

Exploring the Impact of Dietary and Environmental Factors on Hypertension and Mercury Exposure

2024-12-15

Danish Maknojia, Mona Saeed, Leonard Eshun, Su Zhang

Abstract

Investigating factors influencing blood pressure and mercury levels in the US population is crucial for public health interventions and environmental policy-making. This study aims to examine the relationships between dietary sodium intake and blood pressure levels, as well as factors that influence blood mercury levels. The dataset analyzed is derived from the Center for Disease Control (CDC) survey data collected through interviews, blood tests, and medical examinations from August 2021 to August 2023. Multinomial regression was used to model blood pressure levels, and linear regression was applied to model blood mercury levels. The analysis revealed no significant association between dietary sodium consumption and blood pressure levels. However, age emerged as a significant predictor of blood pressure, with older individuals being more likely to fall into the “Normal” category ($p < 0.05$). Gender and race were also significant, with Non-Hispanic White individuals showing a significantly higher likelihood of being classified as “Normal” ($p = 0.000$). Additionally, females were found to be more likely to develop Stage 1 Hypertension ($p < 0.05$). Tap water intakes, age, and income level have significant positive associations with blood mercury level ($p < 0.001$). Non-Hispanic Asians have significantly higher mercury level (289.5%) compared with reference group of Mexicans Americans. Females on average have mercury levels 8% lower compared to males. Income impacts mercury levels differently across racial groups, with Non-Hispanic Whites showing a significant increase (10.2%) and Non-Hispanic Asians showing a significant decrease (10.5%) in mercury levels as income rises. Our study identified key water intake and demographic factors affecting blood mercury levels, which can inform public health policies regarding water consumption and exposure to environmental toxins.

Introduction

Hypertension and environmental exposure to mercury are significant public health concerns worldwide (National Center for Biotechnology Information, 2021). With hypertension, affecting over 1.28 billion adults worldwide with a majority in low- and middle-income countries, and environmental mercury exposure, linked to cardiovascular and neurological issues such as elevated blood pressure even at low levels, are significant global public health concerns (World Health Organization; Houston, p 621). Understanding the factors that influence these conditions is critical to improving health outcomes and developing targeted interventions. This study leverages data from the U.S Centers for Disease Control and Prevention (CDC), a nationally representative dataset, to explore two key research questions that address these pressing health issues.

The first research question examines the association between dietary sodium intake and blood pressure levels, with a particular focus on whether the effect varies by factors such as gender and race. Blood pressure is an ordinal variable, categorized into four levels according to CDC guidelines, ranging from normal to Stage 2 hypertension (CDC, 2023). High dietary sodium intake has long been implicated as a risk factor for hypertension, a leading contributor to cardiovascular diseases globally (World Health Organization, 2021). Gender and race may further modulate this relationship due to physiological and social determinants of health, making it essential to explore these interactions. Answering this question provides insights that can inform dietary guidelines and public health policies aimed at reducing hypertension-related morbidity and mortality.

The second research question investigates the impact of drinking water sources, such as tap or bottled water, on blood mercury levels. Blood mercury is a continuous variable that reflects exposure to mercury, a toxic heavy metal with known adverse effects on neurological and cardiovascular health (Agency for Toxic Substances and Disease Registry, 2023). Factors such as race may influence this relationship, potentially due to disparities in environmental exposure or differences in water consumption patterns. This question is vital because understanding the relationship between water sources and mercury exposure can guide environmental regulations and public health recommendations, particularly in vulnerable populations.

These research questions address two critical dimensions of health: dietary and environmental factors. By exploring the interplay between dietary sodium, drinking water sources, and key demographic factors, this study aims to provide actionable insights into mitigating the risks of hypertension and mercury toxicity. The findings have the potential to inform individual behavioral changes, public health interventions, and broader environmental and dietary policies.

Methods

Data Overview and Preprocessing

The dataset used for the first research question consists of information on 7,801 observations, including variables such as sodium intake, blood pressure categories, gender, race, age, and alcohol consumption. The outcome variable, **blood pressure category**, is ordinal and categorized as *Normal*, *Elevated*, *Stage 1 Hypertension*, and *Stage 2 Hypertension* based on both the systolic and diastolic blood measurements as illustrated in the table below. An additional **Not Classified** class was added to account for measurements that don't fall into any of the ranges below. There were 1830 observations with missing values for alcohol consumption and sodium intake and 323 observations have missing values in systolic and diastolic blood pressure measurements.

Blood Pressure Category	Systolic Blood Pressure	Diastolic Blood Pressure
Normal	<120 mmHg	and <80 mmHg
Elevated	120-129 mmHg	and <80 mmHg
Stage 1 Hypertension	130-139 mmHg	or 80-89 mmHg
Stage 2 Hypertension	140 mmHg	or 90 mmHg

The systolic and diastolic blood measurements used in this analysis represent the mean of the three measurements taken for each participant in the original dataset. These measurements were taken on three consecutive days. Predictor variables include continuous variables (e.g., sodium intake, age, alcohol consumption) and categorical variables (e.g., gender, race).

For the second research question on blood mercury level, the data has 7785 observations, among which 2870 observations have missing values in either predictor or outcome variables. This dataset has missing values in variables of mercury level, income level, and water intakes of bottled water and tap water.

Stochastic regression imputations were used to handle missing values because (1) dataset size is large enough to build reliable imputation models; (2) missing values appear in both predictors and outcome variables; (3) maintains the natural associations in the data; (4) better than median or mean imputations, which are more arbitrary and underestimate variance. Extreme outliers in sodium intake and alcohol consumption were identified using visualizations (boxplots) and summary statistics.

Model Fitting and Evaluation

To examine the relationships between sodium intake and blood pressure categories, a multinomial logistic regression model was developed. The initial model included main effects for sodium intake, age, gender, race, and alcohol consumption. Interaction terms between sodium intake and gender, as well as sodium intake and race, were later introduced to assess whether the effect of sodium intake on blood pressure categories varied across these demographic subgroups. This allowed for the evaluation of potential effect modification.

Model comparison was performed using the Akaike Information Criterion (AIC), which balances goodness of fit with model complexity. By comparing AIC values for multinomial regression models with and without interaction terms, we determined whether the improved fit provided by the interaction terms justified the added complexity. Additionally, a confusion matrix was used to evaluate the predictive performance of the multinomial logistic regression model.

For modeling blood mercury levels, a multiple linear regression (MLR) model was employed. Residual plots were analyzed to assess linearity and homoscedasticity, while adjusted R-squared was used to evaluate the model's explanatory power. Variance Inflation Factors (VIFs) were calculated to detect multicollinearity among predictors. We also utilized Cook's distance to examine influential points, and refitted models after removing the influential points to see if that improved the model performance.

All statistical analyses and modeling were conducted using the R programming language. Model fit for both the multinomial and linear regression models was evaluated using AIC and deviance as key metrics.

Results

Research Question 1: Contributing Factors of Blood Pressure Category

Exploratory data analysis (EDA) revealed several important insights. First, the sodium intake variable contained outliers with extremely high values, which could potentially skew the results. To address this, we limited sodium intake values to a maximum of 15,000 mg. This threshold was chosen based on the distribution of the data and practical dietary considerations. Second, we observed that the racial group "Other Race" had no observations in the dataset, leading to its exclusion from further analysis. Consequently, the analysis was restricted to six racial groups to ensure meaningful and interpretable results. Third, we noted that the "Normal" blood pressure category had the highest number of observations among all categories. This imbalance suggests that the multinomial regression model may estimate higher probabilities for the "Normal" category compared to others, reflecting the

underlying distribution of the data. The table below shows the summary statistics of our continuous variables.

Variable Name	Minimum	Maximum	Mean	Standard Deviation
Sodium Intake (mg)	0	14603	3054.1	1373.85
Alcohol Consumption	-0.585	448.1	6.02	17.68
Age	8	80	44.83	22.72

The multinomial regression model was applied to analyze the impact of various predictors on five blood pressure categories: Elevated, Normal, Stage 1 Hypertension, Stage 2 Hypertension, and Not Classified. The predictors included Sodium intake, Age, Gender, Race, and Alcohol consumption. The model shows varying effects for different predictors; **Sodium intake** does not significantly influence the probability of being classified into any blood pressure category, as indicated by a coefficient of zero for all categories except “Normal” (p-value > 0.05). Similarly, **Age** does not significantly impact the likelihood of being in these categories (p-value > 0.05), but for the “Normal” category, the model shows a small positive coefficient, suggesting older individuals are slightly more likely to be classified as “Normal”. **Gender** plays a significant role in predicting blood pressure categories, especially for “Stage 1 Hypertension” (p-value = 0). The positive coefficient for females indicates they may have a higher likelihood of being classified into this category compared to other blood pressure categories. Gender also shows smaller negative effects for other categories, suggesting some gender-related differences, but with relatively low magnitude (p-value < 0.05).

Race has the most substantial influence on health classifications. **Non-Hispanic White** individuals are significantly more likely to be classified as “Normal” (coefficient = 3.458, p-value = 0.000), while **Non-Hispanic Black**, **Non-Hispanic Asian**, and **Multiracial** categories show smaller but significant coefficients, suggesting race influences blood pressure classifications. **Other Hispanic** shows a modest positive effect for the “Normal” category (p-value < 0.05), while non-White groups are slightly more likely to be classified into higher stages of hypertension. Finally, **Alcohol consumption** has a minor but significant effect on blood pressure categories, particularly in the “Normal” and “Stage 1 Hypertension” categories. A small negative coefficient indicates that as alcohol consumption increases, the likelihood of being classified as “Normal” or “Stage 1 Hypertension” slightly decreases (p-value < 0.05). For a more detailed coefficients and p-values tables refer to **Appendix 1-A**.

The model’s performance, as shown by the confusion matrix below, reveals that the “Normal” category was most accurately predicted, with 3689 correct predictions. However, the model struggled with **Elevated** and **Not Classified** categories, failing to predict any

instances correctly. It also misclassified **Stage 1 Hypertension** and **Stage 2 Hypertension**, though with lower frequency. Overall, the model has difficulty distinguishing between certain categories, particularly **Elevated** and **Not Classified**.

To assess multicollinearity, Variance Inflation Factors (VIFs) were calculated for each predictor (Appendix 1-C). All VIF values were below the threshold of 5, indicating that multicollinearity is not a major issue in the model. **Non-Hispanic White** had the highest VIF (3.46), suggesting a moderate level of correlation with other predictors, but it does not indicate a severe problem.

Interaction terms between sodium intake and race as well as sodium intake and gender were examined; it was found that they did not improve the model’s performance. Both models, with and without interaction terms, achieved the same AIC of 16826.49, suggesting no improvement in the model fit. Additionally, introducing these interaction terms led to high VIFs, indicating multicollinearity (Appendix 1-D). Given these issues, we decided to retain the model without interaction terms. For detailed coefficients, p-values, confusion matrix, and VIF tables of the multinomial model with interaction terms, please refer to **Appendix 1-B**.

Table 3: Confusion Matrix

	Elevated	Normal	Not Classified	Stage 1 Hypertension	Stage 2 Hypertension
Elevated	0	0	0	0	0
Normal	627	3688	58	1201	364
Not Classified	0	0	0	0	0
Stage 1 Hypertension	323	535	34	532	373
Stage 2 Hypertension	9	9	1	18	25

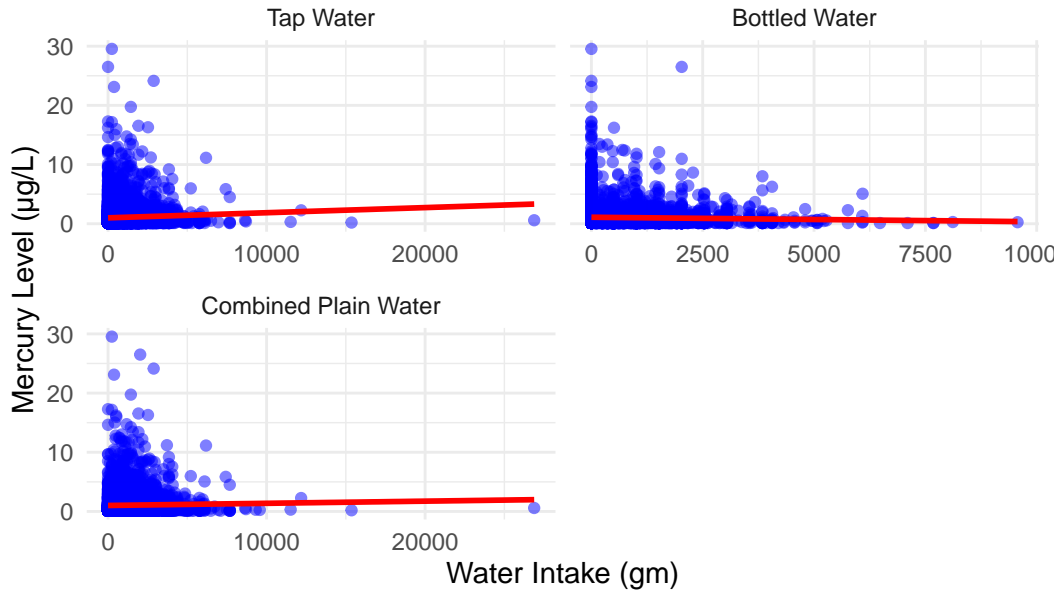
Research Question 2: Contributing Factors of Blood Mercury Level

Our statistical approach employed multiple linear regression with log-transformed mercury levels as the dependent variable, addressing both the non-normal distribution of mercury concentrations and the presence of detection limits in the measurements. The model incorporated water consumption variables (tap and bottled water intake), demographic characteristics (race, gender, age), and socioeconomic status (income-to-poverty ratio) and interaction term between race and income level to capture the complex relationships between these factors. Below is a summary descriptive statistics of continuous predictor and outcome variables. For categorical variables, we included a summary statistics table in Appendix 2-A.

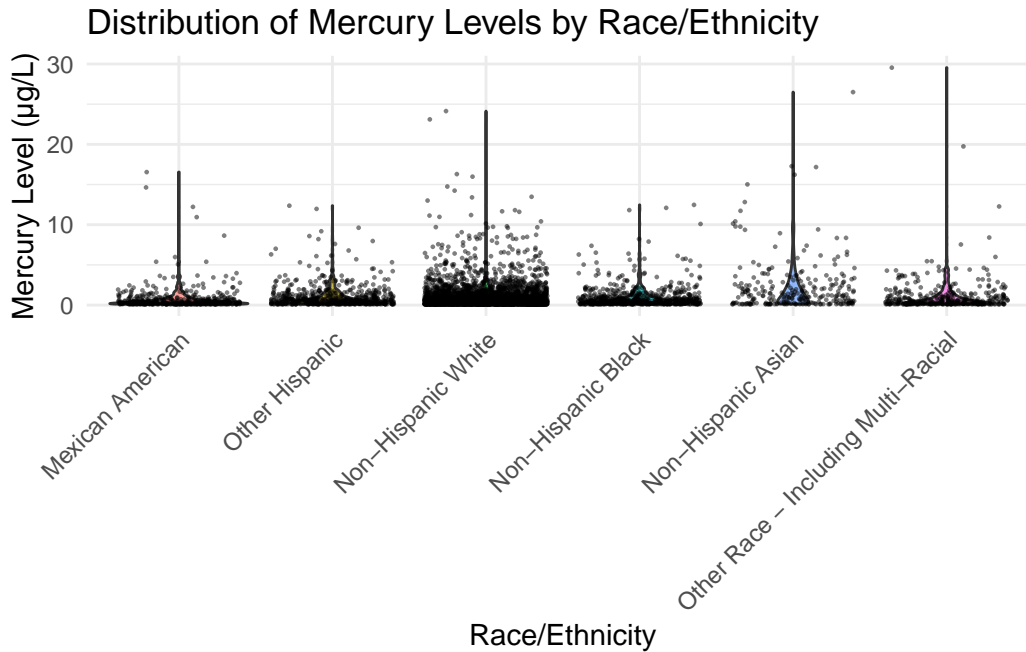
Table 4: Summary Statistics of Variables

Variable	Mean	Standard Deviation	Median	Missing Values
Age	44.83	22.72	47.00	0
Ratio of Family Income to Poverty	2.79	1.66	2.61	1000
Tap Water Drank Yesterday (gm)	732.11	1138.27	240.00	1830
Plain Water Drank Yesterday (gm)	1248.70	1204.79	1014.00	1830
Bottled Water Drank Yesterday (gm)	516.60	891.93	0.00	1830
Blood Mercury Level (ug/L)	1.04	1.84	0.45	672

Effect of kinds of water consumed on Mercury Levels



1. **Mercury Levels vs. Water Consumption** The faceted scatter plot examined the relationships between mercury levels and water intake across three types of water consumption (tap water, bottled water, and combined plain water). Linear trend lines showed that higher consumption of tap water might led to a slightly increase in blood mercury levels in people versus bottled water which had no change in blood mercury levels as more is consumed.



2. **Mercury Levels by Race/Ethnicity** The violin plot was used to compare mercury distributions across racial/ethnic groups. This graph shows that mercury levels vary significantly across racial and ethnic groups, with Non-Hispanic Asians & Other Race - Including Multi-Racial having the highest levels and the widest distribution. We can see that most distributions are asymmetrical, meaning that most individuals in all racial categories have low mercury levels, but there are noticeable outliers, especially among Non-Hispanic Asians, Other Race - Including Multi-Racial & Non-Hispanic Blacks.

We initially fit a multiple linear regression model towards the blood mercury levels. However, in the residuals vs fitted graph (Appendix 2-B), the spread of residuals increases as the fitted values grow, indicating heteroscedasticity. The Q-Q plot (Appendix 2-B) deviates significantly from the 45-degree reference line, particularly at the tails of the distribution, indicating that the residuals are not normally distributed. These suggest a potential violation of the normality assumption. Therefore, we log-transformed the dependent variable to stabilize the variance and help normalize the residuals.

When fitting the model, we included interaction term between race and family income to poverty level, because the influence of income on health outcomes such as mercury levels might vary by race due to differences in access to healthcare, dietary habits, or exposure to environmental risk factors and we want to take the heterogeneous effects into considerations. We also examined the influential points using Cook's distance and marked the top 5 influential points (Appendix 2-C). We then refit the model again after removing

these influential points (the model results outputs and residual plots can be found in Appendix 2-D). However, removing influential points did not significantly change the model results (adjusted R-squared and p-values). Therefore, we decided to keep these influential points in the model. Hence, the final model was to fit logged blood mercury level using age, gender, race, income levels, tap water intake, bottle water intake, and interaction term between race and income levels. Refer to **Appendix 2-1** for a detailed model output.

The model summary output shows that tap water intake has a significant positive association ($p < 0.001$), whereas bottled water intake has no significant effect ($p = 0.133$). For the variable of race, compared to the reference level of Mexican Americans, non-Hispanic Asians have 289.5% higher mercury levels, while non-Hispanic Black and other races have 17.1% and 3.9% higher mercury levels, respectively. Non-Hispanic White individuals, however, have mercury levels that are 28.6% lower than the reference group. For demographic factors, both age and income level have a positive association with mercury levels. Holding all other variables constant, for every unit increase in the income-to-poverty ratio, mercury levels increase by 7.6% ($p = 0.020$), which may seem counter-intuitive, as families with higher income levels have higher blood mercury levels. Each additional year of age is associated with a 1.4% increase in mercury levels ($p < 0.001$). For gender, compared to males, females have mercury levels that are 7.9% lower ($p = 0.002$).

Regarding the interaction between race and income level, the effects of income on mercury levels vary across racial groups. For Non-Hispanic White individuals, a higher income-to-poverty ratio is associated with a 10.2% increase in blood mercury levels ($p = 0.004$). Conversely, for Non-Hispanic Asians, a higher income-to-poverty ratio is associated with a 10.5% decrease in mercury levels ($p = 0.016$). These significant findings suggest that income affects mercury exposure differently across racial groups. However, for Other Hispanic, Non-Hispanic Black, and Other/Multi-Racial groups, the interaction effects between income and mercury levels are not statistically significant, with p-values of 0.439, 0.307, and 0.145, respectively.

The model is statistically significant (F-statistic = 81.04, $p < 0.001$) and explains 13.34% of the variance in blood mercury levels with an adjusted R-squared of 0.1334. The model's low adjusted R-squared value is probably because of two possible reasons: (1) There are outstanding outliers of blood mercury levels, but our model's predictor variables on demographics and water drinking types are unlikely to be the key reasons that cause these extremely high mercury levels. The key reason could be environmental factors, such as serious chemical pollution. (2) For the mercury exam results, there is lower detection limit. When the actual mercury level is less than this detection limit of 0.21 $\mu\text{g/L}$, the value would be recorded as 0.12 $\mu\text{g/L}$. This creates a large cluster of identical values and true values below detection limit are unknown, which makes it harder for the model to explain variation that's artificially removed.

We also evaluated the residual plots of the final model (Appendix 2-E). For the residual plot, there are straight lines in the middle, which is caused by the detection limit of mercury level. For the Q-Q plot, the points follow the diagonal line closely, suggesting good normality in the central part of the distribution. However, there are some deviations in both tails, which are likely due to the outliers in the mercury level. VIF analysis (Appendix 2-F) is done and there is no multicollinearity issue detected in this model.

Conclusion

This report sheds light on some of the key factors influencing blood pressure and mercury levels in the U.S. population. While we didn't find a significant link between dietary sodium intake and blood pressure, age stood out as a major predictor of hypertension. Additionally, females were found to be more likely to develop Stage 1 Hypertension, suggesting that efforts to prevent high blood pressure should focus more on age-related and gender-specific risk factors. When it comes to mercury exposure, tap water consumption, race, and income emerged as important contributors. Non-Hispanic Asian populations showed significantly higher mercury levels, which raises questions about whether dietary factors cause this or if there is influence from other relationships. To test this we created a model to determine the strength of factors on mercury levels in individuals. The complex relationship between race and income also suggests that environmental exposure risks are not evenly distributed, pointing to areas where public health policies could make a big difference. Overall, these findings suggest a more tailored approach to public health—one that considers individual and community-level differences in exposure and risk. By combining insights from diet, environment, and demographics, we can create more effective strategies to reduce hypertension and mercury toxicity in vulnerable populations.

This study does have some limitations, like the possibility of confounding variables not included in the model (e.g., physical activity, environmental factors, genetic predispositions) could influence blood pressure and blood mercury levels, potentially biasing the results. Moreover, the reliance on observational data also makes it harder to draw clear cause-and-effect conclusions. Future research could consider finding datasets on other relevant variables, such as physical activity, environmental factors, and genetic markers, to capture a more comprehensive view of the determinants of blood pressure and mercury levels. Additionally, identifying thresholds or cutoffs (e.g., alcohol intake levels) that have meaningful clinical implications for blood pressure management would also help increase practical significance of this statistical analysis. Future research could benefit from using long-term data and more granular studies to better understand how individual and environmental factors drive disparities in mercury exposure and hypertension.

References

Centers for Disease Control and Prevention. (2023). High blood pressure facts and statistics. <https://www.cdc.gov/highbloodpressure/data-research/facts-stats/index.html>

Houston, Michael C. “The Role of Mercury in Cardiovascular Disease.” *Journal of Clinical Hypertension*, vol. 13, no. 8, 2011, pp. 621–627, <https://doi.org/10.1111/j.1751-7176.2011.00489.x>.

World Health Organization. “Hypertension.” WHO, 2023, <https://www.who.int/news-room/fact-sheets/detail/hypertension>.

World Health Organization. (2021). Sodium intake and cardiovascular health. <https://www.who.int/news-room/fact-sheets/detail/sodium-intake-for-adults-and-children>

Agency for Toxic Substances and Disease Registry. (2023). Mercury and health. <https://www.atsdr.cdc.gov/toxprofiles/tp46.pdf>

National Center for Biotechnology Information. (2021). The clinical consequences of mercury exposure. PubMed Central. Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC8108748/>

Appendix

Research Question 1

Appendix 1-A: Multinomial Model Coefficients and P-values without Interaction Terms

Table 5: Model Coefficients without Interaction Terms

	(Intercept)	Sodium_intake	Age	Gender	RaceOther Hispanic	RaceNon-Hispanic White	RaceNon-Hispanic Black	RaceNon-Hispanic Asian	RaceMultiracial	Alcohol_consumption
Normal	1.983	0	-0.037	0.815	0.096	0.056	-0.116	-0.026	-0.010	-0.005
Not Classified	-3.029	0	0.013	0.130	-0.168	-0.453	0.092	-0.534	-0.341	0.006
Stage 1 Hypertension	-0.560	0	0.006	0.415	0.061	0.113	0.384	0.120	0.119	0.002
Stage 2 Hypertension	-2.638	0	0.035	0.488	-0.212	-0.389	0.303	-0.367	-0.210	0.004

Table 6: P-Values for Model Coefficients without Interaction Terms

	(Intercept)	Sodium_intake	Age	Gender	RaceOther Hispanic	RaceNon-Hispanic White	RaceNon-Hispanic Black	RaceNon-Hispanic Asian	RaceMultiracial	Alcohol_consumption
Normal	0	0.262	0	0	0	0.042	0	0	0.01	0.011
Not Classified	0	0.665	0	0	0	0.000	0	0	0.00	0.088
Stage 1 Hypertension	0	0.173	0	0	0	0.000	0	0	0.00	0.354
Stage 2 Hypertension	0	0.174	0	0	0	0.000	0	0	0.00	0.088

Appendix 1-B: Multinomial Model Coefficients and P-values with Interaction Terms

Table 7: Model Coefficients with Interaction Terms

	(Intercept)	Sodium_intake	Gender	RaceOther Hispanic	RaceNon-Hispanic White	RaceNon-Hispanic Black	RaceNon-Hispanic Asian	RaceMultiracial	Age	Alcohol_consumption	Sodium_intake:Gender	Sodium_intake:RaceOther Hispanic	Sodium_intake:RaceNon-Hispanic White	Sodium_intake:RaceNon-Hispanic Black	Sodium_intake:RaceNon-Hispanic Asian	Sodium_intake:RaceMultiracial
Normal	1.863	0	0.739	0.561	-0.308	-0.017	-0.241	0.397	-0.037	-0.005	0	0	0	0	0	0
Not Classified	-3.181	0	0.152	-0.248	-0.367	0.347	-0.525	0.225	0.015	0.006	0	0	0	0	0	0
Stage 1 Hypertension	-0.766	0	0.468	0.065	0.289	0.516	-0.598	0.656	0.006	0.002	0	0	0	0	0	0
Stage 2 Hypertension	-2.809	0	0.350	0.677	-0.009	0.716	-1.163	0.295	0.035	0.004	0	0	0	0	0	0

Table 8: P-Values for Model Coefficients with Interaction Terms

	(Intercept)	Sodium_intake	Gender	RaceOther Hispanic	RaceNon-Hispanic White	RaceNon-Hispanic Black	RaceNon-Hispanic Asian	RaceMultiracial	Age	Alcohol_consumption	Sodium_intake:Gender	Sodium_intake:RaceOther Hispanic	Sodium_intake:RaceNon-Hispanic White	Sodium_intake:RaceNon-Hispanic Black	Sodium_intake:RaceNon-Hispanic Asian	Sodium_intake:RaceMultiracial
Normal	0	0.715	0	0	0	0	0	0	0	0.011	0.394	0.004	0.059	0.538	0.394	0.022
Not Classified	0	0.604	0	0	0	0	0	0	0	0.089	0.923	0.868	0.802	0.541	0.984	0.284
Stage 1 Hypertension	0	0.089	0	0	0	0	0	0	0	0.343	0.467	0.983	0.245	0.002	0.002	0.008
Stage 2 Hypertension	0	0.793	0	0	0	0	0	0	0	0.079	0.140	0.000	0.038	0.046	0.007	0.043

Table 9: Confusion Matrix with Interaction Terms

	Elevated	Normal	Not Classified	Stage 1 Hypertension	Stage 2 Hypertension
Elevated	1	1	0	0	0
Normal	633	3686	56	1191	355
Not Classified	0	0	0	0	0
Stage 1 Hypertension	315	528	35	537	380
Stage 2 Hypertension	10	17	2	23	27

Appendix 1-C: VIF for the Multinomial Regression Model without Interaction Terms

Table 10: Variance Inflation Factor (VIF)

	Variable	VIF
Sodium_intake	Sodium_intake	1.070071
Age	Age	1.070983
Gender	Gender	1.060601
RaceOther Hispanic	RaceOther Hispanic	2.041447
RaceNon-Hispanic White	RaceNon-Hispanic White	3.458221
RaceNon-Hispanic Black	RaceNon-Hispanic Black	2.187297
RaceNon-Hispanic Asian	RaceNon-Hispanic Asian	1.586143
RaceMultiracial	RaceMultiracial	1.689048
Alcohol_consumption	Alcohol_consumption	1.029229

Appendix 1-D: VIF for the Multinomial Regression Model with Interaction Terms

Table 11: Variance Inflation Factor (VIF) with Interaction Terms

	Variable	VIF
Sodium_intake	Sodium_intake	23.597787
Gender	Gender	6.317246
RaceOther Hispanic	RaceOther Hispanic	12.566711
RaceNon-Hispanic White	RaceNon-Hispanic White	22.067362
RaceNon-Hispanic Black	RaceNon-Hispanic Black	12.920453
RaceNon-Hispanic Asian	RaceNon-Hispanic Asian	10.592776
RaceMultiracial	RaceMultiracial	10.539913
Age	Age	1.073655
Alcohol_consumption	Alcohol_consumption	1.030666
Sodium_intake:Gender	Sodium_intake:Gender	12.183091
Sodium_intake:RaceOther Hispanic	Sodium_intake:RaceOther Hispanic	12.348642
Sodium_intake:RaceNon-Hispanic White	Sodium_intake:RaceNon-Hispanic White	27.698201
Sodium_intake:RaceNon-Hispanic Black	Sodium_intake:RaceNon-Hispanic Black	13.629163
Sodium_intake:RaceNon-Hispanic Asian	Sodium_intake:RaceNon-Hispanic Asian	11.599994
Sodium_intake:RaceMultiracial	Sodium_intake:RaceMultiracial	11.098754

Research Question 2

Appendix 2-1: Multiple Linear Regression Model Coefficients

To make the coefficients easier to interpret, we performed a back-transformation by exponentiating the regression coefficients and their confidence intervals to interpret the results on the original scale of mercury levels rather than their logged value. This allows us to describe how each predictor multiplicatively affects mercury levels, providing more meaningful and practical insights.

Table 12: Blood Mercury Levels MLR Model Summary Table (Back-Transformed Coefficients)

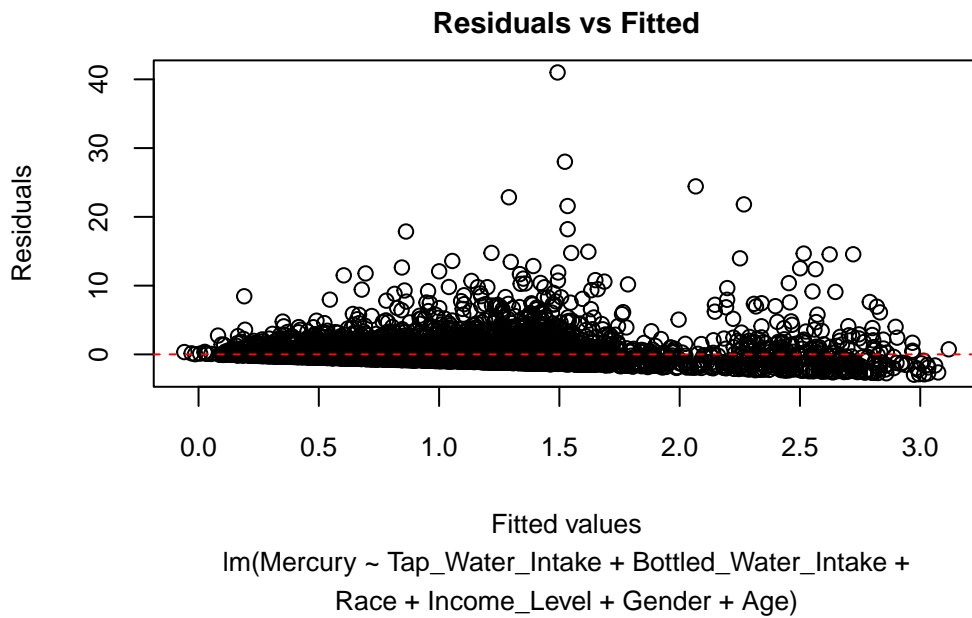
Predictor Variables	Exp(Beta)	SE	95% CI (Exp)	P-value
Intercept	0.182	0.085	0.154, 0.215	< 0.001
Tap Water Intake	1.000	0.000	1.000, 1.000	< 0.001
Bottled Water Intake	1.000	0.000	1.000, 1.000	0.133
Race: Other Hispanic	1.361	0.104	1.110, 1.669	0.003
Race: Non-Hispanic White	0.714	0.090	0.599, 0.852	< 0.001
Race: Non-Hispanic Black	1.171	0.099	0.963, 1.422	0.113
Race: Non-Hispanic Asian	3.895	0.147	2.919, 5.198	< 0.001
Race: Other/Multi-Racial	1.039	0.118	0.823, 1.310	0.749
Income-to-Poverty Ratio	1.076	0.032	1.011, 1.145	0.020
Gender: Female	0.921	0.027	0.873, 0.971	0.002
Age (years)	1.014	0.001	1.013, 1.015	< 0.001
Other Hispanic \times Income	1.032	0.041	0.952, 1.119	0.439
Non-Hispanic White \times Income	1.102	0.034	1.032, 1.177	0.004
Non-Hispanic Black \times Income	1.042	0.040	0.963, 1.127	0.307
Non-Hispanic Asian \times Income	0.895	0.046	0.819, 0.980	0.016
Other/Multi-Racial \times Income	1.067	0.045	0.978, 1.165	0.145

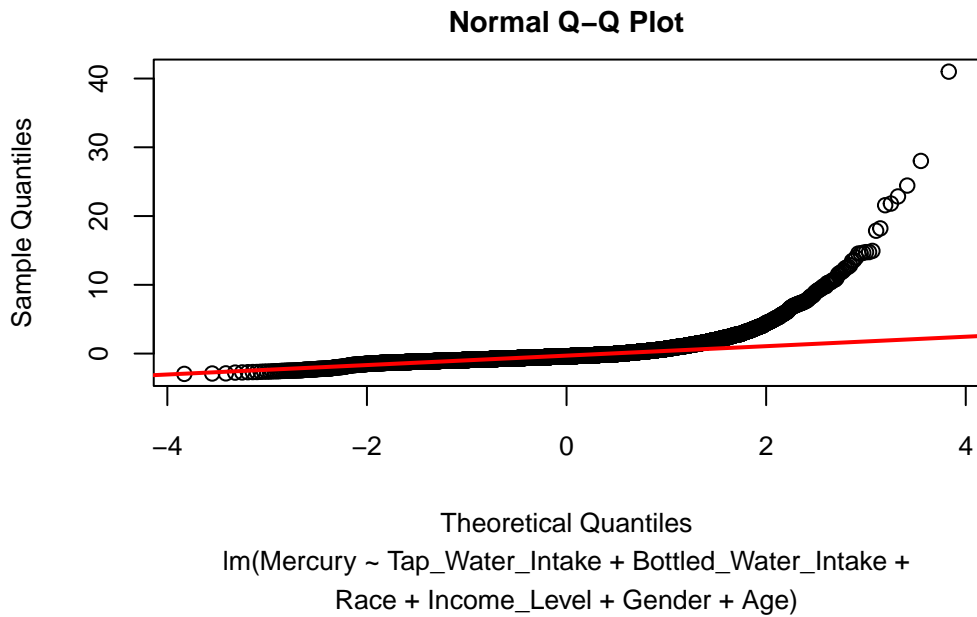
Appendix 2-A: Summary table for categorical variables

Appendix 2-B: Residual plots for original model without log transformation and without interaction term

Table 13: Summary Table for Categorical Variables

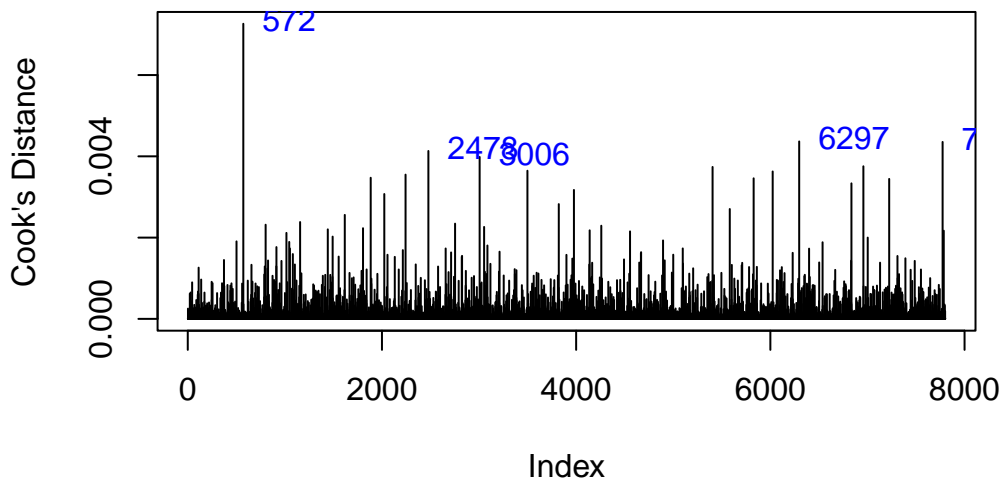
	Category	Count	Percentage
Mexican American	Mexican American	674	8.6
Other Hispanic	Other Hispanic	870	11.2
Non-Hispanic White	Non-Hispanic White	4256	54.6
Non-Hispanic Black	Non-Hispanic Black	1005	12.9
Non-Hispanic Asian	Non-Hispanic Asian	453	5.8
Other Race - Including Multi-Racial	Other Race	543	7.0
Male	Male	3595	46.1
Female	Female	4206	53.9
Tap Water	Tap Water	2778	35.6
Bottled Water	Bottled Water	1969	25.2
Both	Both	498	6.4
None	None	2556	32.8





Appendix 2-C: Cook's distance analysis for top 5 influential points

Influential Points Measured by Cook's Distance



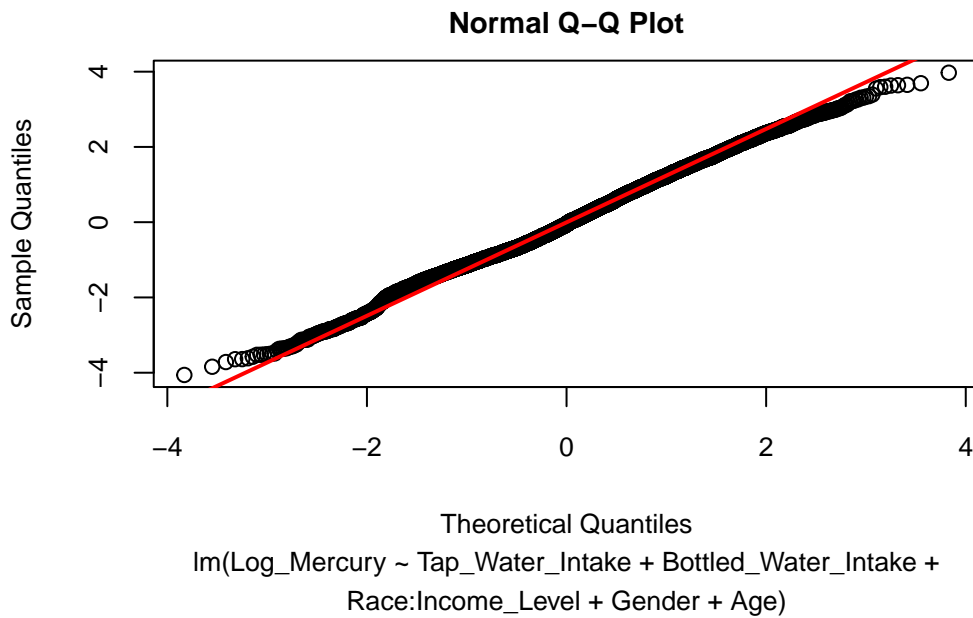
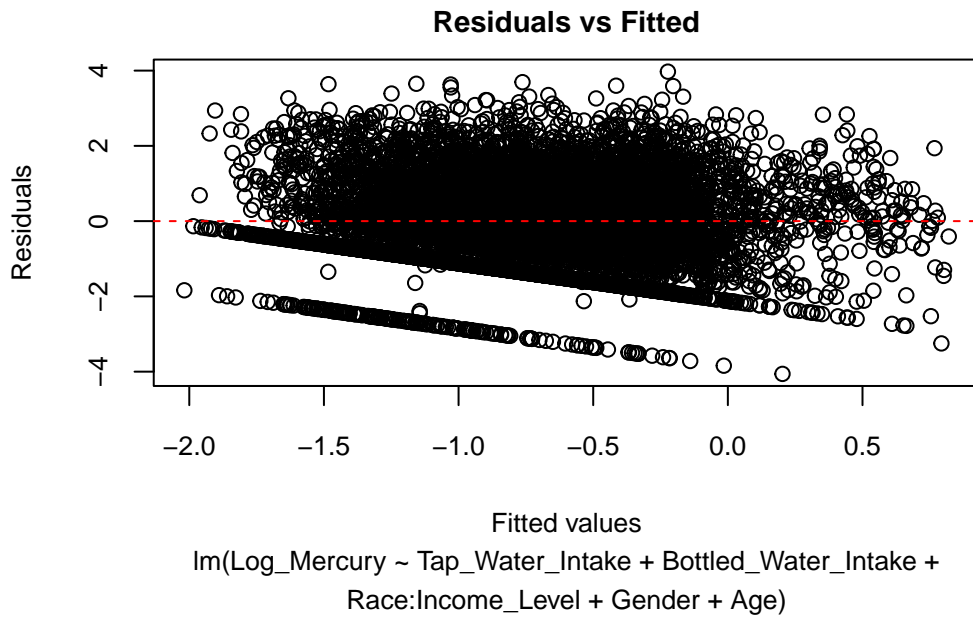
Appendix 2-D: MLR model results and residual plots after removing top 5 influential points examined through Cook's distance

Table 14: Blood Mercury Levels MLR Summary Table After Removing Influential Points

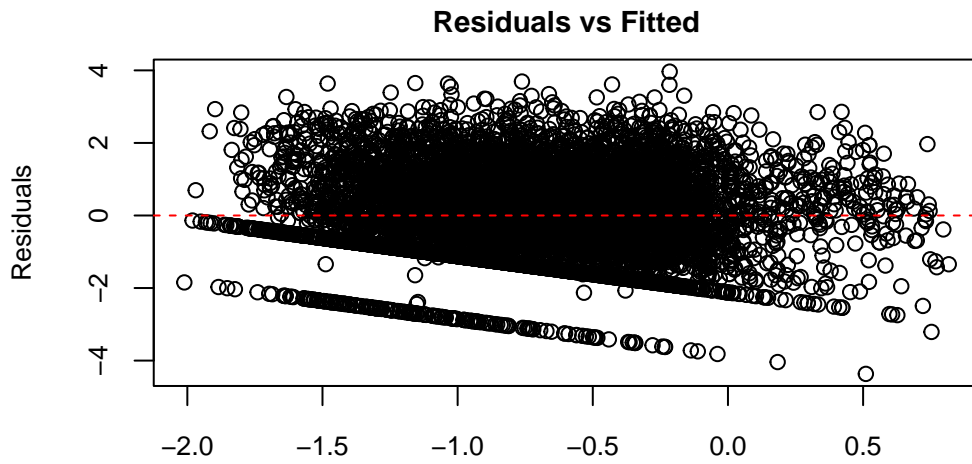
Predictor Variables	Exp(Beta)	SE	95% CI (Exp)	P-value
Intercept	0.180	0.085	0.153, 0.213	< 0.001
Tap Water Intake	1.000	0.000	1.000, 1.000	< 0.001
Bottled Water Intake	1.000	0.000	1.000, 1.000	0.095
Race: Other Hispanic	1.365	0.104	1.113, 1.673	0.003
Race: Non-Hispanic White	0.715	0.090	0.600, 0.853	< 0.001
Race: Non-Hispanic Black	1.174	0.099	0.967, 1.426	0.106
Race: Non-Hispanic Asian	3.988	0.147	2.988, 5.324	< 0.001
Race: Other/Multi-Racial	1.032	0.118	0.818, 1.301	0.793
Income-to-Poverty Ratio	1.078	0.032	1.013, 1.147	0.017
Gender: Female	0.919	0.027	0.872, 0.969	0.002
Age (years)	1.014	0.001	1.013, 1.015	< 0.001
Other Hispanic \times Income	1.030	0.041	0.950, 1.116	0.473
Non-Hispanic White \times Income	1.099	0.034	1.029, 1.174	0.005
Non-Hispanic Black \times Income	1.040	0.040	0.961, 1.124	0.332
Non-Hispanic Asian \times Income	0.892	0.046	0.816, 0.976	0.013
Other/Multi-Racial \times Income	1.072	0.045	0.982, 1.170	0.122

Table 15: Refitted Model Summary Metrics

Metric	Value
R-squared	0.1369
Adjusted R-squared	0.1353
F-statistic	82.29
P-value	< 1e-04

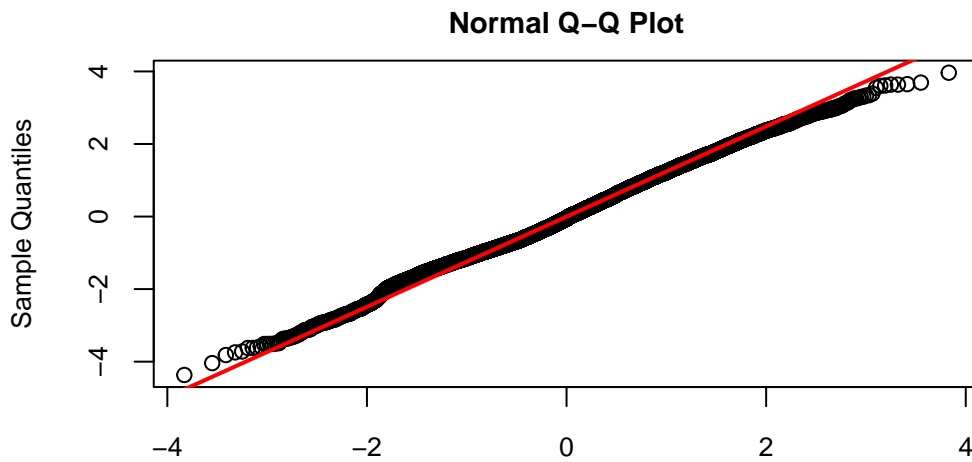


Appendix 2-E: Residual plots for finalized model



Fitted values

lm(Log_Mercury ~ Tap_Water_Intake + Bottled_Water_Intake +
Race:Income_Level + Gender + Age)



Theoretical Quantiles

lm(Log_Mercury ~ Tap_Water_Intake + Bottled_Water_Intake +
Race:Income_Level + Gender + Age)

Appendix 2-F: VIF analysis of final MLR model that includes interaction term between race and income level

Table 16: Variance Inflation Factor (VIF) for Predictors

	Predictor	VIF.GVIF	VIF.Df	VIF.GVIF..1..2.Df..
Tap_Water_Intake	Tap_Water_Intake	1.149	1	1.072
Bottled_Water_Intake	Bottled_Water_Intake	1.152	1	1.073
Race	Race	462.503	5	1.847
Income_Level	Income_Level	15.517	1	3.939
Gender	Gender	1.007	1	1.003
Age	Age	1.085	1	1.042
Race:Income_Level	Race:Income_Level	2477.870	5	2.185