes, labels=c("Yes", "No")))


hypertension.c11 <- Hypertension11 %>% select(SEQN, BPQ020) %>% rename(have_hypertnesion = BPQ020)%>%filter(have_hypertnesion == 1 | have_hypertnesion == 2) %>%
  mutate(have_hypertnesion = factor(have_hypertnesion, labels=c("Yes", "No")))

alcohol.c11 <- alcohol11 %>% select(SEQN, ALQ101) %>% rename(have_drink = ALQ101)%>%filter(have_drink == 1 | have_drink == 2) %>%
  mutate(have_drink = factor(have_drink, labels=c("Yes", "No")))
```

```{r}
# primary outcome
one11 <- nhanesAnalysis11 %>%
  left_join(Creatinine11, by = "SEQN")   %>% drop_na()

#one

one11 <- one11 %>%
  left_join(lead_cad11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(arsenic11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(pfas11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(bmi.c11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(smoking.c11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(diabete.c11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(hypertension.c11, by = "SEQN")   %>% drop_na()

one11 <- one11 %>%
  left_join(alcohol.c11, by = "SEQN")   %>% drop_na()
```

one11 #556
```
```{r}

one11 #556
one13 #222
one15 #205

final_data <- one11 %>%
   full_join(one13) %>%
  full_join(one15)

final_data
```


```{r}
calculate_eGFR <- function(Scr, Age, gender_f) {

  Scr <- as.numeric(as.character(Scr))
  Age <- as.numeric(as.character(Age))

  isfemale <- ifelse(gender_f == "Women", TRUE, FALSE)
  kappa <- ifelse(isfemale, 0.7, 0.9)
  alpha <- ifelse(isfemale, -0.241, -0.302)
  female <- ifelse(isfemale, 1.012, 1)

  eGFR <- 142 * (min(Scr / kappa, 1) ** alpha) * (max(Scr / kappa, 1) ** -1.2) * (0.994 ** Age)
* female
  return(eGFR)
}

#calculate_eGFR(0.68,43,"Women")

eGFR_values <- apply(final_data, 1, function(x) calculate_eGFR(x['Creatinine.serum'], x['age'],
x['gender']))

final_data$eGFR <- eGFR_values
final_data



final_data1 <- final_data %>%
  mutate(group = ifelse(eGFR < 60, 1, 0)) %>%
  mutate(CKD_condition = factor(group, level = c(0, 1), labels = c("no CKD", "CKD")))%>%

```
  mutate(subsampleweight6=1/3 * subsampleweight2)

final_data1

#sum(final_data1$arsenic == 0.88)


```

```{r}

#final_data1[final_data1$strata == 129,]

final_data1$strata <- ifelse(final_data1$strata == 128, 129, final_data1$strata)


final_data1$smoking_status <-factor(final_data1$smoking_status, levels = c("Not at all", "Some days", "Every day"))
final_data1$have_diabetes <-factor(final_data1$have_diabetes, levels = c("No", "Yes"))
final_data1$have_hypertnesion <-factor(final_data1$have_hypertnesion, levels = c("No", "Yes"))
final_data1$have_drink <-factor(final_data1$have_drink, levels = c("No", "Yes"))




nhanesDesign <- svydesign(id      = ~psu,
                          strata  = ~strata,
                          weights = ~subsampleweight6,
                          nest    = TRUE,
                          data    = final_data1)




table1(~. , data = final_data1%>%select(-SEQN,-strata,-psu,-subsampleweight2, -subsampleweight6,-group))



summarise(final_data1, mean=mean(eGFR), sd=sd(eGFR))
summarise(final_data1, mean=mean(arsenic), sd=sd(arsenic))
summarise(final_data1, mean=mean(BMI), sd=sd(BMI))
summarise(final_data1, mean=mean(age), sd=sd(age))
summarise(final_data1, mean=mean(Creatinine.urine), sd=sd(Creatinine.urine))
```

#hist(final_data1$Creatinine.serum)

```

```{r}

svymean(~eGFR, nhanesDesign, na.rm = TRUE)#97.906
sqrt(coef(svyvar(~eGFR, nhanesDesign)))#19.56278

svymean(~lead, nhanesDesign, na.rm = TRUE)#0.69587
sqrt(coef(svyvar(~lead, nhanesDesign)))#1.125008

svymean(~cadmium, nhanesDesign, na.rm = TRUE)#0.51189
sqrt(coef(svyvar(~cadmium, nhanesDesign)))#0.7294955

svymean(~arsenic, nhanesDesign, na.rm = TRUE)#12.711
sqrt(coef(svyvar(~arsenic, nhanesDesign)))#24.40041

svymean(~PFOA, nhanesDesign, na.rm = TRUE)#2.6557
sqrt(coef(svyvar(~PFOA, nhanesDesign)))#4.552953

svymean(~PFOS, nhanesDesign, na.rm = TRUE)#7.3932
sqrt(coef(svyvar(~PFOS, nhanesDesign)))#8.568516

svymean(~age, nhanesDesign, na.rm = TRUE)#48.252
sqrt(coef(svyvar(~age, nhanesDesign)))#15.5736

svymean(~BMI, nhanesDesign, na.rm = TRUE)#28.748
sqrt(coef(svyvar(~BMI, nhanesDesign)))#6.713616

svymean(~Creatinine.urine, nhanesDesign, na.rm = TRUE)
sqrt(coef(svyvar(~Creatinine.urine, nhanesDesign)))

svymean(~CKD_condition, nhanesDesign, na.rm = TRUE)
svymean(~gender, nhanesDesign, na.rm = TRUE)
svymean(~smoking_status, nhanesDesign, na.rm = TRUE)
svymean(~have_diabetes, nhanesDesign, na.rm = TRUE)
svymean(~have_hypertnesion, nhanesDesign, na.rm = TRUE)
svymean(~have_drink, nhanesDesign, na.rm = TRUE)

```
```

```{r}
# model_lead <- svyglm(CKD_condition ~ lead +
#                        age + gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
#                      family = quasibinomial,
#                      design = nhanesDesign,
#                      data = final_data1)


model_lead <- svyglm(CKD_condition ~ lead +
                       gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
                     family = quasibinomial,
                     design = nhanesDesign,
                     data = final_data1)

summ(model_lead,
 digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
 model.fit = getOption("summ-model.fit", FALSE),
 confint = getOption("summ-confint", TRUE),
 ci.width = getOption("summ-ci.width", 0.95))



# model_lead <- svyglm(eGFR ~ arsenic +
#                        age + gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
#                      design = nhanesDesign,
#                      data = final_data1)


summary(model_lead)

```

```{r}
# model_lead <- svyglm(CKD_condition ~ lead +
#                     age +
#                     gender
#                    + BMI
#                   + Creatinine.urine
#                  +have_diabetes
#                + have_hypertnesion
#                   + have_drink
```

```
#                   ,
#                         family = quasibinomial,
#                         design = nhanesDesign,
#                         data = final_data1)
#
#
#
# summ(model_lead,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))
```

```{r}

age_lead <- svyglm(CKD_condition ~ lead,
                         family = quasibinomial,
                         design = nhanesDesign,
                         data = final_data1)



summ(age_lead,
 digits = getOption("jtools-digits", default = 5), exp = T, vifs = F,
 model.fit = getOption("summ-model.fit", FALSE),
 confint = getOption("summ-confint", TRUE),
 ci.width = getOption("summ-ci.width", 0.95))




# summary(glm(lead ~ age, data = final_data1))
# summary(glm(age ~ lead, data = final_data1))

```

```{r}
# model_cad <- svyglm(CKD_condition ~ cadmium +
#                       age + gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
#                     family = quasibinomial,
#                     design = nhanesDesign,
#                     data = final_data1)


model_cad <- svyglm(CKD_condition ~ cadmium +
                    gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
                    family = quasibinomial,
                    design = nhanesDesign,
                    data = final_data1)

summary(model_cad)

# exp(-4.243212)
# 1-exp(-4.243212)

summary_model <- summary(model_cad)


coefficients <- summary_model$coefficients[, 1]  # Estimates are usually in the first column
std_errors <- summary_model$coefficients[, 2]    # Std. Errors are usually in the second column

odds_ratios <- exp(coefficients)

ci_lower <- exp(coefficients - 1.96 * std_errors)
ci_upper <- exp(coefficients + 1.96 * std_errors)

results <- data.frame(
  Odds_Ratio = odds_ratios,
  CI_Lower = ci_lower,
  CI_Upper = ci_upper
)

print(results)
round(results, 2)

```



```{r}
```

```
# model_cad <- svyglm(CKD_condition ~ cadmium +
#                           age + gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
#                           family = quasibinomial,
#                           design = nhanesDesign,
#                           data = final_data1)
#
#
# summ(model_cad,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))
#
#
#
#
# model_cad <- svyglm(CKD_condition ~ age,
#                   family = quasibinomial,
#                           design = nhanesDesign,
#                           data = final_data1)
#
#
# summ(model_cad,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = F,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))
#
#
# summary(glm(cadmium ~ age, data = final_data1))
# summary(glm(age ~ cadmium, data = final_data1))

```
```{r}
# model_ars <- svyglm(CKD_condition ~ arsenic +
#                           age + gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
```

```
#                       family = quasibinomial,
#                       design = nhanesDesign,
#                       data = final_data1)


model_ars <- svyglm(CKD_condition ~ arsenic +
                    gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
                    family = quasibinomial,
                    design = nhanesDesign,
                    data = final_data1)


summary(model_ars)

# exp(-4.243212)
# 1-exp(-4.243212)

summary_model <- summary(model_ars)


coefficients <- summary_model$coefficients[, 1]  # Estimates are usually in the first column
std_errors <- summary_model$coefficients[, 2]    # Std. Errors are usually in the second column

odds_ratios <- exp(coefficients)

ci_lower <- exp(coefficients - 1.96 * std_errors)
ci_upper <- exp(coefficients + 1.96 * std_errors)

results <- data.frame(
  Odds_Ratio = odds_ratios,
  CI_Lower = ci_lower,
  CI_Upper = ci_upper
)

print(results)
round(results, 2)

```


```{r}

# model_ars <- svyglm(CKD_condition ~ arsenic + age
#                       + gender + BMI + Creatinine.urine+have_diabetes + have_hypertnesion +
have_drink,
#                       family = quasibinomial,
```

```
#                         design = nhanesDesign,
#                         data = final_data1)
#
#
# summ(model_ars,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))
#
#
#
# summary(glm(arsenic ~ age, data = final_data1))
# summary(glm(age ~ arsenic, data = final_data1))
#
#
# model_ars <- svyglm(CKD_condition ~ arsenic + age
#                           + gender + BMI + Creatinine.urine+have_diabetes + have_hypertnesion +
have_drink,
#                         family = quasibinomial,
#                         design = nhanesDesign,
#                         data = final_data1)
#
#
# summ(model_ars,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))


```

```{r}
# model_pfoa <- svyglm(CKD_condition ~PFOA +


```

```
#                           age + gender + BMI + Creatinine.urine+smoking_status+have_diabetes +
have_hypertnesion + have_drink,
#                         family = quasibinomial,
#                         design = nhanesDesign,
#                         data = final_data1)


model_pfoa <- svyglm(CKD_condition ~PFOA +
                       gender + BMI + Creatinine.urine+smoking_status+have_diabetes +
have_hypertnesion + have_drink,
                     family = quasibinomial,
                     design = nhanesDesign,
                     data = final_data1)

# summ(model_pfoa,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))



# summary(glm(PFOA ~ age, data = final_data1))
# summary(glm(age ~ PFOA, data = final_data1))

# model_pfoa <- svyglm(CKD_condition ~PFOA,
#                         family = quasibinomial,
#                         design = nhanesDesign,
#                         data = final_data1)
#
summary(model_pfoa)




summary_model <- summary(model_pfoa)


coefficients <- summary_model$coefficients[, 1]  # Estimates are usually in the first column
std_errors <- summary_model$coefficients[, 2]    # Std. Errors are usually in the second column

odds_ratios <- exp(coefficients)

ci_lower <- exp(coefficients - 1.96 * std_errors)
ci_upper <- exp(coefficients + 1.96 * std_errors)
```

```
results <- data.frame(
  Odds_Ratio = odds_ratios,
  CI_Lower = ci_lower,
  CI_Upper = ci_upper
)

print(results)
round(results, 2)
```

```{r}
# model_pfos <- svyglm(CKD_condition ~  PFOS +
#                        age + gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
#                      family = quasibinomial,
#                      design = nhanesDesign,
#                      data = final_data1)

# model_pfos <- svyglm(CKD_condition ~  PFOS,
#                      family = quasibinomial,
#                      design = nhanesDesign,
#                      data = final_data1)


model_pfos <- svyglm(CKD_condition ~  PFOS +
                gender + BMI + Creatinine.urine +smoking_status+have_diabetes +
have_hypertnesion + have_drink,
                     family = quasibinomial,
                     design = nhanesDesign,
                     data = final_data1)

# summ(model_pfos,
#  digits = getOption("jtools-digits", default = 5), exp = T, vifs = T,
#  model.fit = getOption("summ-model.fit", FALSE),
#  confint = getOption("summ-confint", TRUE),
#  ci.width = getOption("summ-ci.width", 0.95))
```

```
# summary(glm(PFOS ~ age, data = final_data1))
# summary(glm(age ~ PFOS, data = final_data1))


summary(model_pfos)



summary_model <- summary(model_pfos)


coefficients <- summary_model$coefficients[, 1]  # Estimates are usually in the first column
std_errors <- summary_model$coefficients[, 2]    # Std. Errors are usually in the second column

odds_ratios <- exp(coefficients)

ci_lower <- exp(coefficients - 1.96 * std_errors)
ci_upper <- exp(coefficients + 1.96 * std_errors)

results <- data.frame(
  Odds_Ratio = odds_ratios,
  CI_Lower = ci_lower,
  CI_Upper = ci_upper
)

print(results)
round(results, 2)
```



```{r}
ggplot(data=final_data1, aes(x=eGFR, y=lead)) +
  geom_point() +
  labs(title="eGFR and lead") + xlab("eGFR (mL/min/1.73m^2)") + ylab("lead (µg/L)")



ggplot(data=final_data1, aes(x=eGFR, y=cadmium)) +
  geom_point() +
  labs(title="eGFR and cadmium") + xlab("eGFR (mL/min/1.73m^2)") + ylab("cadmium (µg/L)")

ggplot(data=final_data1, aes(x=eGFR, y=arsenic)) +
  geom_point() +
  labs(title="eGFR and arsenic") + xlab("eGFR (mL/min/1.73m^2)") + ylab("arsenic (µg/L)")
```

```
ggplot(data=final_data1, aes(x=eGFR, y=PFOA)) +
  geom_point() +
  labs(title="eGFR and PFOA") + xlab("eGFR (mL/min/1.73m^2)") + ylab("PFOA (µg/L)")

ggplot(data=final_data1, aes(x=eGFR, y=PFOS)) +
  geom_point() +
  labs(title="eGFR and PFOS") + xlab("eGFR (mL/min/1.73m^2)") + ylab("PFOS (µg/L)")



summary(final_data1$arsenic)



```


```{r}
ggplot(data=final_data1, aes(x=eGFR, y=lead)) +
  geom_point() +
  labs(title="eGFR and lead") + xlab("eGFR (mL/min/1.73m^2)") + ylab("lead (µg/L)")

ggplot(data=final_data1, aes(x=eGFR, y=cadmium)) +
  geom_point() +
  labs(title="eGFR and cadmium") + xlab("eGFR (mL/min/1.73m^2)") + ylab("cadmium (µg/L)")

ggplot(data=final_data1, aes(x=eGFR, y=log10(arsenic))) +
  geom_point() +
  labs(title="eGFR and arsenic") + xlab("eGFR (mL/min/1.73m^2)") + ylab("arsenic (µg/L)")

ggplot(data=final_data1, aes(x=eGFR, y=PFOA)) +
  geom_point() +
  labs(title="eGFR and PFOA") + xlab("eGFR (mL/min/1.73m^2)") + ylab("PFOA (µg/L)")

ggplot(data=final_data1, aes(x=eGFR, y=PFOS)) +
  geom_point() +
  labs(title="eGFR and PFOS") + xlab("eGFR (mL/min/1.73m^2)") + ylab("PFOS (µg/L)")
```



```{r}

metals <- c(names(final_data1)[7:8], names(final_data1)[10],names(final_data1)[12:13])
metals
```

```
test <- final_data1
test[metals] <- lapply(test[metals], log10)
#test$eGFR <- log10(test$eGFR)


# ggplot(data=test, aes(x=eGFR, y=lead)) + geom_point()
# ggplot(data=test, aes(x=eGFR, y=cadmium)) + geom_point()
# ggplot(data=test, aes(x=eGFR, y=arsenic)) + geom_point()
# ggplot(data=test, aes(x=eGFR, y=PFOA)) + geom_point()
# ggplot(data=test, aes(x=eGFR, y=PFOS)) + geom_point()

# second_outcome <- gwqs(eGFR ~ wqs+
#                          age+ gender + BMI +Creatinine.urine+
#                          have_diabetes+smoking_status + have_hypertnesion + have_drink,
#                        mix_name = metals, data = test,
#                        q = 5, validation = 0.6, b = 10, b1_pos = FALSE, rh = 10,
#                        family = "gaussian", seed = 3024)


second_outcome <- gwqs(eGFR ~ wqs+
                         age+ gender + BMI +Creatinine.urine+
                         have_diabetes+smoking_status +have_hypertnesion+ have_drink,
                       mix_name = metals, data = test,
                       q = 10, validation = 0.6, b = 100, b1_pos = TRUE,
                       family = "gaussian", seed = 2017)

gwqs_scatterplot(second_outcome)


summary(second_outcome)

gwqs_weights_tab(second_outcome)

gwqs_barplot(second_outcome)


```
```{r}

metals <- c(names(final_data1)[7:8], names(final_data1)[10],names(final_data1)[12:13])
metals


test <- final_data1
```

```
test[metals] <- lapply(test[metals], log10)

test1 <- test %>% filter(have_hypertnesion == "No")


second_outcome <- gwqs(eGFR ~ wqs+
                         age+ gender + BMI +Creatinine.urine+
                         have_diabetes+smoking_status + have_drink,
                       mix_name = metals, data = test1,
                       q = 10, validation = 0.6, b = 100, b1_pos = TRUE,
                       family = "gaussian", seed = 2017)

gwqs_scatterplot(second_outcome)


summary(second_outcome)

gwqs_weights_tab(second_outcome)

gwqs_barplot(second_outcome)

```

Bibliography

Altamura, S., Pietropaoli, D., Lombardi, F., Del Pinto, R., & Ferri, C. (2023). An Overview of Chronic Kidney Disease Pathophysiology: The Impact of Gut Dysbiosis and Oral Disease. *Biomedicines*, *11*(11). https://doi.org/10.3390/biomedicines11113033

American Diabetes, A. (2010). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, *33 Suppl 1*(Suppl 1), S62-69. https://doi.org/10.2337/dc10-S062

Carrico, C., Gennings, C., Wheeler, D. C., & Factor-Litvak, P. (2015). Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. *J Agric Biol Environ Stat*, *20*(1), 100-120. https://doi.org/10.1007/s13253-014-0180-3

Cassy, S. R., Natário, I., & Martins, M. R. (2016). Logistic Regression Modelling for Complex Survey Data with an Application for Bed Net Use in Mozambique. Open Journal of Statistics, 6(5), Article 5. https://doi.org/10.4236/ojs.2016.65074

Centers for Disease Control and Prevention. (2024, April.). *Nhanes tutorials - sample design module*. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/nchs/nhanes/tutorials/SampleDesign.aspx

Centers for Disease Control and Prevention. (2017). *BIOPRO_I*. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/BIOPRO_I.htm

Centers for Disease Control and Prevention. (2018). *PFAS_I*. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/PFAS_I.htm

Centers for Disease Control and Prevention. (2018). *UMS_I*. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/UMS_I.htm

Centers for Disease Control and Prevention. (2022, March 24). *Data sources and data analysis*. Centers for Disease Control and Prevention. https://www.cdc.gov/exposurereport/data_sources_analysis.html

Conway, B. N., Badders, A. N., Costacou, T., Arthur, J. M., & Innes, K. E. (2018). Perfluoroalkyl substances and kidney function in chronic kidney disease, anemia, and diabetes. *Diabetes Metab Syndr Obes*, *11*, 707-716. https://doi.org/10.2147/DMSO.S173809

Delgado, C., Baweja, M., Crews, D. C., Eneanya, N. D., Gadegbeku, C. A., Inker, L. A., Mendu, M. L., Miller, W. G., Moxey-Mims, M. M., Roberts, G. V., St Peter, W. L., Warfield, C., & Powe, N. R. (2022). A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. *Am J Kidney Dis*, *79*(2), 268-288 e261. https://doi.org/10.1053/j.ajkd.2021.08.003

Hsueh, Y. M., Chung, C. J., Shiue, H. S., Chen, J. B., Chiang, S. S., Yang, M. H., Tai, C. W., & Su, C. T. (2009). Urinary arsenic species and CKD in a Taiwanese population: a case-control study. *Am J Kidney Dis*, *54*(5), 859-870. https://doi.org/10.1053/j.ajkd.2009.06.016

Inker, L. A., Astor, B. C., Fox, C. H., Isakova, T., Lash, J. P., Peralta, C. A., Kurella Tamura, M., & Feldman, H. I. (2014). KDOQI US commentary on the 2012 KDIGO clinical practice guideline for the evaluation and management of CKD. *Am J Kidney Dis*, *63*(5), 713-735. https://doi.org/10.1053/j.ajkd.2014.01.416

Jin, R., Zhu, X., Shrubsole, M. J., Yu, C., Xia, Z., & Dai, Q. (2018). Associations of renal function with urinary excretion of metals: Evidence from NHANES 2003-2012. *Environ Int*, *121*(Pt 2), 1355-1362. https://doi.org/10.1016/j.envint.2018.11.002

Kidney disease: The basics. National Kidney Foundation. (2023, November 16). https://www.kidney.org/news/newsroom/fsindex#:~:text=Kidney%20disease%2C%20also%20known%20as,than%201%20in%207%20adults)

Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Suppl (2011)*, *12*(1), 7-11. https://doi.org/10.1016/j.kisu.2021.11.003

Mayo Foundation for Medical Education and Research. (2023, September 6). *Chronic kidney disease*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/chronic-kidney-disease/symptoms-causes/syc-20354521

Nan, Y., Yang, J., Yang, J., Wei, L., & Bai, Y. (2023). Associations Between Individual and Combined Metal Exposures in Whole Blood and Kidney Function in U.S. Adults Aged 40 Years and Older. *Biol Trace Elem Res*. https://doi.org/10.1007/s12011-023-03722-z

Renzetti, S., Gennings, C., & Calza, S. (2023). A weighted quantile sum regression with penalized weights and two indices. *Front Public Health*, *11*, 1151821. https://doi.org/10.3389/fpubh.2023.1151821

Sanders, A. P., Mazzella, M. J., Malin, A. J., Hair, G. M., Busgang, S. A., Saland, J. M., & Curtin, P. (2019). Combined exposure to lead, cadmium, mercury, and arsenic and kidney health in adolescents age 12-19 in NHANES 2009-2014. *Environ Int*, *131*, 104993. https://doi.org/10.1016/j.envint.2019.104993

Weaver, V. M., Vargas, G. G., Silbergeld, E. K., Rothenberg, S. J., Fadrowski, J. J., Rubio-Andrade, M., Parsons, P. J., Steuerwald, A. J., Navas-Acien, A., & Guallar, E. (2014). Impact of urine concentration adjustment method on associations between urine metals and estimated glomerular filtration rates (eGFR) in adolescents. *Environ Res*, *132*, 226-232. https://doi.org/10.1016/j.envres.2014.04.013