No More Needles:
Characterizing the Relationship Between LDL Cholesterol Levels and Urinary Heavy Metal Concentrations

Thomas Jefferson High School for Science and Technology
Research Statistics 3, Period 5
Dr.Scott
June 8, 2022

Rationale

In the US, cardiovascular disease continues to be the leading cause of death for men, women, and people of most racial and ethnic groups. In particular, coronary artery disease (CAD), or the inability of the major blood vessels of the heart to supply enough blood, oxygen and nutrients to the heart due to low-density lipoprotein (LDL) cholesterol deposits, claimed the lives of 360,900 Americans in 2019, and 16% of the world's deaths (Heart Disease Facts).

Currently, the only method used to assess LDL levels in adults is a lipid panel blood test, which is time consuming, produces significant medical waste requiring special disposal (hazardous sharps), and can carry some risks including bleeding, infection, and bruising (Lipid Panel, 2020). Furthermore, while blood tests are feasible in limited numbers, the prevalence of CAD among Americans suggests that lipid tests should be conducted more often. Thus, alternative methods of assessing LDL levels must be considered.

One promising method of assessing LDL levels in adults is using nontraditional urinary biomarkers, particularly urinary metal concentration. Toxic heavy metals can be absorbed through the skin, breathed in from vehicle or cigarette smoke, or eaten. Urinary heavy metal tests check for high levels of these substances throughout the body (Cleveland Clinic, 2020).

A number of research studies have suggested that urinary metal content may have an impact on cardiovascular health. In a retrograde observational study analyzing 790 participants, a group found that heavy metals such as urinary cadmium, tungsten, and antimony (Sb) were associated with a greater risk of peripheral arterial disease (PAD), marked by occluded blood flow in the muscular arteries of lower extremities. The group found that subjects with PAD had 36% higher levels of cadmium in their urine and 49% higher levels of tungsten compared with non cases, and also found that PAD risk increased

sharply at lower levels of antimony and remained elevated beyond levels as little as 0.1 μg/L (Navas-Acien et al., 2005).

In another retrograde observational study surveying up to 32,012 participants, a research group used data from the 1999-2016 National Health and Examination Survey, or NHANES, and concluded that urinary antimony and dyslipidemia risk had a linear, positive relationship independent with obesity. The group also found that high urinary uranium was associated with a 30% higher odds for type 2 diabetes, which generally corresponds to higher LDL cholesterol levels (Swayze et al., 2021).

However, these trends are not limited to the US alone. Using Korean data, another group found that exposure to (and subsequent intake of) heavy metals like lead, cadmium, and mercury, can affect lipid metabolism. It is important to mention that the group considered serum lead whereas they used urinary mercury and cadmium levels as urinary lead is widely considered to present inaccurate measurements. After adjusting for demographic and socioeconomic factors, the group concluded that increases in urinary mercury levels were associated with an increase in serum lipid levels with a p-value less than 0.05; however, the group found no significant association between the cadmium levels and dyslipidemia in Korean adults (Kim et al., 2022).

Similarly, a research group surveying 1440 adults from Spain found that urine barium (Ba), cadmium (Cd), chromium (Cr), molybdenum (Mo), vanadium (V), and zinc (Zn) levels were positively associated with oxidative stress biomarkers like glutathione. The group found that older participants had, on average, lower urinary concentrations of Co, Cr, and V, and higher concentrations of Cu, Ba, and Zn. Following up on this observation, the group created single-metal adjusted models and found that increases in Mo, Ba, Cr, and V had the greatest association with increases in glutathione (Domingo-Relloso et al., 2019). This is particularly troubling as previous in vitro studies have found that oxidative stress can lead to cholesterol

accumulation in the smooth muscle cells of the vasculature, or blood vessels, by impeding their cholesterol metabolism pathways and creating cholesterol-filled "foam cells" (Gesquière et al., 1999).

As evident from the research, only a few urinary heavy metals, namely Mo, Ba, Cd, Cr, V, Sb, and Zn have been studied; however, the heavy metal urine profile typically also contains data about other metals like Tin and Tungsten, which have been recently implicated in increasing the risk for strokes (Tyrrell et al., 2013).

Thus, in this study, we aim to analyze the relationship between concentrations of previously studied urinary heavy metals of Barium, Cadmium, Cesium, Molybdenum, Magnesium, and Antimony as well as other commonly found heavy metals like Tin, Thallium, Tungsten, and Uranium with low-density lipoprotein (LDL) concentrations. Through this study, we aim to identify a robust model that can be generalized to the population of American adults and elderly of both genders, defined as males and females of the ages 20-64 and 65-80.

Methods

Data Source

Data used in this study was obtained from the National Health and Nutrition Examination Survey (NHANES), which is a longitudinal bi-annual survey conducted by the Centers for Disease Control and Prevention (CDC) since 1999.

Sampling

Each NHANES survey includes demographics, nutrition, medication, and laboratory data on 10,000 participants. To ensure that these participants are representative of the civilian, non-institutionalized American population, NHANES uses a multistage probability sampling design to select participants.

First, primary sampling units are selected. These are typically single counties, and are selected such that sampling units with larger populations are more likely to be selected than those with smaller populations. Then, the sampled PSUs are divided up into segments such as a city block. Households within each segment are listed, and a sample is randomly drawn. As such, segments with greater proportions of particular age, ethnic, or income groups are more likely to oversample these groups in that particular segment, but not so overall. Finally, individuals are selected to participate in NHANES from a list of all persons residing in the selected householdsm in regards to their age, sex, and ethnicity. On average, 2 samples are selected per household (NHANES Tutorials - Module 2 - Sample Design, n.d.).

Measures

As large-scale data analysis would require a large amount of computing power and data storage, which is not feasible given our resources; however, due to the volatile nature of NHANES data and inherent sampling variability, randomization is essential. Thus, before

collecting data, we randomly selected an examination cycle between 2000-2001 (when the CDC first started characterizing high LDL as a public health threat) and 2017-2018. More recent years are not included as COVID-19 may have been a confounding factor as asymptomatic individuals may have altered levels of urinary metals due to their temporary inability to maintain homeostasis. Furthermore, certain communities heavily affected by COVID-19 may not have been represented in the survey in an effort to protect the surveyors' health.

Each cycle was assigned a number between 0 and 18 and a random number generator on Google was used to select a particular survey. The number 14 was drawn, indicating the 2013-2014 examination cycle. This data was obtained from the cleaned dataset posted by the CDC on Kaggle (CDC, 2017). Using the data explorer, we downloaded the demographics.csv and labs.csv files.

Following this, we imported the files onto Google Sheets and used the VLOOKUP function with the following sample syntax '=VLOOKUP(A2,Demographics!A2:E10176,5,false)' to combine the age and gender columns from the demographics dataset with the labs dataset with respect to the patient's sequence number. Following this, we deleted all unnecessary columns, keeping only URXUBA (Barium), URXUCD (Cadmium), URXUCO (Cobalt), URXUCS (Cesium), URXUMN (Manganese), URXUMO (Molybdenum), URXUPB (Lead), URXUSB (Antimony), URXUSN (Tin), URXUSR (Strontium), URXUTL (Thallium), URXUTU (Tungsten), and URXUUR (Uranium), along with LDL.

Following this, we used filter views to split the dataset into 4 spreadsheets by age group (20-64 and 65-80) and gender (male and female).

Data Analysis

Descriptive statistics and boxplots were obtained using the single-variable descriptive statistics function on MiniTab 19 for the overall dataset and for each of the four gender-age datasets.

ANOVA analysis was performed using MiniTab 19's built-in one-way anova feature. Multiple regression was performed using MiniTab 19's built-in multiple regression feature. The explanatory and response variables are detailed in the multiple regressions section of the data analysis.

Data Analysis and Results

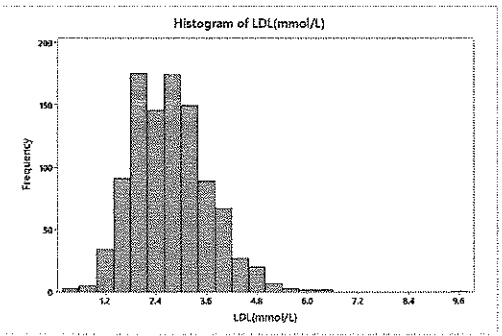
Overall Descriptive Statistics and Histograms:

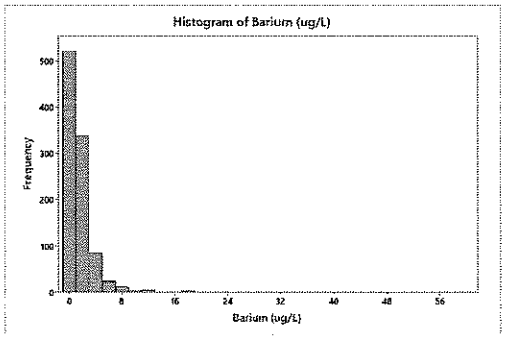
Summary statistics for each urinary metal (ug/L) and LDL. n=994:

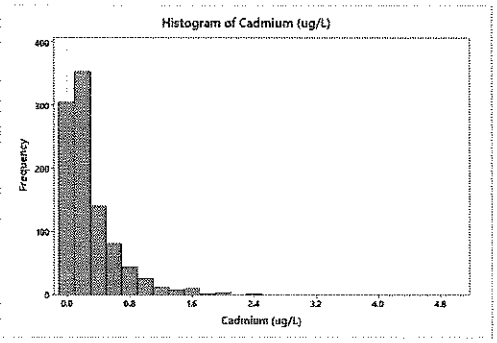
Statistics

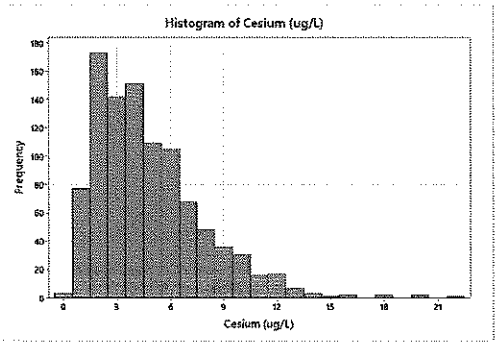
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
LDL(mmol/L)	994	0	2.7605	0.0295	0.9289	0.3620	2.0950	2.6890	3.2905
Barium (ug/L)	994	0	1.6636	0.0888	2.8006	0.0420	0.4600	0.9250	1.9500
Cadmium (ug/L)	994	0	0.3221	0.0139	0.4373	0.0250	0.0810	0.1740	0.3862
Cesium (ug/L)	994	0	4.7998	0.0960	3.0280	0.3870	2.4775	4.1090	6.3515
Molybdenum (ug/L)	994	0	46.13	1.22	38.44	2.12	19.28	35.98	61.52
Magnesium (ug/L)	994	0	0.1432	0.0115	0.3621	0.0920	0.0920	0.0920	0.1300
Antimony (ug/L)	994	0	0.06925	0.00379	0.11964	0.01600	0.02300	0.04300	0.07400
Tin (ug/L)	994	0	1.527	0.192	6.051	0.064	0.200	0.425	0.942
Thallium (ug/L)	994	0	0.19020	0.00446	0.14056	0.01300	0.09300	0.15800	0.25325
Tungsten (ug/L)	994	0	0.09819	0.00497	0.15654	0.01300	0.02200	0.05100	0.10900
Uranium (ug/L)	994	0	0.01176	0.00110	0.03464	0.00140	0.00260	0.00500	0.01043
AGE	994	0	42.567	0.648	20.425	12.000	23.000	41.500	60.000
GENDER	994	0	1.5040	0.0159	0.5002	1.0000	1.0000	2.0000	2.0000

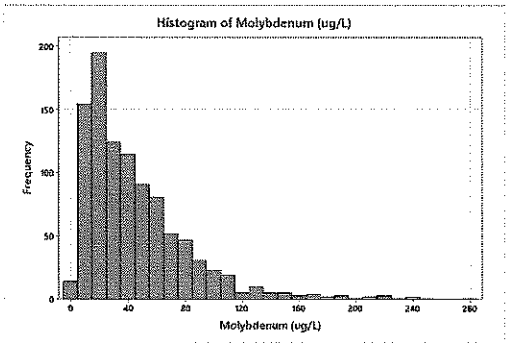
Variable	Maximum
LDL(mmol/L)	9.6980
Barium (ug/L)	60.1100
Cadmium (ug/L)	5.0600
Cesium (ug/L)	21.8000
Molybdenum (ug/L)	281.00
Magnesium (ug/L)	10.4200
Antimony (ug/L)	2.03000
Tin (ug/L)	90.970
Thallium (ug/L)	1.21000
Tungsten (ug/L)	2.47600
Uranium (ug/L)	0.82220
AGE	80.000
GENDER	2.0000

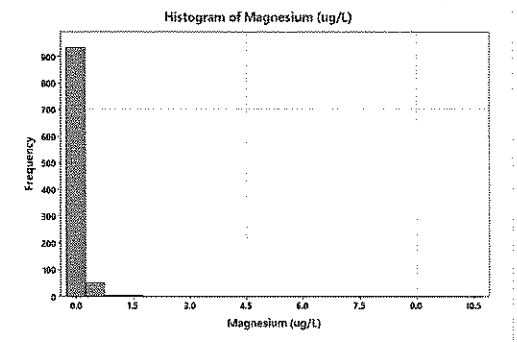


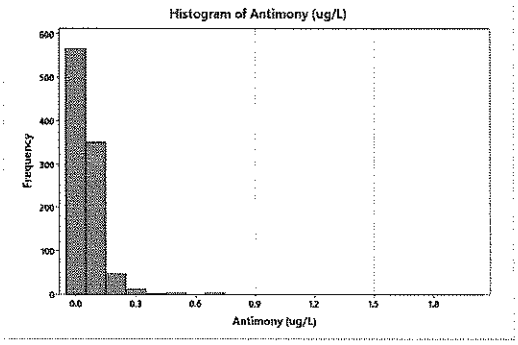


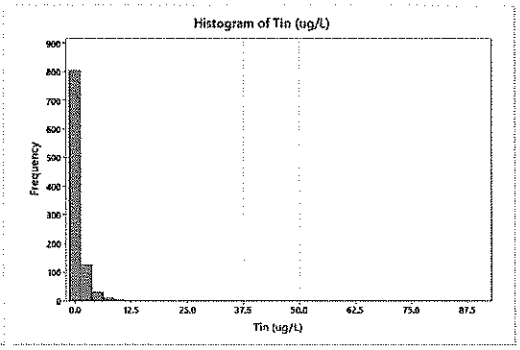

Histogram of Cadmium (ug/L)
Frequency
Cadmium (ug/L)

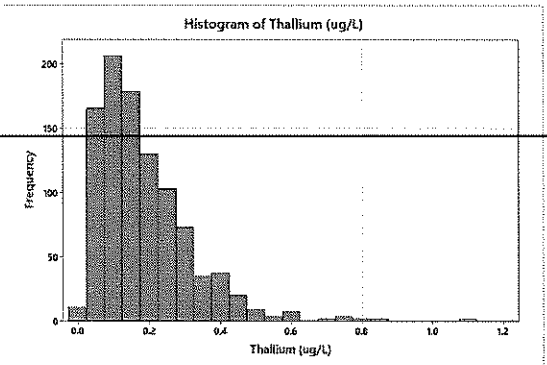

Histogram of Cesium (ug/L)
Frequency
Cesium (ug/L)

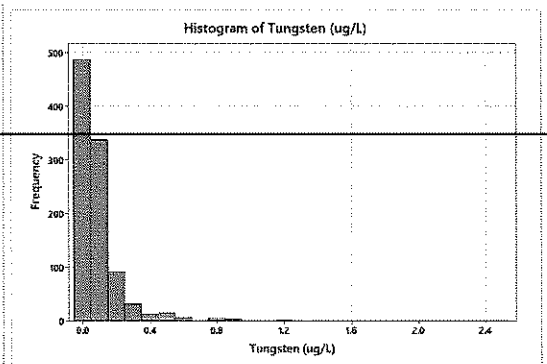

Histogram of Molybdenum (ug/L)
Frequency
Molybdenum (ug/L)

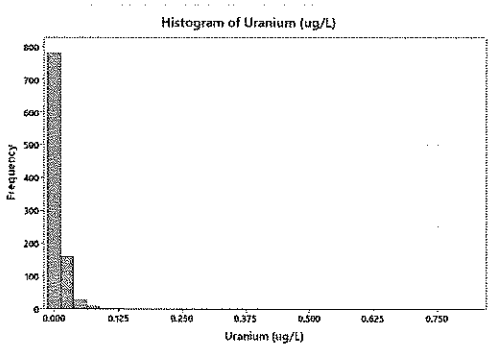

Histogram of Magnesium (ug/L)
Frequency
Magnesium (ug/L)


Histogram of Antimony (ug/L)
Frequency
Antimony (ug/L)


Histogram of Tin (ug/L)
Frequency
Tin (ug/L)


Histogram of Thallium (ug/L)
Frequency
Thallium (ug/L)


Histogram of Tungsten (ug/L)
Frequency
Tungsten (ug/L)



From the data above, it appears that all variables, except for overall LDL concentration in mmol/L, are right-skewed. This means that most data points are on the lower end, although there are a few high data points. Thus, resistant measures such as median and interquartile range must be used to minimize the effect of potential outliers. As all of the units of the metal concentrations are the same, it appears that the median amount of molybdenum in the body is orders of magnitude greater than most of the other metals other than cesium and perhaps barium. Similarly, uranium has the lowest concentration in the body. There appears to be a roughly even distribution of ages, which is essential for further analysis.

*IQR and Outlier Interval (1.5*IQR +/- Median)*

Metal	IQR	Outlier Range
Barium	1.49	(-0.565 , 2.415*)

Cadmium	0.3052	(-0.1312, 0.4792*)
Cesium	3.874	(0.235 ,7.983*)
Molybdenum	42.24	(-6.26 , 78.22*)
Magnesium	0.038	(0.054 , 0.13*)
Antimony	0.051	(-0.008 ,0.094*)
Tin	0.742	(-0.317 ,1.167*)
Thallium	0.16025	(-0.00225,0.31825*)
Tungsten	0.087	(-0.036 , 0.138*)
Uranium	0.00783	(-0.00283,0.01283)

In the above table, it's clear that Barium, Cesium, and Molybdenum have extremely high interquartile ranges relative to the scale of their median, indicating that these explanatory variables are quite variable. Stars indicate the presence of outliers. Here, it's clear that there are no lower outliers, but there are upper outliers which would inflate the mean. This is consistent with the right-skewed distributions for almost every metal. The only metal with no outliers is Uranium.

Descriptive Statistics for Gender/Age Groups:

Summary statistics for each urinary metal (ug/L) and LDL for females aged 21-64. n=313:

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
LDL	313	0	2.9354	0.0532	0.9410	0.9310	2.3270	2.8190	3.3880	9.6980
Barium	313	0	1.3535	0.0987	1.7457	0.0420	0.4350	0.7400	1.6300	17.2000
Cadmium	313	0	0.3779	0.0247	0.4361	0.0250	0.1180	0.2570	0.5030	4.4700
Cesium	313	0	4.742	0.174	3.070	0.405	2.220	4.040	6.520	19.521
Molybdenum	313	0	38.05	1.60	28.28	2.12	17.12	30.80	52.67	173.00
Magnesium	313	0	0.12815	0.00521	0.09223	0.09200	0.09200	0.09200	0.14000	1.25000
Antimony	313	0	0.05976	0.00425	0.07521	0.01600	0.01600	0.03800	0.07000	0.73400
Tin	313	0	1.766	0.433	7.662	0.064	0.200	0.420	0.905	89.880
Thallium	313	0	0.18796	0.00734	0.12982	0.01300	0.08900	0.15900	0.25550	0.84400
Tungsten	313	0	0.07447	0.00641	0.11347	0.01300	0.01300	0.03900	0.08600	0.94000
Uranium	313	0	0.00991	0.00112	0.01985	0.00140	0.00235	0.00450	0.00970	0.24580

Summary statistics for each urinary metal (ug/L) and LDL for females aged 65-80. n=85:

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
LDL	85	0	2.9137	0.0918	0.8465	1.1120	2.2760	2.9740	3.4265	5.3270
Barium	85	0	1.294	0.235	2.163	0.140	0.375	0.820	1.380	18.500
Cadmium	85	0	0.4752	0.0476	0.4390	0.0250	0.1535	0.3730	0.6145	2.4600
Cesium	85	0	4.871	0.364	3.351	0.387	2.625	3.990	6.015	20.300
Molybdenum	85	0	38.34	3.37	31.09	2.87	15.22	32.80	49.40	192.00
Magnesium	85	0	0.260	0.122	1.122	0.092	0.092	0.092	0.150	10.420
Antimony	85	0	0.04845	0.00661	0.06097	0.01600	0.01600	0.03400	0.05800	0.53600
Tin	85	0	1.319	0.155	1.427	0.064	0.370	0.670	1.835	6.660
Thallium	85	0	0.1569	0.0122	0.1125	0.0130	0.0815	0.1340	0.1930	0.7080
Tungsten	85	0	0.06574	0.00810	0.07466	0.01300	0.01300	0.04400	0.08550	0.50700
Uranium	85	0	0.00967	0.00132	0.01215	0.00140	0.00270	0.00530	0.01015	0.06460

Summary statistics for each urinary metal (ug/L) and LDL for males aged 21-64. n=297:

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
LDL	297	0	2.9809	0.0564	0.9727	0.3880	2.2110	2.9480	3.7370	6.3100
Barium	297	0	1.916	0.226	3.888	0.042	0.485	1.100	2.300	60.110
Cadmium	297	0	0.3019	0.0246	0.4241	0.0250	0.0805	0.1580	0.3425	5.0600
Cesium	297	0	4.818	0.176	3.039	0.467	2.470	4.020	6.365	18.100
Molybdenum	297	0	46.37	2.22	38.34	2.47	19.40	35.70	63.33	244.00
Magnesium	297	0	0.1232	0.0117	0.2008	0.0920	0.0920	0.0920	0.0920	3.4600
Antimony	297	0	0.0897	0.0110	0.1904	0.0160	0.0265	0.0460	0.0805	2.0300
Tin	297	0	1.205	0.356	6.141	0.064	0.160	0.310	0.695	90.970
Thallium	297	0	0.19172	0.00769	0.13261	0.01300	0.09650	0.16100	0.26050	0.80900
Tungsten	297	0	0.1016	0.0109	0.1885	0.0130	0.0220	0.0500	0.1070	2.4760
Uranium	297	0	0.01129	0.00177	0.03049	0.00140	0.00240	0.00470	0.01060	0.39230

Summary statistics for each urinary metal (ug/L) and LDL for males aged 65-80. n=92:

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
LDL	92	0	2.4958	0.0819	0.7858	0.3620	2.0040	2.4180	2.9935	4.6030
Barium	92	0	1.310	0.146	1.404	0.060	0.425	0.925	1.580	8.460
Cadmium	92	0	0.5476	0.0696	0.6675	0.0930	0.2150	0.3135	0.6055	4.8890
Cesium	92	0	5.671	0.356	3.413	1.310	3.288	4.930	7.100	21.800
Molybdenum	92	0	47.83	4.29	41.14	5.73	21.98	38.70	58.38	281.00
Magnesium	92	0	0.11224	0.00505	0.04847	0.09200	0.09200	0.09200	0.09200	0.38000
Antimony	92	0	0.05632	0.00470	0.04503	0.01600	0.02525	0.04450	0.07550	0.30400
Tin	92	0	2.661	0.690	6.618	0.064	0.330	0.785	1.995	48.630
Thallium	92	0	0.1942	0.0212	0.2038	0.0130	0.0820	0.1405	0.2117	1.2100
Tungsten	92	0	0.07258	0.00757	0.07260	0.01300	0.03000	0.04900	0.07700	0.34700
Uranium	92	0	0.01385	0.00252	0.02420	0.00140	0.00303	0.00575	0.01350	0.17180

Overall, all four sample sizes are different, with females aged 21-64 having the greatest number of datapoints (313) and females aged 65-80 having the fewest number of datapoints (85). Males aged 21-64 appear to have the highest levels of barium with a median of 2.300 ug/L, antimony at 0.085 ug/L, and Tungsten at 0.0500 ug/L. Similarly, men aged 65-80 appear to have substantially higher levels of tin at a median of 0.785 ug/L.

One-Way ANOVA

Prior to conducting our multiple regression analysis, we conducted two one-way ANOVAs testing whether there is a significant difference in LDL levels between the two gender or two age groups to justify splitting the dataset into the aforementioned four blocks. First, we combined the dataset using the previously described filters method.

Age

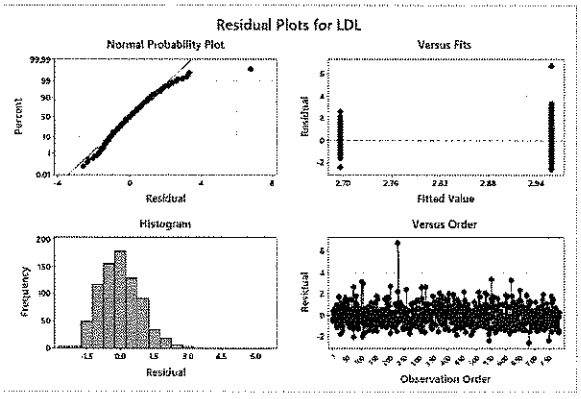
For the one-way anova for age, the hypotheses are H_0: $\mu_y = \mu_o$ and H_A: $\mu_y \neq \mu_o$ where the y stands for young, or the 21-64 age group and the o stands for old, or the 65-80 year old age group.

Then, we assessed the conditions for ANOVA. The graphs of the residuals are given below. The normal probability plot and the histograms show that the data is right-skewed, thereby failing to meet the normal condition of ANOVA. It is reasonable to assume that the two samples are independent of each other, since each sample is represented by one data point, and collecting the urine of one sample does not impact collecting the urine of another.

According to MiniTab, the standard deviation of the 21-64 age group is 0.9560, and that of the 65-80 age group is 0.8397. The ratio of the highest to the lowest standard deviation is 1.138, which is less than 2, meaning that the equal variance condition is met. Thus, since all conditions have not been met, we must proceed with caution.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
AGE	1	9.353	9.3532	10.79	0.001
Error	785	680.725	0.8672		
Total	786	690.079			



Since the p-value of 0.001 is less than the alpha value of 0.05, we reject the null hypothesis. The data suggests that there is a difference in the mean LDL level of the 21-46 age group and the 65-80 age group.

Gender

For the one-way anova for gender, the hypotheses are H_0: $\mu_m = \mu_f$ and H_A: $\mu_m \neq \mu_f$ where the m stands for male and y stands for female.

Then, we assessed the conditions for ANOVA. The graphs of the residuals are given below. The normal probability plot and the histograms show that the data is right-skewed, thereby failing to meet the normal condition of ANOVA. It is reasonable to assume that the two samples are independent of each other, since each sample is represented by one data point, and collecting the urine of one sample does not impact collecting the urine of another.

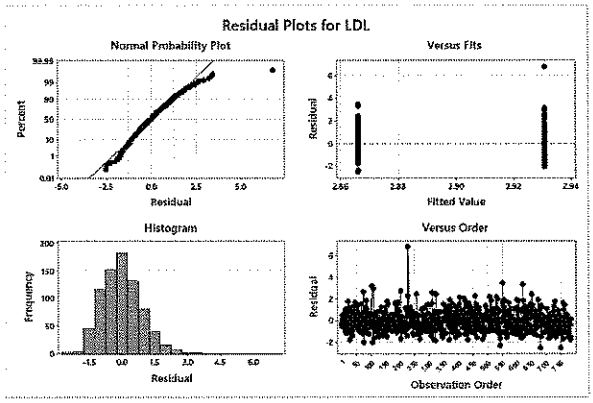
According to MiniTab, the standard deviation of the male group is 0.9535, and that of the female group is 0.9206. The ratio of the highest to the lowest standard deviation is 1.036, which is less than 2, meaning that the equal variance condition is met. Thus, since all conditions have not been met, we must proceed with caution.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Gender	1	0.821	0.8213	0.94	0.334
Error	785	689.257	0.8780		
Total	786	690.079			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.937035	0.12%	0.00%	0.00%



Since the p-value of 0.334 is greater than the alpha value of 0.05, we fail to reject the null hypothesis. The data suggests that there is no difference in the mean LDL level of males and females.

While in the real world we would potentially disregard grouping the data by gender, we will do so in our analysis for the sake of simplicity and working with fewer data points per analysis.

Multiple Regression

Until now, we have considered a number of different metal biomarkers as potential predictors of LDL levels; however, using all of these variables would cause overfitting, which would artificially elevate the R-squared value while introducing redundant information. Due to the number of variables, individually creating added variable plots would be tedious. To avoid this, we have chosen to utilize forward selection with a significance threshold of $\alpha = 0.25$. This is much greater than the typical $\alpha = 0.05$, however, after a preliminary study, it appeared that this significance level eliminated all variables. This is also true for stepwise regression. All following analyses will use R-sq(adj) because of outliers.

Female 21-64

LDL(hat) = 2.902 + 0.0387*Barium* + 0.192*Cadmium* - 0.859*Magnesium* + 0.835*Tungsten* - 4.46*Uranium*

Coefficients

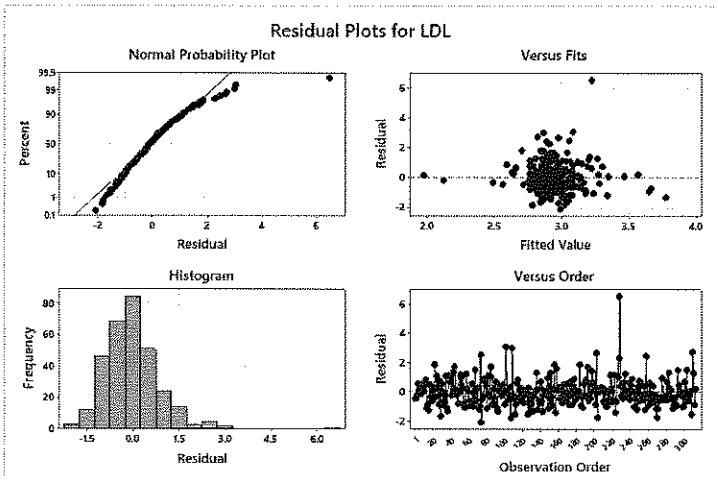
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.902	0.100	29.01	0.000	
Barium	0.0387	0.0317	1.22	0.223	1.09
Cadmium	0.192	0.126	1.53	0.128	1.07
Magnesium	-0.859	0.609	-1.41	0.160	1.13
Tungsten	0.835	0.503	1.66	0.097	1.16
Uranium	-4.46	2.88	-1.55	0.122	1.17

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	8.405	1.6810	1.93	0.090
Barium	1	1.302	1.3022	1.49	0.223
Cadmium	1	2.036	2.0360	2.33	0.128
Magnesium	1	1.734	1.7336	1.99	0.160
Tungsten	1	2.411	2.4108	2.76	0.097
Uranium	1	2.095	2.0953	2.40	0.122
Error	307	267.847	0.8725		
Total	312	276.252			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.934058	3.04%	1.46%	0.00%



From the plots above, it appears that the residuals are right-skewed and clustered, with little variance. This suggests that a transformation — either logarithmic or square root — may be necessary. After the transformation, we found that the logarithmic transformation gave a model with an R-squared (adjusted) value of 0.71%, and the square root transformation gave a model with an R-squared value of 1.34%. This means that the linear relationship of the explanatory variables of the square root multiple regression model explained 1.34% of variability in the LDL concentrations among women aged 21-64.

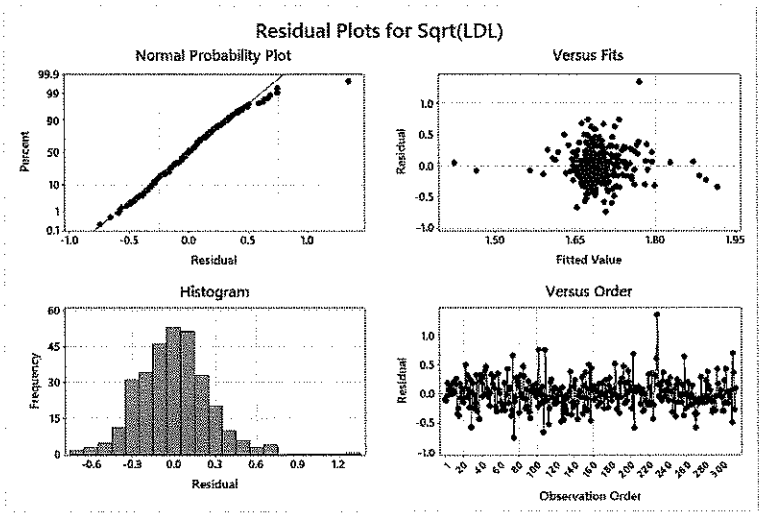
After square root transformation:
Sqrt(LDL)hat = 1.6858 + 0.01118*Barium* + 0.0512*Cadmium* - 0.240*Magnesium* + 0.216*Tungsten* - 1.243*Uranium*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	0.6241	0.12481	1.85	0.103
Barium	1	0.1086	0.10863	1.61	0.206
Cadmium	1	0.1451	0.14509	2.15	0.144
Magnesium	1	0.1356	0.13565	2.01	0.158
Tungsten	1	0.1612	0.16123	2.39	0.123
Uranium	1	0.1629	0.16289	2.41	0.122
Error	307	20.7424	0.06756		
Total	312	21.3665			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.6858	0.0278	60.55	0.000	
Barium	0.01118	0.00882	1.27	0.206	1.09
Cadmium	0.0512	0.0350	1.47	0.144	1.07
Magnesium	-0.240	0.170	-1.42	0.158	1.13
Tungsten	0.216	0.140	1.54	0.123	1.16
Uranium	-1.243	0.800	-1.55	0.122	1.17



In both models, no explanatory variable had a significant slope at the $\alpha = 0.05$ level, indicating that both models should be interpreted with caution; however, following the square root transformation, the residuals appear to be far more normally distributed, even if they still appear to be clustered.

The same process was repeated for the following groups, although only the final transformed multiple regression equation is shown for the sake of conserving space.

Female 65-80

Log(LDL)hat = 0.9434-0.1301*Cadmium* + 0.0248*Cesium* - 0.0650*Magnesium* + 0.749*Antimony*

Model Summary

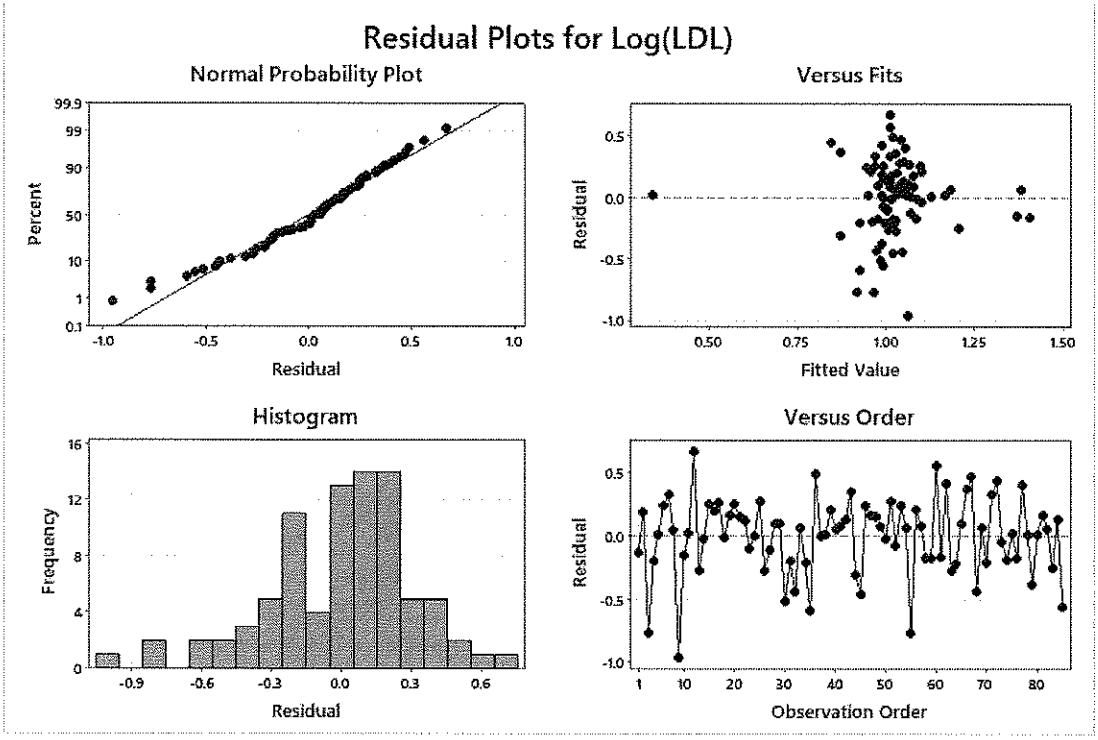
S	R-sq	R-sq(adj)	R-sq(pred)
0.309311	13.09%	8.75%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.9434	0.0653	14.45	0.000	
Cadmium	-0.1301	0.0860	-1.51	0.134	1.25
Cesium	0.0248	0.0113	2.21	0.030	1.25
Magnesium	-0.0650	0.0303	-2.15	0.035	1.01
Antimony	0.749	0.561	1.34	0.186	1.03

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	1.1531	0.28828	3.01	0.023
Cadmium	1	0.2190	0.21902	2.29	0.134
Cesium	1	0.4658	0.46581	4.87	0.030
Magnesium	1	0.4412	0.44122	4.61	0.035
Antimony	1	0.1706	0.17056	1.78	0.186
Error	80	7.6539	0.09567		
Total	84	8.8070			



The R-squared value is slightly greater (8.75%), indicating a model with greater fit, and the residuals appear to be less clustered together. Furthermore, both Cesium and Magensium have p-values (0.030 and 0.035) that are lower than $\alpha = 0.05$. Thus, the data suggests that there is a significant linear relationship between these two variables and LDL levels; however, we must not disregard cadmium and antimony as removing either one makes both variables insignificant again.

Male 21-64

Sqrt(LDL) = 1.6910 + 0.01417*Cesium* - 0.000968*Molybdenum* - 0.2012*Tungsten* + 0.738*Uranium*

Model Summary

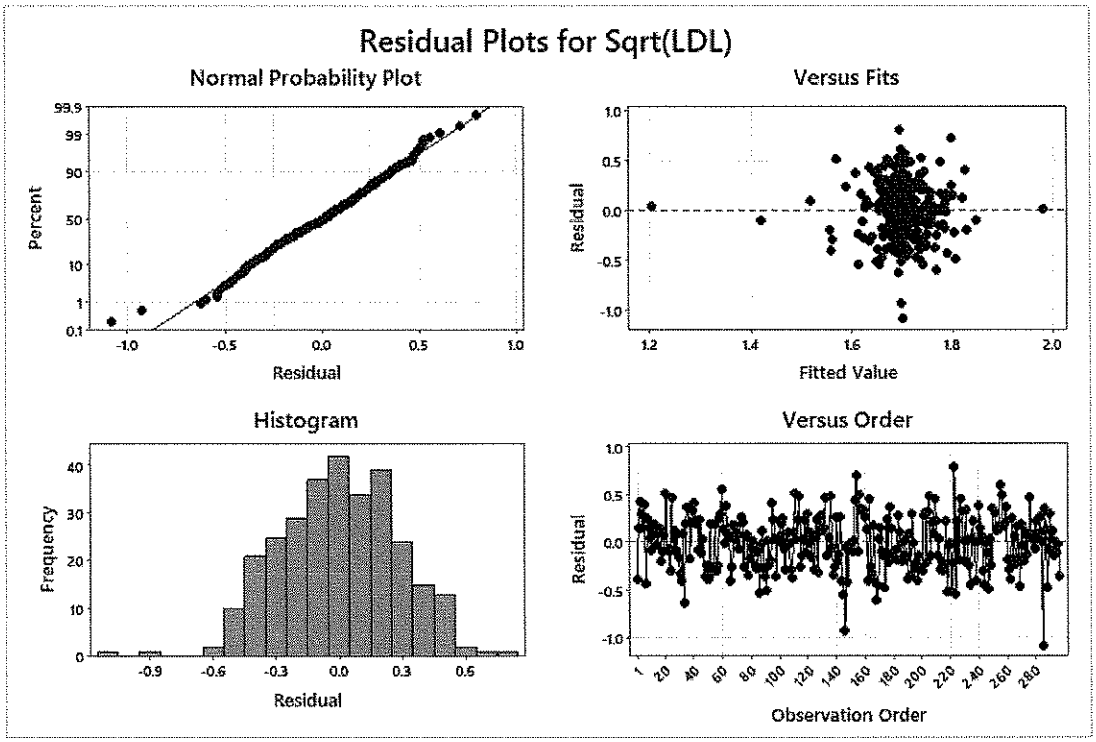
S	R-sq	R-sq(adj)	R-sq(pred)
0.284802	4.01%	2.70%	1.81%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.6910	0.0319	53.08	0.000	
Cesium	0.01417	0.00690	2.05	0.041	1.60
Molybdenum	-0.000968	0.000553	-1.75	0.081	1.64
Tungsten	-0.2012	0.0908	-2.22	0.027	1.07
Uranium	0.738	0.559	1.32	0.187	1.06

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	0.9906	0.24764	3.05	0.017
Cesium	1	0.3424	0.34243	4.22	0.041
Molybdenum	1	0.2489	0.24886	3.07	0.081
Tungsten	1	0.3985	0.39846	4.91	0.027
Uranium	1	0.1417	0.14168	1.75	0.187
Error	292	23.6848	0.08111		
Total	296	24.6753			



Compared to the previous two groups, the residuals of male 21-64 appear to be the most linear and homoscedastic. This is not reflected in the R-squared value, however, which is only 2.70%. In this model, the slopes of Cesium and Tungsten appear to be significant at the $\alpha = 0.05$ level.

Male 65-80

Sqrt(LDL)hat = 1.4903 + 0.0306*Barium* + 0.890*Magnesium* - 1.269*Antimony*

Model Summary

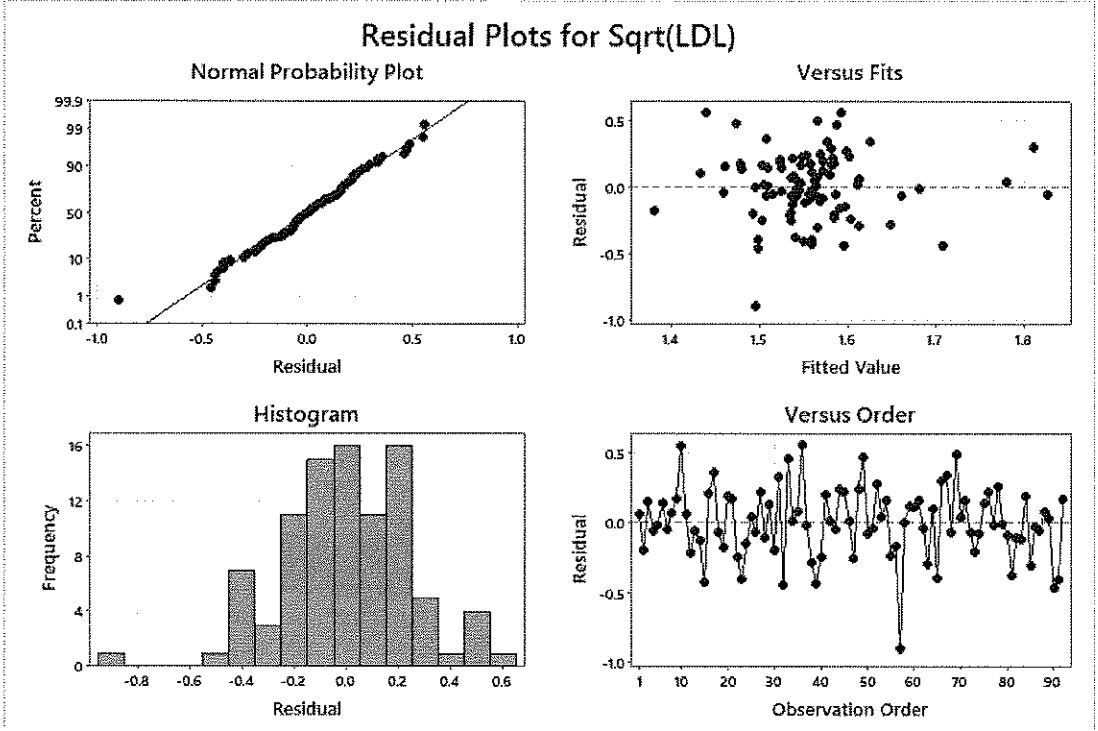
S	R-sq	R-sq(adj)	R-sq(pred)
0.252595	7.00%	3.83%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.4903	0.0690	21.61	0.000	
Barium	0.0306	0.0195	1.57	0.120	1.07
Magnesium	0.890	0.588	1.51	0.134	1.16
Antimony	-1.269	0.643	-1.97	0.051	1.19

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.4224	0.14079	2.21	0.093
Barium	1	0.1576	0.15759	2.47	0.120
Magnesium	1	0.1462	0.14617	2.29	0.134
Antimony	1	0.2487	0.24867	3.90	0.051
Error	88	5.6148	0.06380		
Total	91	6.0372			



Among men aged 65-80, the variables in the model are radically different from their younger counterparts. There are no variables that are significant at an $\alpha = 0.05$ level, although antimony is very close. After the square root transformation, the residuals appear to be the most evenly distributed, although there appears to be one low outlier.

The relationship between the explanatory variables in the model explain 3.83% of the variability in the square root of the LDL levels in men aged 65-80.

Discussion

Due to the sampling method of NHANES, it should be noted that single-year data can be quite volatile and have incredibly large variance between years. This is because NHANES is limited by the number of locations it can survey due to the cost and time involved in moving the mobile examination centers, and participants refusing to give laboratory samples; however, this variance is also a major shortcoming for our study. Since previous studies surveyed data from a range of years or utilized hospital lab-drawn samples rather than mobile-lab samples, minute relationships between LDL and heavy metal concentrations could be more easily distinguished [(Swayze et al., 2021), (Domingo-Relloso et al., 2019)].

Furthermore, medical statistics such as these vary drastically from person to person. For instance, zinc is known to have a strong genetic component involved in its homeostasis (Kambe et al., 2008). Thus, much of the variation in our data, and our model's subsequent inability to explain this variation, may be accounted for by this random variation from observational unit to observational unit.

The metals implicated in this study were Barium, Cadmium, Magnesium, Tungsten, Uranium, Antimony, and Molybdenum. Metals studied but not included in a model were Tin and Thallium. Thus far, no research has been published relating dyslipidemia to tin exposure, although a study has implicated tin fumes with left ventricle diastolic dysfunction and sclerosis of the aortic valve. This suggests that ingested Tin, unlike previously studied tin fumes, may not have major dyslipidemic effects. This is promising because tinfoil is commonplace in our daily lives: it's used as a coating on copper kitchen appliances, as a wrapper, and to create consumer goods (Gunay et al., 2006). The results for Thallium are also positive. No research has been published on Thallium's effects on cholesterol thus far, even though it is commonly used as a radioisotope in cardiovascular stress tests and is actually used to diagnose coronary heart disease. Thallium has several reported side effects including

severe allergic reactions and gastrointestinal disturbances; however, these side effects seem to involve the immune and digestive systems, and not the vasculature or the heart [source].

Interestingly, some of these metals had a negative slope during the multiple regression analysis. This means that the presence of these metals actually decreased the level of LDL in the body. For instance, uranium — the only metal with no outliers — appeared to have a negative slope in the regression line for Females aged 21-64. In Males 21-64, these variables were Molybdenum and Tungsten. In males aged 65-80, Antimony had a negative slope.

These findings are partially consistent with previous studies. For instance, it was found that urinary antimony and tungsten were actually higher for individuals with dyslipidemia and PAD (caused by cholesterol build-up) [(Kim et al., 2022), (Navas-Acien et al., 2005)]; however, research in rats yielded clinically relevant doses of Molybdenum (up to 0.0125 mg/kg) would raise the ceiling of normal heart function and prevent metabolism disorders, which are responsible for cholesterol buildup. Since urinary metal content is reflective of metallic intake, this could mean that greater molybdenum supplements may be necessary for Males 21-64 to lower LDL rates [source]. Finally, uranium is radioactive, and has chemical properties which damage the kidneys. Studies suggesting its potential in lowering LDL levels throughout the body are yet to be published and even if there was a correlation, detecting uranium in urine is not a viable diagnostic method for LDL and is indicative of greater problems (Centers for Disease Control and Prevention, 2017).

In the future, further demographic factors should be considered in the study. For instance, heavy metal exposure, and thereby heavy metal accumulation in the urine, isn't just a marker of the body's inability to maintain metal homeostasis: it's also a marker of the body's increased efforts to flush out metals in people with high environmental exposure. For instance, in the past, only barium, cadmium, and lead (not studied) were found in higher levels in smoker's urine samples compared to non-smokers (Prokopowicz et al., 2020);

however, mining, second-hand smoke, car fumes, and decreased water quality in certain populations. As mentioned earlier, future studies should consider data from a number of years, taking care to account for the fact that some participants may be repeated multiple times. This would exponentially increase the number of samples being used for analysis, which in turn would minimize some of the variance and highlight minor trends. Finally, we suggest pursuing this research as an in-vivo experiment involving mice or other model organisms. Unfortunately, observational data such as this is highly volatile and may have a number of different confounding variables. An experiment can allow researchers to utilize randomization and equal group sizes, which would potentially allow them to use a balanced two-way ANOVA.

But to conclude once again, coupled with the failure of the data in meeting the conditions set forth for conducting an ANOVA and multiple regression means that overall, the presented relationship between LDL and urinary metal content should be taken with a grain of salt.

Reflection

Overall, I'm glad that I was able to complete this project as I did not get the chance to do so during RS2. This project forced me to expand and improve upon my current research knowledge, and tested my perseverance. For instance, I spent over 3 class periods looking for cleaned medical datasets as high-quality medical data is often protected due to HIPAA policies. I used multiple links from the library resources and scoured Google to ultimately locate my dataset on Kaggle, and I was lucky that the CDC had already published the 2013-2014 data. I also learned how to work with large datasets, particularly using G Suite applications. Google Sheets is inherently slower and far more prone to crashing than Excel. However, I learned how to prevent that and how to combine data from two sheets based on

index number without having to manually copy and paste these values. This project helped me become a better researcher overall, and taught me skills essential for me to complete my senior research project next year.

Thank you for the wonderful year, Dr.Scott and I can't wait to see you next year!

Works Cited

Centers for Disease Control and Prevention. (2017a). *National biomonitoring program.* https://www.cdc.gov/biomonitoring/Uranium_BiomonitoringSummary.html

Centers for Disease Control and Prevention. (2017, January 26). *National health and nutrition examination survey.* Kaggle. https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey?select=demographic.csv

Cleveland Clinic. (2020). *Heavy metal test: What it is, purpose, procedure & results.* https://my.clevelandclinic.org/health/diagnostics/22797-heavy-metal-test

Domingo-Relloso, A., Grau-Perez, M., Galan-Chilet, I., Garrido-Martinez, M. J., Tormos, C., Navas-Acien, A., Gomez-Ariza, J. L., Monzo-Beltran, L., Saez-Tormo, G., Garcia-Barrera, T., Dueñas Laita, A., Briongos Figuero, L. S., Martin-Escudero, J. C., Chaves, F. J., Redon, J., & Tellez-Plaza, M. (2019). Urinary metals and metal mixtures and oxidative stress biomarkers in an adult population from Spain: The Hortega Study. *Environment International*, 123, 171–180. https://doi.org/10.1016/j.envint.2018.11.055

Gesquière, L., Loreau, N., Minnich, A., Davignon, J., & Blache, D. (1999). Oxidative stress leads to cholesterol accumulation in vascular smooth muscle cells. *Free Radical Biology and Medicine*, 27(1–2), 134–145. https://doi.org/10.1016/s0891-5849(99)00055-6

Gunay, N., Aksoy, M., Davutoglu, V., Yildirim, C., & Ege, I. (2006). Cardiac Damage Secondary to Occupational Exposure to Tin Vapor. Inhalation Toxicology, 18(1), 53–56. https://doi.org/10.1080/08958370500282928

Heart disease facts | cdc.gov. (2022, February 7). Centers for Disease Control and Prevention. https://www.cdc.gov/heartdisease/facts.htm

Kambe, T., Weaver, B. P., & Andrews, G. K. (2008). The genetics of essential metal homeostasis during development. *Genesis*, 46(4), 214–228. https://doi.org/10.1002/dvg.20382

Kim, D. W., Ock, J., Moon, K. W., & Park, C. H. (2022). Association between heavy metal exposure and dyslipidemia among korean adults: From the korean national environmental health survey, 2015–2017. *International Journal of Environmental Research and Public Health*, 19(6), 3181. https://doi.org/10.3390/ijerph19063181

Lipid panel. (2020, December 4). Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lipid-panel

Navas-Acien, A., Silbergeld, E. K., Sharrett, A. R., Calderon-Aranda, E., Selvin, E., & Guallar, E. (2005). Metals in urine and peripheral arterial disease. *Environmental Health Perspectives*, 113(2), 164–169. https://doi.org/10.1289/ehp.7329

NHANES tutorials - module 2 - sample design. (n.d.). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/nchs/nhanes/tutorials/module2.aspx

Navas-Acien, A., Silbergeld, E. K., Sharrett, A. R., Calderon-Aranda, E., Selvin, E., & Guallar, E. (2005). Metals in Urine and Peripheral Arterial Disease. *Environmental Health Perspectives*, 113(2), 164–169. https://doi.org/10.1289/ehp.7329

Prokopowicz, A., Sobczak, A., Szdzuj, J., Grygoyć, K., & Kośmider, L. (2020). Metal concentration assessment in the urine of cigarette smokers who switched to electronic cigarettes: A pilot study. *International Journal of Environmental Research and Public Health*, 17(6), 1877. https://doi.org/10.3390/ijerph17061877

Sherkov, Z., Kurdanov, Kh., & Sherkova, L. (2012). Effects of molybdenum and hypoxia on the cardiovascular system of experimental animals. Aviakosm Ekolog Med, 46(5), 51-4. https://pubmed.ncbi.nlm.nih.gov/23405421/

Swayze, S., Rotondi, M., & Kuk, J. L. (2021). The associations between blood and urinary

concentrations of metal metabolites, obesity, hypertension, type 2 diabetes, and dyslipidemia among US adults: NHANES 1999–2016. *Journal of Environmental and Public Health*, 2021, 1–13. https://doi.org/10.1155/2021/2358060

Thallium-201. (2022, May 8). Poudyal B., Pallabi S., & Chowdhury, Y. https://www.ncbi.nlm.nih.gov/books/NBK560586/

Tyrrell, J., Galloway, T. S., Abo-Zaid, G., Melzer, D., Depledge, M. H., & Osborne, N. J. (2013). High urinary tungsten concentration is associated with stroke in the national health and nutrition examination survey 1999–2010. *PLoS ONE*, 8(11). https://doi.org/10.1371/journal.pone.0077546