Associations Between Body Mass Index and The Prevalence of Low Micronutrient Intake Among US Adults, 2017-2018

Liang Li[1, †, *], Xuanzhuang Lu[2, †]

[1]Global Health Research Center, Duke Kunshan University, Kunshan, Jiangsu, China

[2]The Third Clinical School, Guangzhou Medical University, Guangzhou, Guangdong, China

[*]Corresponding author: ll341@duke.edu, 2021111355@stu.gzhmu.edu.cn

[†]These authors contributed equally to this work.

Keywords: BMI, micronutrient, NHANES.

Abstract: Obesity as a growing public health threat, is associated with low micronutrient intake. It is important to explore the relationship between the two for updated dietary recommendations. This study aims at exploring the association between BMI and micronutrient intake levels by analyzing a nationally representative sample of US adults aged 20 or elder from the National Health and Nutrition Examination Survey (NHANES) 2015-2018. With linear and nonlinear regression as well as model optimization, results have indicated that a higher BMI is associated with more intake of sodium and magnesium, and conversely with lower intake of total beta-carotene, folate, vitamin C, vitamin D and vitamin E. These findings help to update the current relationship between micronutrient intake and BMI and highlight the specific micronutrient, directly helping to improve existing dietary intake recommendations for U.S. adults.

1. Introduction

Obesity has become a public health threat with acceleratingly increased prevalence [8]. In 2018, 42.5% of US adults are obese and another 31.1% are overweight [6], and by 2030 almost half of US adults are estimated to have obesity [26]. A simple and convenient way to define obesity and overweight raised by the World Health Organization (WHO) and the National Institutes of Health (NIH) is based on the body mass index (BMI), which is calculated by dividing one's weight (kg) by the square of one's height (meters). BMI (kg/m^2) between 25 and 29.9 is considered overweight and BMI ≥ 30 is considered obese [20]. Previous literatures have identified the associations the between the increased BMI and higher risks of all-cause mortality and multiple chronic diseases including CVD, diabetes and some types of cancer [23].

The association between low level of micronutrients and increased BMI has also been revealed [15], which could be resulted from the diet pattern that is rich in energy but poor in nutrient level [11]. Various biological mechanisms behind such association have been proposed including that micronutrient supplements alter energy expenditure and fat oxidation [16], mineral micronutrients regulate appetite [17], and micronutrients alters common obesity by intervening in metabolic syndrome (MetS) [9]. Inspired by the potential benefit of micronutrients, 33.9% US adults reported ever attempting to take micronutrient supplements for weight loss [22], however, the efficacy of micronutrient supplements is of limited evidence [14], which emphasized the importance of updated scientific evidence.

Given the high prevalence of inadequate micronutrient intake and obesity, as well as the wild use of micronutrient supplements, the objective in this study is to determine the specific micronutrients associated with lower BMI with focus on 2017 to 2018 wave of the National Health and Nutrition Examination Survey (NHANES), which is a major nation-wide survey series [10].

2. Methods

2.1 Study population

The present study used the data from the 2017-18 wave (2020 published) and 2015- 16 wave (2018 published) of NHANES which is a program of studies designed to assess the health and nutritional status in the United States, with both interviews and physical examinations. NHANES is a major program of the National Center for Health Statistics (NCHS), belonged to the Centers for Disease Control and Prevention (CDC). The NHANES program began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics. In 1999, the survey became a continuous program that has a changing focus on a variety of health and nutrition measurements to meet emerging needs. The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country, 15 of which are visited each year. Details of this survey have been published elsewhere.

For these two sets of data, samples which are reported both two days in each set of data and which are estimated reliable and meet the minimum criteria have been selected. When compared food consumed yesterday to usual, only the usual samples which may represent for usual diet were chosen. In addition, the samples which are on special diet were abandoned because it would be ambiguous when they began their special diet, which may have great impact on their BMI. Besides, after excluding participants due to missing data, the final analytical sample included 4247 participants aged over 20 years old.

2.2 Body Mass Index (BMI)

The BMI data were collected, in the Mobile Examination Center (MEC), by trained health technicians. The health technician was assisted by a recorder during the body measures examination. The BMI is defined by the formula BMI = weight(kg) /height $(m)^2$.

2.3 Assessment of micronutrients intake

NHANES participants were assessed for their micronutrient's intake by self-reported data. The detailed self-reported dietary intake information from NHANES participants was obtained and used to estimate the types and amounts of foods and beverages (including all types of water) consumed during the 24-hour period prior to the interview (midnight to midnight), and to estimate intakes of energy, nutrients, and other food components from those foods and beverages. Here, we select all kinds of micronutrients as our independent variables. It is worth noting that as for folate consumed, we choose the variable of folate as dietary folate equivalents instead of folic acid in total or food folate because dietary folate equivalents is a more scientific index to indicate people's folate intake [25]. Next, we plus Vitamin B12 and Added Vitamin B12 into total Vitamin B12 intake. And we plus Vitamin E as alpha-tocopherol and Added alpha-tocopherol (Vitamin E) into total Vitamin E intake. Finally, we calculate the average of two days intake in order to make our data more representative to one's usual diet which may actually influence BMI.

2.4 Covariates

During each interview the interviewers measured a range of demographic and formative environment covariates ascertained from the survey of NHANES in 2015- 2016 and 2017- 2018. we consider potential confounders, including age, gender, race/ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian and Other Race - Including Multi-Racial) and education level (1-5 level). All NHANES participants are eligible for two 24-hour dietary recall interviews, the respondents were encouraged to answer more questions as soon as possible and these questions were collected by trained research staff members. The interviewers were monitored throughout the data collection period. If they were unable to answer questions, we would leave the option blank. What's more, we also include Caffeine, Theobromine, Alcohol intake and total water consumed yesterday as our covariates (we plus Total plain water drank yesterday (including plain tap water, water from a drinking fountain, water from a water cooler, bottled water, and spring

water), Total tap water drank yesterday (including filtered tap water and water from a drinking fountain) and Total bottled water drank yesterday into total water consumed yesterday).

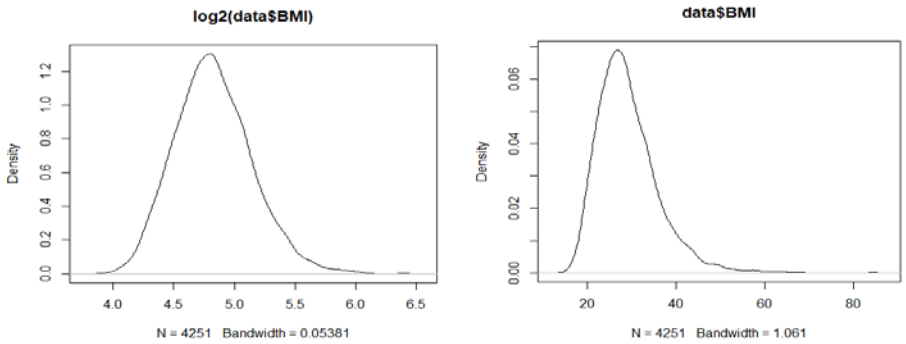
2.5 Statistical analysis

Initially, we conduct exploratory data analysis on our data. As for continuous variables, we draw boxplots to see their distribution while for category variables, we draw histogram (see Figure S1, Figure S2). As we can see, the age of our sample shows perfectly normal distribution while the sample size of male is a little bit larger than that of female, which we think is reasonable. What's more, CAFF, ALCO, THEO, CRYP, ACAR and LZ are of skewness to some degree. Overall, we think our data is fairly uniform because most of variables are almost normally distributed. In addition, we use summary function to calculate summary statistics of our data (see Table S1) and use sd function as well var function to calculate standard deviation and variance of BMI (see Table S2). What's more, we can see our data distribution of categorical variables clearly and value descriptions in Table 1.

Table 1. Data Distribution of Categorical Variables

Item	**Value**	**n = 4247**	
		n	%
Gender (Gen)			
Female	0	n=2078	48.90
Male	1	n=2169	51.10
Race/ethnicity (RETH)			
Mexican American	1	n=604	14.22
Other Hispanic	2	n=442	10.41
Non-Hispanic White	3	n=1702	40.08
Non-Hispanic Black	4	n=800	18.84
Non-Hispanic Asian	6	n=532	12.53
Another Race - Including Multi-Racial	7	n=167	3.93
Education level (Edu)			
Less than 9th grade	1	n=349	8.22
9-11th grade (Includes 12th grade with no diploma)	2	n=433	10.20
High school graduate/GED or equivalent	3	n=978	23.03
Some college or AA degree	4	n=1297	30.54
College graduate or above	5	n=1190	28.02

Then, we make a density plot to check the skewness of BMI, and the result (Figure 1) suggests that the plot is a little right-skewed. So, we perform a log2 transformation to normalize the data.



previous after transformation

Figure 1. log2 Transformation

Next, we conduct a pairs function to see the correlation between our continues variables and transform BMI (y) (see Figure S3). However, it's not proper to use the scatter plot in the bivariate

analysis of a categorical variable and continuous variable. Thus, the box is normally used plot to demonstrate the relationship between categorical variable and continuous variable (see Figure S4).

Then is our regression part. We firstly use linear regression model as our baseline model to fit our data. Because RETH and Edu are multiple class variable, we need to transform them into dummy variable, and r can achieve it when we add those two factors into our model. What's worth mentioning is that r set the first value of those two variables, which are 1(Mexican American) for RETH and 1(Less than 9th grade) for Edu as the reference value automaticly. However, no matter multiple R-squared or adjusted R-squared is quite small, which means the ability for our dependent variables to explain Y is insufficient. Therefore, we conduct non- linear transformations to enlarge our R- squared as well explore the impact for square, cube, fourth power or interaction of X on Y, which is our model 2. But there are so many non- important variables (with p-value > 0.05) in model 2, so we need to conduct variable selection to select important variable. Currently, we have two effective selection method, which is all-subsets selection and stepwise selection. But it's generally acknowledged that stepwise regression may not show us the best regression model while all-subsets selection assesses all possible model so its result is more reliable [18]. So, we conduct only all-subsets selection here to create our model 3. In order to select a maximum likelihood estimation model with simplicity, we use BIC criteria [3] in all of the three-selection method above.

In addition, we also adopt Shrinkage Method to optimize our model. Because r couldn't transfer factors into dummy variable when we conduct ridge regression and lasso regression, we first transfer Edu and RETH into characters and use model. matrix on the whole data, which will not have impact on continuous variables and binary variables but only transfer character ones into dummy variables. By default, the glmnet () function performs ridge regression for an automatically selected range of λ (tuning parameter) values. However, here we have chosen to implement the function over a grid of values ranging from λ = 1000 to λ = 10^ (-2), essentially covering the full range of scenarios from the null model containing only the intercept, to the least squares fit. Note that by default, the glmnet () function standardizes the variables so that they are on the same scale. Here, we don't turn off this default setting. We divide our data into two parts, upper and lower halves of our data as training data and test data and use cross-validation method to choose the best λ. Considering that $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ (yi is y of test dataset), the best λ lead min Mean- Square Error of a model and Mean- Square Error is a standard to measure a model's ability to predict y. The smaller Mean- Square Error is, the stronger prediction ability the model has. Moreover, normally, no coefficients in ridge regression is equal to 0, that is to say, all the variables in our data are including in this model, so their maybe many insignificant variables in this model. That's why we conduct lasso regression (model 5) besides ridge regression (model 4), which can actually select important variables. Moreover, we use $\mathrm{rsq} = 1 - \mathrm{cv.out\$cvm/var(y)}$ to calculate the R- squared of ridge regression and lasso regression.

In addition, we also make a comparison of all the 5 models and finally we consider model 3 and 5 as the best models.

We consider p values of less than 0.05 to be statistically significant. We analyse all data using R , version 4.1.1.

3. Result

3.1 Regression

We first use linear regression model to fit our data. We can view the result in Table 2.

Table 2. the Result of Linear Regression

Residuals:					
Min	1Q	Median	3Q	Max	
-0.86783	-0.20345	-0.01286	0.18672	1.47202	
Coefficients:					
	Estimate	Std. error	t value	Pr (>| t |)	
(Intercept)	4.808e+00	2.854e-02	168.493	< 2e-16	***
CHOL	1.433e-04	7.641e-05	1.875	0.060842	.
TVE	-5.508e-04	8.411e-04	-0.655	0.512623	
RET	2.087e-04	3.108e-03	0.067	0.946474	
VARA	-2.392e-04	3.109e-03	-0.077	0.938683	
ACAR	1.617e-06	1.302e-04	0.012	0.990094	
BCAR	1.899e-05	2.591e-04	0.073	0.941553	
CRYP	-2.193e-05	1.316e-04	-0.167	0.867637	
LYCO	-1.039e-07	8.257e-07	-0.126	0.899845	
LZ	7.400e-06	4.611e-06	1.605	0.108650	.
VB1	-2.013e-02	1.381e-02	-1.458	0.145024	.
VB2	-2.422e-02	1.083e-02	-2.236	0.025423	*
NIAC	8.910e-04	1.051e-03	0.848	0.396519	
VB6	-1.524e-02	8.359e-03	-1.824	0.068282	.
FDFE	-1.730e-05	3.004e-05	-0.576	0.564592	
CHL	-2.047e-04	1.163e-04	-1.761	0.078368	.
TVB12	4.204e-03	1.703e-03	2.469	0.013598	*
VC	-7.838e-05	8.624e-05	-0.909	0.363467	
VD	-4.279e-03	1.514e-03	-2.827	0.004728	**
VK	-1.701e-04	9.709e-05	-1.752	0.079913	.
Ca	-2.437e-05	2.372e-05	-1.027	0.304412	
P	1.139e-04	3.268e-05	3.485	0.000497	***
Mg	-3.263e-04	1.146e-04	-2.848	0.004420	**
Fe	2.931e-03	1.475e-03	1.987	0.046937	*
Zn	-1.700e-03	1.488e-03	-1.142	0.253393	
Cu	-5.470e-03	1.318e-02	-0.415	0.678203	
Na	1.474e-05	6.797e-06	2.169	0.030117	*
K	2.109e-05	1.349e-05	1.563	0.118130	.
Se	-4.189e-05	2.110e-04	-0.199	0.842605	
CAFF	1.044e-04	3.417e-05	3.055	0.002264	**
THEO	-1.373e-04	8.959e-05	-1.533	0.125464	
ALCO	-1.630e-04	2.819e-04	-0.578	0.563179	
TWC	2.336e-05	2.638e-06	8.853	< 2e-16	***
Age	1.099e-03	2.888e-04	3.805	0.000144	***
Gen	-2.530e-02	1.027e-02	-2.464	0.013791	*
as. factor (RETH)2	-7.666e-02	1.931e-02	-3.969	7.33e-05	***
as. factor (RETH)3	-8.559e-02	1.606e-02	-5.330	1.04e-07	***
as. factor (RETH)4	-3.734e-02	1.763e-02	-2.118	0.034269	*

as. factor (RETH)6	-2.413e-01	2.034e-02	-11.866	< 2e-16	***
as. factor (RETH)7	-4.199e-02	2.726e-02	-1.540	0.123545	
as. factor (Edu)2	-1.408e-02	2.227e-02	-0.632	0.527454	
as. factor (Edu)3	1.554e-02	2.000e-02	0.777	0.437038	
as. factor (Edu)4	3.064e-02	1.963e-02	1.561	0.118626	
as. factor (Edu)5	-2.600e-02	2.024e-02	-1.285	0.198941	
Residual standard error:	Residual standard error: 0.3005 on 4203 degrees of freedom				
Multiple R-squared:	0.1164		Adjusted R-squared:	0.1073	
F-statistic:	12.87 on 43 and 4203 DF		P- value:	< 2.2e-16	

* Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

We can find that high level of TVB12, P, Fe, Na, CAFF, TWC intake and low level VB2, VD, Mg intake are statistically significant to high level BMI. Moreover, elder groups and female are more possible to risk in high level BMI. Compared to Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian may have lower BMI. We call this model as baseline model, which is our model 1.

Next, we conduct non- linear transformation. Considering guarding against overfitting [2], we try square, cube, and no more than fourth power of each predictor and polynomial no more than fourth power as well in our model to improve the prediction power and estimation accuracy. We retain the replacement which could greatly increase R- squared to some degree or its p- value. We can see the abstract comparison in Table 3.

Table 3. Abstract Comparison between Baseline Model and Transformed Model

Baseline Model				Transformed Model		
	Pr (>| t |)				Pr (>| t |)	
VARA	0.938683			I(VARA^2)	0.027730	*
NIAC	0.396519			I(NIAC^4)	0.036103	*
TVB12	0.013598	*		I(TVB12^2)	0.004943	**
Ca	0.304412	.		I(Ca^2)	0.039926	*
Mg	0.004420	**		poly (Mg, 2)1	0.000508	***
				poly (Mg, 2)2	0.068406	.
Zn	0.253393			poly (Zn, 2)1	0.306714	
				poly (Zn, 2)2	0.024270	*
Cu	0.678203			I(Cu^2)	0.011716	*
Na	0.030117	*		poly (Na, 2)1	0.006308	**
				poly (Na, 2)2	0.000116	***
K	0.118130			I(K^2)	0.008289	**
ALCO	0.563179			poly (ALCO, 4)1	0.277290	
				poly (ALCO, 4)2	0.199224	
				poly (ALCO, 4)3	0.133287	
				poly (ALCO, 4)4	0.002895	**
Age	0.000144	***		poly (Age, 2)1	0.000265	***
				poly (Age, 2)2	2.88e-16	***
Multiple R-squared: 0.1164				Multiple R-squared: 0.14		
Adjusted R-squared: 0.1073				Adjusted R-squared: 0.1297		
p-value: < 2.2e-16				p-value: < 2.2e-16		

* Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

Besides, we also want to explore the impact of interaction on the regression, so we calculate the full interaction of our independent variables. For simplicity, only those interaction with more than 2 stars (p- value ≤ 0.001) were taken into consideration (see Table S3). When we add all of the 10 interactions into our transformed model, we find that only the interactions of (NIAC * Age) (p- value = 0.021659) and (Age * Gen) (p- value = 0.020507) are statistically significant. Finally, we figure out a best non-linear regression model to some extend (result see Table S4), which is our model 2. In this model, we can find that the interaction of Age and NIAC, CHOL, TVB12^2, P, Fe, K^2 and Zn^2, Na^1, ALCO^4 in polynomials shows strong and statistically significant positive correlation with BMI while the interaction of Age and Gen, VB1, NIAC^4, VB6, CHL, VD, Cu^2, Age^2 and Mg^1, Na^2 in polynomials shows statistically significant negative correlation with BMI. Compared to Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian may have lower BMI. In addition, the R square of model 2 is 0.1312.

Next, we conduct all- subsets selection (result see Table 4).

Table 4. the Result of All- Subset Selection

subset	Chosen Variables	R^2
1	as. factor (RETH)6	0.05714483
2	poly (Age, 3)2, as. factor (RETH)6	0.07591892
3	TWC, poly (Age, 3)2, as. factor (RETH)6	0.08523962
4	TWC, poly (Age, 3)2, as. factor (RETH)6, as. factor (Edu)5	0.09460479
5	TWC, poly (Mg, 2)1, poly (Na, 2)1, poly (Age, 3)2, as. factor (RETH)6	0.10275602
6	TWC, poly (Mg, 2)1, poly (Na, 2)1, poly (Age, 3)2, as. factor (RETH)6, as. factor (Edu)5	0.10887167
7	TWC, poly (Mg, 2)1, poly (Na, 2)1, poly (ALCO, 4)4, poly (Age, 3)2, as. factor (RETH)6, as. factor (Edu)5	0.11087201
8	TWC, poly (Mg,2)1, poly (Na,2)1, poly (Age, 3)2, as. factor (RETH)2, as. factor (RETH)3, as. factor (RETH)6, as. factor (Edu)5	0.11310718

It returns subsets and each size up to 8. We print R2 of each selection and make the plot of it (see Figure S5). As expected, the R^2 statistic increases monotonically as more variables are included. Then, we plot RSS, adjusted R2, Cp, and BIC for all of the models to decide which model to select (see Figure S6). As for adjusted R2 and bic, the larger they are, the better the model is. Instead, for Cp, the smaller it is, the better the model is. As the plots shows, all those 4 methods suggest that the subset with 8 independent variables is the best one. The coefficient estimates associated with this model are in Table 5 and we call this model as model 3.

Table 5. Coefficient Estimates of model 3

(Intercept)	TWC	poly (Mg, 2)1	poly (Na, 2)1	poly (Age, 3)2
4.850072e+00	1.937641e-05	-2.769208e+00	2.112013e+00	-2.786726e+00
as. factor (RETH)2	as. factor (RETH)3	as. factor (RETH)6	as. factor (Edu)5	
-5.621064e-02	-4.124967e-02	-2.352730e-01	-5.463195e-02	

In this model, we can find that there is a positive correlation between TWC, Na^1 in polynomial and BMI, and there is also a negative correlation between Mg^1, Age^2 in polynomial and BMI. What's more, compared to Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Asian have more possibility to have lower BMI and compared to those whose education level are Less than 9th grade, those with education level of College graduate or above may have lower BMI.

As for Shrinkage Method, first one is Ridge regression. We calculate the dimension of ridge model, which has 47 independent variables and 100 groups of βj (coefficients) . We plot this result in Figure S7. As we can see, with enlargement of λ, βj separate gradually. And the result of cross-validation

method to choose the best λ is showed in Figure S8. The first dotted line in Figure S8 is the best λ with min MSE of the model. We calculate the exact best λ is equal to 0.03065084 and we make the plot of y in test dataset and y predict (see Figure 2). If x = y, we can say it's a perfect model. We call this ridge model as our model 4 and the coefficients are in Table S5. Here, we will not go into details of the result. As expected, no coefficients is equal to 0.

Next one is lasso regression model. We use the cross-validation method to choose the best λ (result in Figure S9) as well. Use the standard above, the best λ is 0.002213239. We also make the plot of y in test dataset and y predict (see Figure 3).

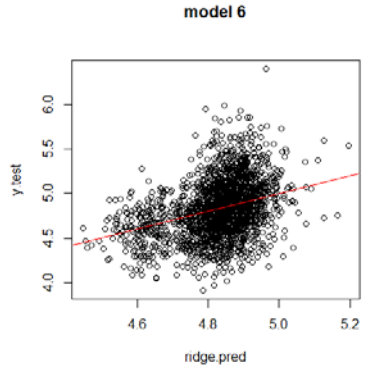


Figure 2. y predicts and y test of ridge regression model

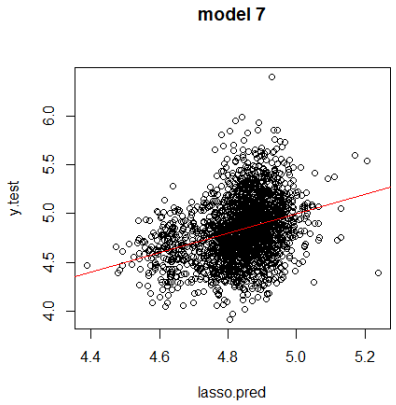


Figure 3. y predicts and y test of lasso regression model

We call this model as model 5 and the coefficients which are not equal to 0 are in Table 6, because those variables whose coefficients are equal to 0 means not important and we don't need to include them into our model.

Table 6. Coefficient Estimates of model 5

(Intercept)	CHOL	TVE	BCAR	FDFE	VC
4.909137e+00	5.176403e-05	-2.442991e-06	-2.110041e-06	-5.870075e-06	-7.579831e-05
VD	poly (Mg, 2)1	poly (Na, 2)1	TWC	RETHNon-HispanicAsian	
-1.550160e-04	7.874318e-06	1.455158e-05	1.274234e-05	9.554800e-05	

In this model, we can find CHOL, TWC and Mg^1, Na^1 in polynomial have statistically significant positive correlation with BMI, while TVE, BCAR, FDFE, VC and VD have negative correlation with BMI. Besides, Non-Hispanic Asian have more possibility to have higher level BMI than Mexican American. We can see the tendency of R- squared changes with lambda increasing in Figure S10. And we input the Mean- squared error (cvm) of model 4 and model 5 in the formula to calculate the exact R- squared of them, which is 0.07109264 for model 4 and 0.06858774 for model 5.

3.2 Comparison

Now, we have 5 models and next, we make a comparison of them. Firstly, we plot 2 figures for the former two models, which is fitted value vs residuals and fitted value vs true value (see Figure 4).

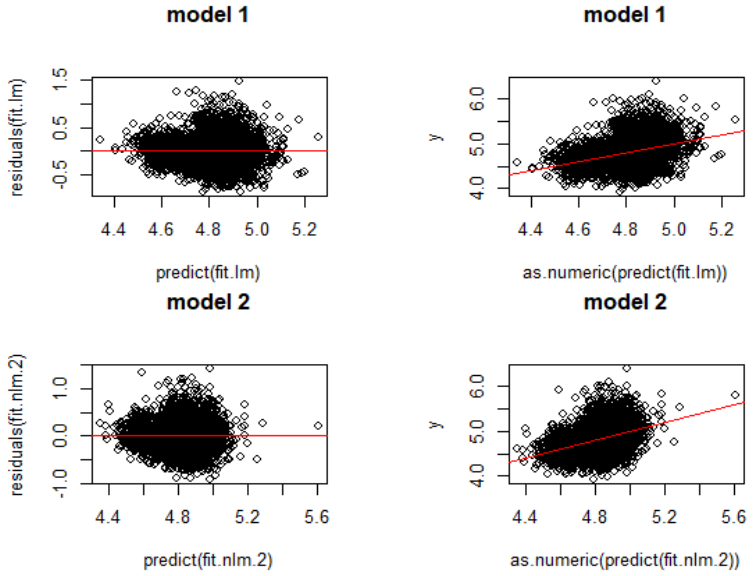


Figure 4. Fitted Value vs Residuals and Fitted Value vs True Value of Model 1, 2

In the residual's figures, if the dots distribute homogeneously on each side of the red line, then we say it is a good residual. Otherwise, if dots distribute in homogeneously from left to right, we say there is heteroscedasticity in this model. Moreover, we summary the R- squared of all the 5 models in Table 7.

Table 7. R- squared of 6 models

	R-squared:	Adjusted R-squared
Model 1	0.1164	0.1073
Model 2	0.142	0.1312
Model 3	0.11310718	0.11143300
Model 4	0.07109264	/
Model 5	0.06858774	/

Overall, as we see, although model 2 has larger R- squared than model 1 so that model 2 could explain y better, the residuals don't perform well as model 1's and true y departs even further from predict y. As for model 3, 4, and 5, the R- squared of model 3 is a little larger than that of model 4 and 5. When it comes to model 4 and 5, we can compare Figure 2 and 3. We can find that in these two figures, y predict performs nearly as well as each other. But the R- squared of model 4 is a little larger than that of model 5 although they are both in low level. But as what Dahlgren's work suggests [7], lasso regression could only include important variables in the model while ridge regression couldn't result in variable selection so that it may has a lot of noises.

To sum up, model 3 (all- subsets regression) and model 5 (lasso regression) are the best model. But model 5 includes more micronutrients, which is more related to our topic.

4. Discussion

In this cross-sectional analysis of a nationally representative sample of 4,247 US adults aged 20 and older, we evaluated the association between micronutrient intake and BMI, and further determined micronutrients with the strongest association with BMI among all NHANES micronutrients. Our main finding was that a higher BMI was associated with more intake of sodium and magnesium, and conversely with lower intake of total vitamin E, beta-carotene, folate, vitamin D, and vitamin D. Our

results were consistent with previous studies that have assessed the association between higher BMI and limited nutrient levels.

Previous studies have revealed that an important feature of people with obesity is systemic low-grade chronic inflammation, with cells in adipose tissue involved in the production and release of inflammation-related biomarkers [4]. Increased abdominal fat mass is associated with a chronic increase in circulating concentrations of inflammatory mediators, including several acute phase inflammatory proteins such as CRP, pro-inflammatory and anti-inflammatory cytokines, adhesion molecules, and pro-thrombotic molecules [4]. The effects of dietary micronutrients on the level of inflammatory biomarkers in human body as well as body weight have been shown with extensive exploratory in their biological mechanisms.

A variety of studies have shown that vitamins, including β carotenoids, vitamin C, vitamin D and vitamin E, are involved in regulation of obesity by functioning as inflammation inhibitors, which are in line with our observation. According to a study conducted by Silveira et al., decreased C-reactive protein level was observed after daily intake of 750 mL red orange juice, indicating a decreased inflammatory state and an increased serum antioxidant capacity [24], where the red orange juice is a source of provit A carotenoids (β-cryptoxanthin and β-carotene) which may help improve metabolic parameters [1]. Several cell studies in vitro have also demonstrated the ability of β-carotenoids to reduce reactive oxygen species and positively regulate inflammation and oxidative stress [13]. As an effective water-soluble antioxidant, vitamin C mainly functions as ascorbic acid, which has antioxidant function and plays an important role in inflammation and prevention of oxidative damage. Increased oxidative stress associated with obesity may lead to vitamin C destruction, leading to the lower vitamin C levels observed in obese individuals [4]. Vitamin E is an effective chain breaking antioxidant, preventing the propagation of the reaction initiated by free radicals. In vitro, vitamin E plays a series of anti-inflammatory roles in proinflammatory cytokine and arachidonic production and monocyte adhesion interactions with endothelial cells [4], which may explain the negative correlation between total vitamin E and BMI observed in this study. The negative association between vitamin D and BMI found in this study is consistent with previous studies, and diverse explanations have been proposed. One theory is that vitamin D has a variety of anti-inflammatory effects and that a decrease in its level is associated with an increase in inflammation [4]. Another mechanism of lower vitamin D levels in obese people include reduced bioavailability by fat isolation of vitamin D from skin and dietary sources [15]. It is also suggested that vitamin D evolved as a photoreceptor sensitive to UV-B, and that a drop-in vitamin D stimulates the winter response and promotes fat accumulation [9].

Lower level of folates was often observed in chronic inflammatory diseases including obesity [12]. However, the effects of folates on inflammation are up to dosage. In mice, folates supplementation in the maternal diet has been shown to counteract the effects of a high-fat maternal diet on weight gain in offspring [5], while an excessive intake may lead to an inverse effect [21].

Higher intake of mineral micronutrients such as sodium and magnesium were positively associated with BMI. It has been shown in clinical studies that sodium restriction leads to a decrease in inflammatory biomarkers [2], which may explain the positive association between sodium intake and BMI, given the nature of obesity as a chronic inflammatory disease. The magnesium deficiency is associated with an increase in chronic low-grade inflammation, primarily by increasing extracellular calcium ions to promote inflammation [19], which is inconsistent with our finding. A potential explanation could be excessive intake of magnesium could lead to the increased cell size, which increases the risk of an inflammatory response.

Strength of this study includes the use of data from a large, nationally representative survey of adults in the United States, which included both extensive physical examination and assessment of dietary intakes. However, this study suffers from inevitable limitations including the cross-sectional nature of NHANES, which make it impossible to infer a directional conclusion or causality between dietary micronutrient intake and BMI. In addition, we failed to obtain a complete diet record for each participant, but simulated it with two self-reported 24-hour diet records instead, which leads to recall biases. We also did not address the challenge posed by seasonal variations in participants' diet preference, but such variation is relatively limited, so we still consider our study robust.

5. Conclusion

The findings of the present study provide robust associations between micronutrient intake and BMI that a higher BMI was associated with more intake of sodium and magnesium, and conversely with lower intake of total beta-carotene, folate, vitamin C, vitamin D and vitamin E. These findings are significant in that they help to update the current relationship between micronutrient intake and BMI, directly helping to improve existing dietary intake recommendations for U.S. adults.

References

[1] Amiot, M. J., Riva, C., & Vinet, A. (2016). Effects of dietary polyphenols on metabolic syndrome features in humans: a systematic review. Obesity reviews, 17(7), 573-586.

[2] Ávila-Díaz, M., Ventura, M. D. J., Valle, D., Vicente-Martinez, M., Garcia-Gonzalez, Z., Cisneros, A., ... & Paniagua, R. (2006). Inflammation and extracellular volume expansion are related to sodium and water removal in patients on peritoneal dialysis. Peritoneal dialysis international, 26(5), 574-580.

[3] Bailey L. B. (1998). Dietary reference intakes for folate: the debut of dietary folate equivalents. Nutrition reviews, 56(10), 294–299.

[4] Calder, P. C., Ahluwalia, N., Brouns, F., Buetler, T., Clement, K., Cunningham, K., ... & Winklhofer-Roob, B. M. (2011). Dietary factors and low-grade inflammation in relation to overweight and obesity. British Journal of Nutrition, 106(S3), S1-S78. https://doi.org/10.1111/j.1753-4887.1998.tb01662.x

[5] Carlin, J., George, R., & Reyes, T. M. (2013). Methyl donor supplementation blocks the adverse effects of maternal high fat diet on offspring physiology. PloS one, 8(5), e63549.

[6] Centers for Disease Control and Prevention. (2021). Overweight & Obesity. Retrieved from https://www.cdc.gov/obesity/index.html

[7] Dahlgren J. P. (2010). Alternative regression methods are not considered in Murtaugh (2009) or by ecologists in general. Ecology letters, 13(5), E7–E9. https://doi.org/10.1111/j.1461-0248.2010.01460.x

[8] Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. Jama, 307(5), 491-497.

[9] Foss, Y. J. (2009). Vitamin D deficiency is the cause of common obesity. Medical hypotheses, 72(3), 314-321.

[10] Fryar, C. D., Carroll, M. D., & Afful, J. (2020). Prevalence of overweight, obesity, and severe obesity among adults aged 20 and over: United States, 1960–1962 through 2017–2018. NCHS Health E-Stats. 2020. Internet: https://www.cdc.gov/nchs/data/hestat/obesity-adult-17-18/obesity-adult.htm (accessed 15 February 2022).

[11] Fulgoni, V. L., 3rd, Keast, D. R., Bailey, R. L., & Dwyer, J. (2011). Foods, fortificants, and supplements: Where do Americans get their nutrients. The Journal of nutrition, 141(10), 1847–1854. https://doi.org/10.3945/jn.111.142257

[12] Jones, P., Lucock, M., Scarlett, C. J., Veysey, M., & Beckett, E. L. (2019). Folate and Inflammation–links between folate and features of inflammatory conditions. Journal of nutrition & intermediary metabolism, 18, 100104.

[13] Kaulmann, A., & Bohn, T. (2014). Carotenoids, inflammation, and oxidative stress—implications of cellular signaling pathways and relation to chronic disease prevention. Nutrition research, 34(11), 907-929.

[14] Kidambi, S., Batsis, J. A., Donahoo, W. T., Jastreboff, A. M., Kahan, S., Saunders, K. H., & Heymsfield, S. B. (2021). Dietary supplements and alternative therapies for obesity: A Perspective from The Obesity Society's Clinical Committee. Obesity, 29(7), 1095-1098.

[15] Kimmons, J. E., Blanck, H. M., Tohill, B. C., Zhang, J., & Khan, L. K. (2006). Associations between body mass index and the prevalence of low micronutrient levels among US adults. Medscape General Medicine, 8(4), 59.

[16] Li, Y., Wang, C., Zhu, K., Feng, R. N., & Sun, C. H. (2010). Effects of multivitamin and mineral supplementation on adiposity, energy expenditure and lipid profiles in obese Chinese women. International Journal of Obesity, 34(6), 1070-1077.

[17] Major, G. C., Alarie, F. P., Doré, J., & Tremblay, A. (2008). Calcium plus vitamin D supplementation and fat mass loss in female very low-calcium consumers: potential link with a calcium-specific appetite control. British journal of nutrition, 101(5), 659-663.

[18] Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: turning noise into signal pollution. The American naturalist, 173(1), 119–123. https://doi.org/10.1086/593303

[19] Nielsen, F. H. (2018). Magnesium deficiency and increased inflammation: current perspectives. Journal of inflammation research, 11, 25.

[20] North American Association for the Study of Obesity., National Heart, Lung, and Blood Institute., National Institutes of Health (U.S.), & NHLBI Obesity Education Initiative. (2000). The practical guide: Identification, evaluation, and treatment of overweight and obesity in adults. Bethesda, Md.: National Institutes of Health, National Heart, Lung, and Blood Institute, NHLBI Obesity Education Initiative, North American Association for the Study of Obesity.

[21] Pannia, E., Cho, C. E., Kubant, R., Sánchez-Hernández, D., Huot, P. S., Chatterjee, D., ... & Anderson, G. H. (2015). A high multivitamin diet fed to Wistar rat dams during pregnancy increases maternal weight gain later in life and alters homeostatic, hedonic and peripheral regulatory systems of energy balance. Behavioural Brain Research, 278, 1-11.

[22] Pillitteri, J. L., Shiffman, S., Rohay, J. M., Harkins, A. M., Burton, S. L., & Wadden, T. A. (2008). Use of dietary supplements for weight loss in the United States: results of a national survey. Obesity, 16(4), 790-796.

[23] Pi-Sunyer, X. (2009). The medical risks of obesity. Postgraduate medicine, 121(6), 21-33. https://doi.org/10.3810/pgm.2009.11.2074

[24] Silveira, J. Q., Dourado, G. K., & Cesar, T. B. (2015). Red-fleshed sweet orange juice improves the risk factors for metabolic syndrome. International Journal of Food Sciences and Nutrition, 66(7), 830-836.

[25] Vrieze S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychological methods, 17(2), 228–243. https://doi.org/10.1037/a0027127

[26] Ward, Z. J., Bleich, S. N., Cradock, A. L., Barrett, J. L., Giles, C. M., Flax, C., ... & Gortmaker, S. L. (2019). Projected US state-level prevalence of adult obesity and severe obesity. New England Journal of Medicine, 381(25), 2440-2450.

Appendix

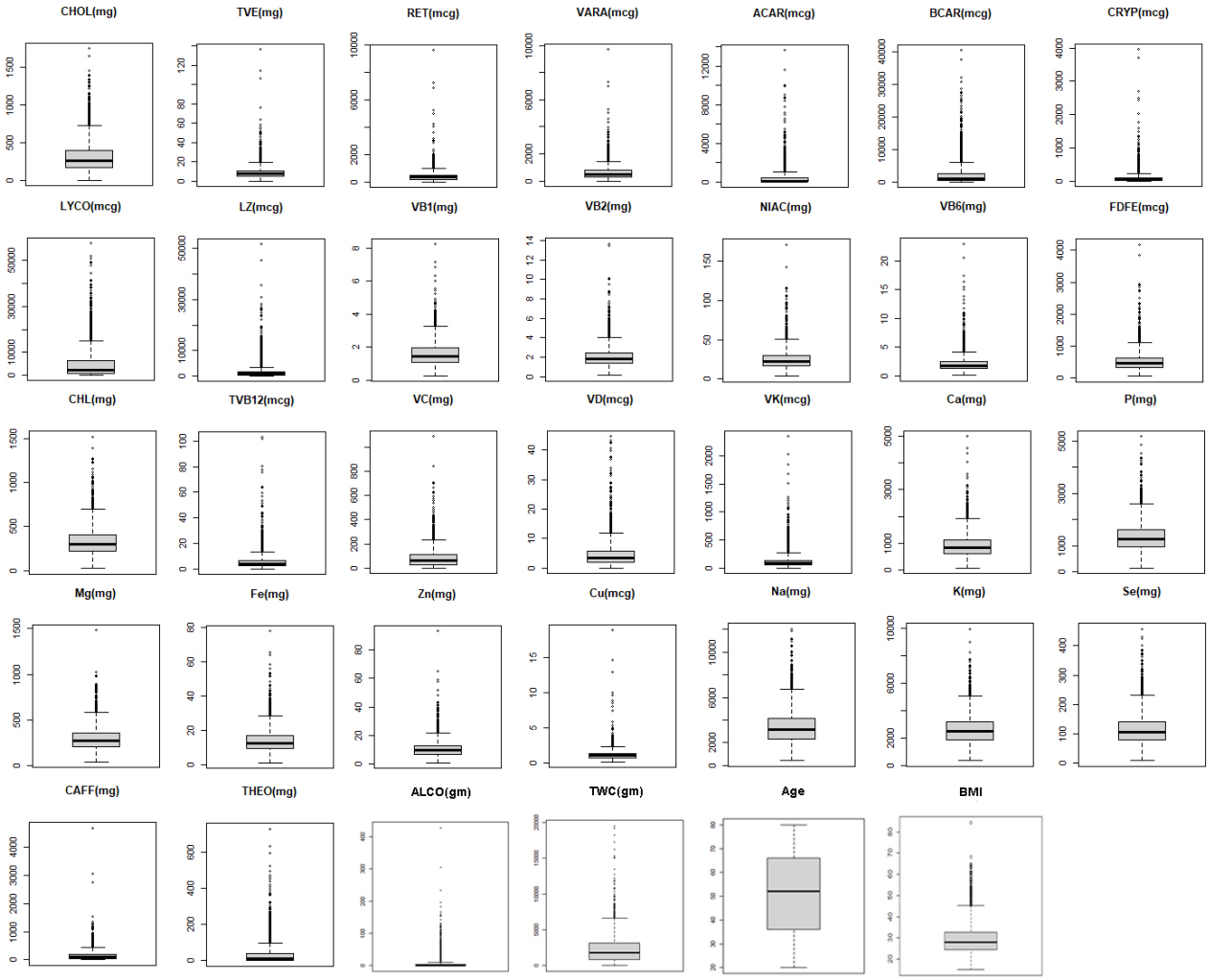


Figure S1. Basic Distribution of Continuous Variables

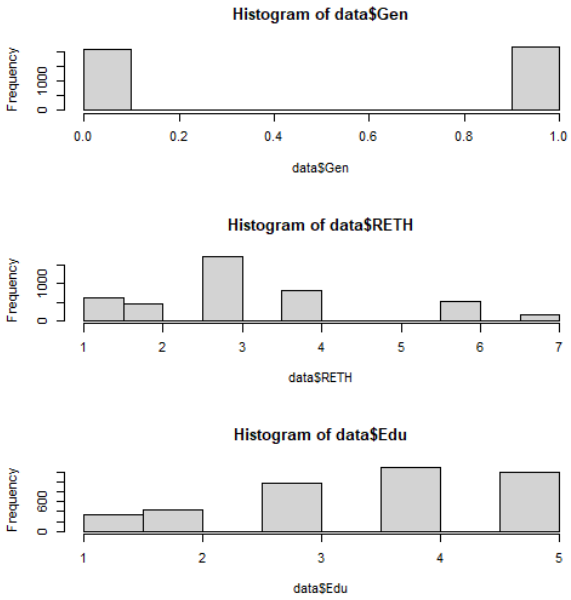


Figure S2. Basic Distribution of Categorical Variables

Table S1. Summary statistics

Item	Stand for	Min	1st Qu	Median	Mean	3rd Qu	Max
BMI	Body Mass Index	15.10	24.30	28.00	29.11	32.70	84.40
CHOL	Cholesterol (mg)	0.0	165.0	260.0	302.4	390.5	1751.0
TVE	total Vitamin E (mg)	0.290	5.225	7.620	9.288	10.912	132.690
RET	Retinol (mcg)	0.0	194.0	327.5	400.6	509.2	9650.5
VARA	Vitamin A, RAE (mcg) *	0.0	323.0	508.5	613.6	774.5	9729.0
ACAR	Alpha-carotene (mcg)	0.0	25.0	84.5	398.3	439.5	13682.0
BCAR	Beta-carotene (mcg)	0.0	467.8	1157.5	2313.6	2726.2	40406.5
CRYP	Beta-cryptoxanthin (mcg)	0.00	16.50	43.50	96.44	107.00	3956.00
LYCO	Lycopene (mcg)	0.0	711.5	2426.5	4762.9	6418.8	57439.0
LZ	Lutein + zeaxanthin (mcg)	0.5	489.8	891.0	1666.5	1664.8	51701.5
VB1	Thiamin (Vitamin B1) (mg)	0.2405	1.0750	1.4565	1.5838	1.9575	8.2275
VB2	Riboflavin (Vitamin B2) (mg)	0.1095	1.3500	1.8165	2.0031	2.4143	13.7010
NIAC	Niacin (mg)	3.371	16.772	22.791	24.998	30.266	171.352
VB6	Vitamin B6 (mg)	0.0365	1.2985	1.8080	2.0660	2.4540	23.0245
FDFE	Folate, DFE (mcg) *	43.5	321.0	446.5	511.1	630.5	4184.5
CHL	Total choline (mg)	22.85	220.38	302.95	330.70	410.05	1518.90
TVB12	total Vitamin B12 (mcg)	0.000	2.415	4.005	5.609	6.715	103.295
VC	Vitamin C (mg)	0.00	29.68	61.85	82.88	111.30	1090.50
VD	Vitamin D (D2 + D3) (mcg)	0.000	1.850	3.450	4.577	5.800	44.700
VK	Vitamin K (mcg)	0.15	51.55	82.50	121.17	136.88	2354.15
Ca	Calcium (mg)	59.0	591.2	832.5	911.7	1127.0	5016.0
P	Phosphorus (mg)	136	962	1257	1335	1614	5186
Mg	Magnesium (mg)	39.5	208.2	275.5	295.4	357.2	1480.0
Fe	Iron (mg)	1.210	9.527	12.785	14.082	17.265	77.970
Zn	Zinc (mg)	0.870	7.130	9.685	10.703	13.090	93.060
Cu	Copper (mg)	0.1465	0.8080	1.0740	1.1942	1.4130	18.8945
Na	Sodium (mg)	418	2374	3180	3376	4125	12056
K	Potassium (mg)	342.5	1871.8	2467.5	2587.5	3155.0	9956.5
Se	Selenium (mcg)	6.7	77.6	104.0	113.2	139.2	458.2
CAFF	Caffeine (mg)	0.0	33.0	101.0	142.4	195.5	4680.0
THEO	Theobromine (mg)	0.0	0.0	9.5	30.8	39.5	727.5
ALCO	Alcohol (gm)	0.000	0.000	0.000	7.304	4.000	427.450
TWC	Total water drank yesterday (gm)	0.0	866.8	1830.0	2278.7	3180.0	19485.0
Age	Age	20.00	36.00	52.00	51.27	66.00	80.00

* Vitamin A, RAE: Vitamin A as retinol activity equivalents

* Folate, DFE: Folate as dietary folate equivalents

Table S2. Standard deviation and Variance of BMI

Standard deviation	**Variance**
6.824075	46.568

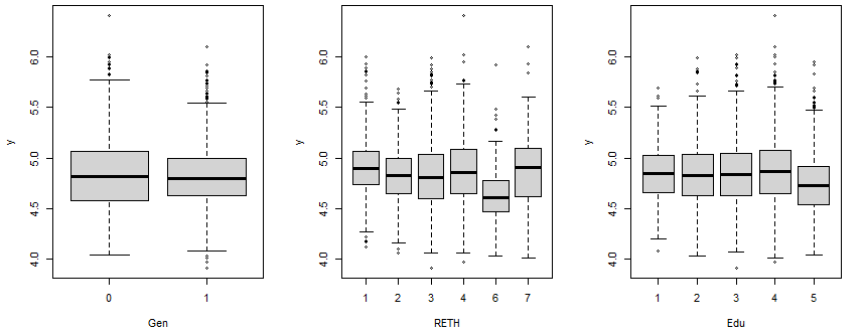


Figure S3: Pairs function

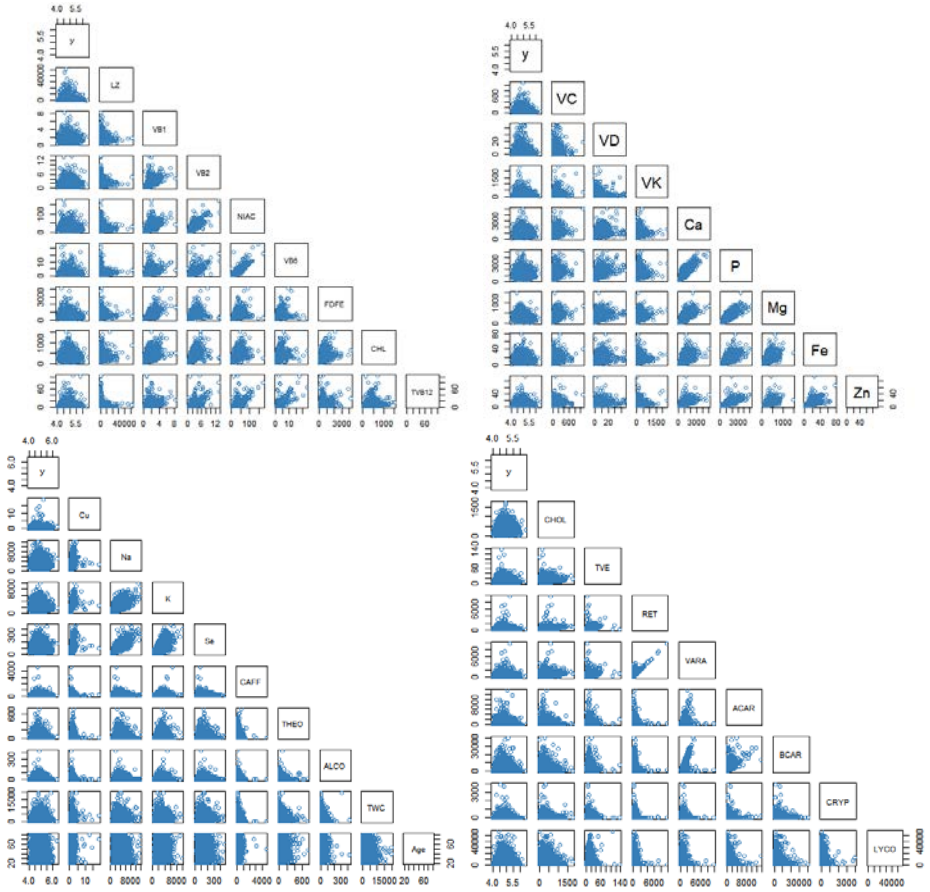


Figure S4: The Correlation between y and Gen, RETH and Edu

Table S3. Part of the Results of Full Interaction

Coefficients:					
	Estimate	Std. error	t value	Pr (>| t |)	
RET: VD	-4.363e-03	1.390e-03	-3.138	0.00172	**
VARA: VD	4.380e-03	1.390e-03	3.152	0.00164	**
ACAR: VD	-1.831e-04	5.808e-05	-3.153	0.00163	**
BCAR: VD	-3.661e-04	1.158e-04	-3.161	0.00159	**
CRYP: VD	-1.851e-04	6.034e-05	-3.067	0.00218	**
LZ:VB2	-5.562e-05	1.981e-05	-2.807	0.00503	**
VB2: VK	1.291e-03	4.078e-04	3.167	0.00156	**
NIAC: Age	2.779e-04	9.244e-05	3.006	0.00267	**
CHL: TVB12	-1.604e-04	5.942e-05	-2.699	0.00700	**
Age: Gen	-1.861e-03	7.066e-04	-2.634	0.00847	**

* Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

Table S4. the Result of Non- Linear Regression

Residuals:					
Min	1Q	Median	3Q	Max	
-0.91777	-0.20184	-0.01151	0.18504	1.41544	
Coefficients:					
	Estimate	Std. error	t value	Pr (>| t |)	
(Intercept)	4.803e+00	4.929e-02	97.444	< 2e-16	***
I(NIAC * Age)	3.761e-05	1.672e-05	2.249	0.024555	*
I(Age * Gen)	-1.225e-03	5.221e-04	-2.346	0.019026	*
CHOL	1.582e-04	7.514e-05	2.106	0.035299	*
TVE	-7.014e-04	8.503e-04	-0.825	0.409467	
RET	-5.843e-05	3.321e-05	-1.759	0.078619	.
I(VARA^2)	1.734e-08	9.961e-09	1.741	0.081832	.
ACAR	-8.340e-06	7.661e-06	-1.089	0.276372	
BCAR	-5.013e-06	3.200e-06	-1.567	0.117283	
CRYP	-3.534e-05	2.777e-05	-1.273	0.203166	
LYCO	-1.799e-07	8.143e-07	-0.221	0.825157	
LZ	6.711e-06	4.548e-06	1.475	0.140162	
VB1	-2.782e-02	1.345e-02	-2.069	0.038627	*
VB2	-1.508e-02	1.014e-02	-1.488	0.136926	
I(NIAC^4)	-9.133e-10	3.840e-10	-2.378	0.017444	*
VB6	-1.512e-02	7.189e-03	-2.104	0.035468	*
FDFE	-1.185e-05	2.995e-05	-0.396	0.692382	
CHL	-2.241e-04	1.141e-04	-1.963	0.049664	*
I(TVB12^2)	8.373e-05	2.662e-05	3.145	0.001674	**
VC	-9.146e-05	8.388e-05	-1.090	0.275617	
VD	-3.525e-03	1.427e-03	-2.470	0.013538	*
VK	-1.399e-04	9.587e-05	-1.459	0.144693	
I(Ca^2)	-1.167e-08	6.874e-09	-1.698	0.089628	.
P	1.157e-04	3.084e-05	3.750	0.000179	***
poly (Mg, 2)1	-3.075e+00	8.431e-01	-3.647	0.000269	***
poly (Mg, 2)2	6.677e-01	3.530e-01	1.891	0.058645	.
Fe	3.255e-03	1.478e-03	2.202	0.027701	*
poly (Zn, 2)1	-6.551e-01	5.518e-01	-1.187	0.235245	
poly (Zn, 2)2	8.093e-01	3.702e-01	2.186	0.028877	*
I(Cu^2)	-4.909e-03	2.141e-03	-2.293	0.021908	*
poly (Na, 2)1	1.483e+00	6.161e-01	2.408	0.016099	*
poly (Na, 2)2	-1.289e+00	3.312e-01	-3.893	0.000101	***
I(K^2)	4.902e-09	1.767e-09	2.774	0.005554	**
Se	-6.636e-05	1.997e-04	-0.332	0.739716	
CAFF	3.878e-05	3.430e-05	1.131	0.258215	
THEO	-1.506e-04	8.881e-05	-1.696	0.089968	.
poly (ALCO, 4)1	-4.711e-01	3.621e-01	-1.301	0.193288	
poly (ALCO, 4)2	3.728e-01	3.059e-01	1.219	0.222993	
poly (ALCO, 4)3	-4.542e-01	3.017e-01	-1.506	0.132225	
poly (ALCO, 4)4	9.070e-01	3.000e-01	3.024	0.002514	**
TWC	2.315e-05	2.542e-06	9.106	< 2e-16	***
poly (Age, 3)1	9.349e-01	5.847e-01	1.599	0.109934	
poly (Age, 3)2	-2.483e+00	3.129e-01	-7.933	2.72e-15	***
poly (Age, 3)3	3.154e-01	3.027e-01	1.042	0.297637	
Gen	3.982e-02	2.873e-02	1.386	0.165800	
as. factor (RETH)2	-7.153e-02	1.909e-02	-3.747	0.000181	***

as. factor (RETH)3	-6.929e-02	1.611e-02	-4.302	1.73e-05	***
as. factor (RETH)4	-3.923e-02	1.746e-02	-2.246	0.024739	*
as. factor (RETH)6	-2.427e-01	2.009e-02	-12.079	< 2e-16	***
as. factor (RETH)7	-4.004e-02	2.707e-02	-1.479	0.139179	
as. factor (Edu)2	-1.523e-02	2.199e-02	-0.693	0.488498	
as. factor (Edu)3	1.084e-02	1.978e-02	0.548	0.583889	
as. factor (Edu)4	3.209e-02	1.944e-02	1.651	0.098845	.
as. factor (Edu)5	-2.648e-02	2.009e-02	-1.318	0.187610	
Residual standard error:	0.2965 on 4193 degrees of freedom				
Multiple R-squared:	0.142		Adjusted R-squared:	0.1312	
F-statistic:	13.1 on 53 and 4193 DF		P- value:	< 2.2e-16	

* Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

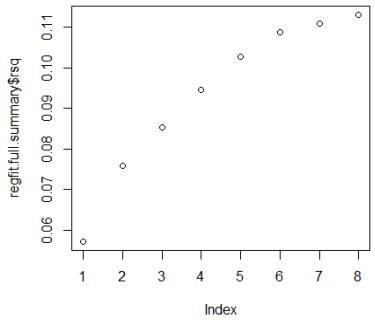


Figure S5. R2 statistic for All- Subset Selection

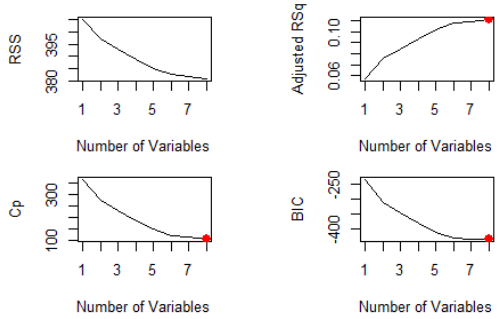


Figure S6. RSS, adjusted R2, Cp, and BIC Plots for model3

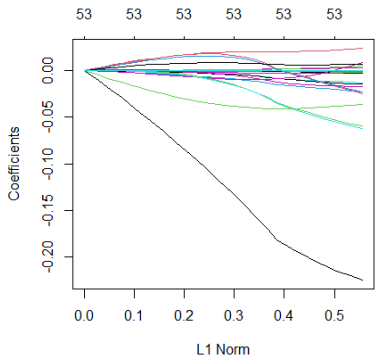


Figure S7. β Change with λ in Ridge Model

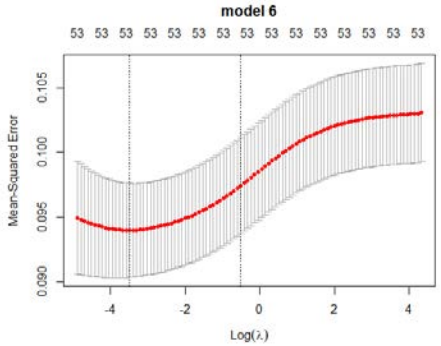


Figure S8. The Result of Cross-validation Method Used in the Ridge Model

Table S5. Coefficient Estimates of model 4

(Intercept)	CHOL	TVE	RET	VARA^2	ACAR	BCAR
4.794188e+00	8.939091e-05	-8.840559e-04	-3.328183e-05	2.694476e-09	-7.612842e-06	-1.850767e-06
CRYP	LYCO	LZ	VB1	VB2	NIAC^4	VB6
-2.581494e-05	-5.227969e-08	2.102930e-06	-1.818663e-02	-1.042731e-02	-7.597400e-10	-1.140706e-02
FDFE	CHL	TVB12^2	VC	VD	VK	Ca^2
-2.195378e-05	-6.869992e-05	5.716520e-05	-1.063162e-04	--3.210806e-03	-7.544735e-05	-1.013746e-09
P	poly (Mg, 2)1	poly (Mg, 2)2	Fe	poly (Zn, 2)1	poly (Zn, 2)2	Cu^2
4.682046e-05	-1.653907e-04	-4.995600e-08	1.998375e-03	-1.627482e-03	2.728328e-05	-1.091797e-03
poly (Na, 2)1	poly (Na, 2)2	K^2	Se	CAFF	THEO	poly (ALCO, 4)1
1.818846e-05	-6.703998e-10	1.873513e-09	-5.674080e-06	5.049535e-05	-1.237458e-04	-5.812274e-04
poly (ALCO, 4)2	poly (ALCO, 4)3	poly (ALCO, 4)4	TWC	poly (Age, 3)1	poly (Age, 3)2	poly (Ag, 3)3
1.308577e-06	-3.906782e-10	2.287027e-12	1.829499e-05	2.199128e-03	--2.289421e-06	-2.389410e-07
Gen	RETHNon-HispanicAsian	VRETHNon-HispanicBlack	RETHNon-HispanicWhite	RETHOther Race - Including Multi-Racial	RETHOtherHispanic	Edulevel2
-5.583435e-03	-2.001893e-01	-8.088100e-03	-4.505125e-02	-8.210742e-03	-4.655999e-02	-1.524621e-02
Edulevel3	Edulevel4	Edulevel5	NIAC.Age	Age.Gen		
5.837692e-03	2.063780e-02	-4.033543e-02	3.796466e-05	-3.586043e-04		

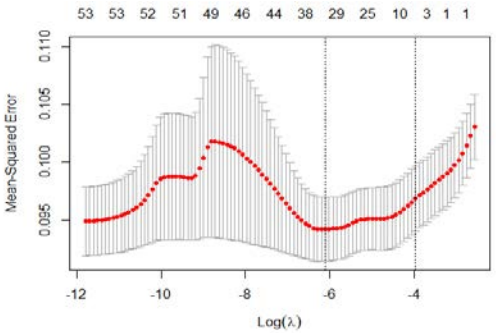


Figure S9. The Result of Cross-validation Method Used in the Lasso Model

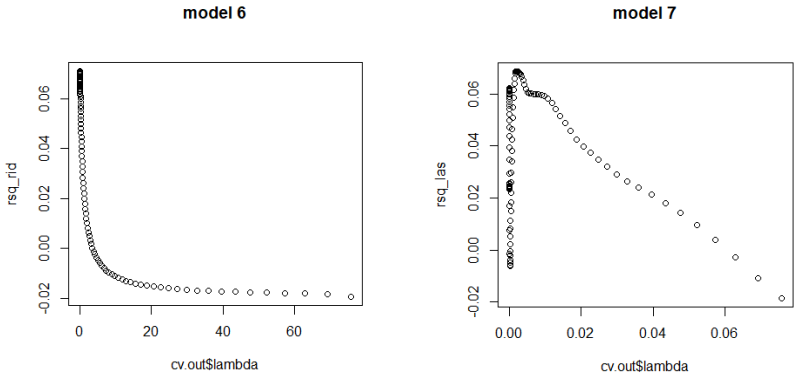


Figure S10. the tendency of R- squared changes with lambda increasing