Table 1 Multivariable linear regression model for blood Lead Level among respondents to the National Health and Nutrition Examination Survey (NHANES), 2015-2016, United States (n=601)

Dependent Variable	Coefficient	Standard Error	*p*-value	95% CI
Do you now smoke cigarettes				
No	*(ref)*			
Every day	0. 071	0.102	0.487	[–0.129;0.271]
Some days	0. 057	0.143	0.693	[-0.225;0.338]
Gender				
Male	*(ref)*			
Female	-0. 214	0. 087	0. 014	[–0.384;–0.044]
Age	0. 025	0. 003	<0.001	[–0.019;0.030]
Race				
Mexican American	*(ref)*			
Other Hispanic	0. 160	0. 159	0. 313	[–0.151;–0.472]
Non-Hispanic White	0. 108	0. 136	0. 429	[–0.160;0.375]
Non-Hispanic Black	0. 267	0. 153	0. 080	[–0.032;0.567]
Other/multi-racial	0. 375	0. 178	0. 036	[–0.025;0.726]
Educational level				
Less than 9th grade	*(ref)*			
9th to 11th grade	–0. 194	0. 188	0. 301	[–0.563;0.174]
High school graduate/GED or equivalent	–0. 255	0. 171	0. 135	[–0.591;0.080]
Some college or Associate degree	–0. 342	0. 172	0. 047	[–0.680;0. 003]
College degree or above	–0. 274	0. 183	0. 133	[–0. 634;0. 084]
Marital status				
Never married	*(ref)*			
Married	–0.009	0.131	0.948	[–0.267;0.250]
Widowed	0.119	0.202	0.558	[–0.278;0.516]
Divorced	0.171	0.163	0.296	[–0.149;0.492]
Separated	0.433	0.226	0.056	[–0.011;0.877]
Living with Partner	0.157	0.160	0.327	[–0.157;0.472]
Blood Cadmium	0.322	0.073	<0.001	[–0.384; 0.044]
Intercept	0.212	0.248	0. 394	[–0.275;0.699]

F (degrees of freedom, residual degrees of freedom): 8.69(18, 582)
Probability > F: <0.001
R-squared: 0.212
Adjusted R-squared: 0.188

Results

As a part of my study to find out the correlation between my main outcome `Blood Lead Level` and my main exposure of interest ` Do you smoke cigarettes now `, I performed a multivariable analysis using the outcome as a response variable and exposure and other potential covariates as predictors. After running an analysis, it is clear that there is no statistically significant association between my exposure of interest and outcome of interest (p-value =0.487 for smoking every day and p-value = 0.693 for some days smoking). And we can say that as compared to a person not smoking at all, people who are smoking every day have higher blood lead levels by 0.07ug/dL and people who are smoking for some days have 0.05ug/dL more blood lead levels.

I also found out that gender is significantly associated with the outcome (p-value = 0.014). As compared to males, females have a lower level of blood lead by 0.21ug/dL. While age has been also significantly correlated with Blood lead levels (p-value < 0.001) where, with an increase in every year of age, there is an increase in blood lead levels by 0.02ug/dL.

While Education level wise, as compared to a person having the highest education level of 9^{th} grade, a person having to study for a college degree is highly affected by blood lead level. As compared to 9^{th} grade or lesser studies person, there is a slight decrease in blood lead by 0.19ug/dL for a person studying in 9^{th} to 11^{th} grade, and decrement of 0.25ug/dL for a person studying in high school, decrement of 0.34ug/dL for a person studying in some college or associate degree and decrease of 0.27ug/dL for a person having a graduate degree.

For the racial origin, as compared to Mexican Americans, other Hispanics have an average of 0.16ug/dL higher Blood lead levels, Non-Hispanic Whites have 0.108ug/dL higher Blood lead levels, while Non-Hispanic Blacks have 0.26ug/dL higher and other races have an average of 0.38ug/dL higher blood lead levels. On the other hand, marital status also played a significant role. As compared to an unmarried person, a married has an average of 0.009 lower blood lead levels, while a widowed has an increase in blood lead level by 0.12ug/dL. Divorcees and separated also have a higher level of blood lead levels by 0.17ug/dL and 0.43ug/dL respectively as compared to an unmarried person, while a person living with a partner has 0.15ug/dL higher blood lead levels.

Blood Cadmium levels also being significantly associated with Blood lead levels (p-value < 0.001). And by observing the coefficient, we can say that with a unit increase in Blood Cadmium level, there will be raise of blood lead level by 0.32ug/dL.

Also, Seeing the value of R-square, we can notice that only 21% of the variance has been explained by the model hence the model does not fit the data well. There is no missing value in our model. Also, the y-intercept of 0.212 shows that this value is constant and that is the minimum value if every other dependent variable is zero.

Discussion

Having all the analysis done including univariate, Bivariate and multivariate modelling, I concluded that our main hypothesis of correlation between outcome `Blood lead level` and exposure `Do you smoke cigarettes now` is wrong and that these both are now significantly associated as per our model while according to bivariate analysis, it was highly correlated to the outcome.

But knowing the fact that `Age`, `Gender`, and ` Blood Cadmium` are significantly associated with our outcome, this is of high importance as per the public health perspective and can be used for further prevention of Blood Lead poisoning in the public. These results raise so many significant questions of interest in the field of Public Health to study. Why with the age there is an increase in blood lead levels? What factors are associated with age to result in an increase in blood lead levels? Why females are less affected than males and what might be the possible reason for sex being a possible confounder?

Chemical and environmental associations between blood lead level and cadmium level are also topics of interest to research. And Why Blacks and other races are more prone to increase blood lead levels than Hispanic Americans and Other Hispanic or Whites? Is that related to Socioeconomic status or that other factors are also involved in Racial based differences?

All these questions draw attention to conducting the research in the direction of these questions.