The association between different nutrients and the risk of iron deficiency: NHANES 2017-2020

Huyang Zhang

Introduction

Iron is required for the survival of most organisms, including bacteria, plants, and humans [1]. The abnormal of iron homeostasis may cause pathophysiological conditions: the defection of iron may lead to anemia, and the overload of iron may lead to genetic disorders, inflammation and infection, cardiovascular diseases, cancer, and neurodegeneration diseases [2]. Given that ferritin level reflects on total iron stores in the blood, and the transferrin saturation indicates the iron transportation ability, the ferritin level and transferrin saturation usually are used to evaluate iron status in the clinical studies [3]. Therefore, this research will choose ferritin level and transferrin saturation as the two response variables to specify the iron status.

Iron status is a complex phenomenon that is influenced by a variety of factors, including nutrient intake. Humans derive iron from their everyday diet, predominantly from plant foods and the rest from foods of animal origin [4]. Iron absorption can vary from 1 to 40% [5] due to the different components of meal. Human generally consume micronutrients along with macronutrients, so this research considers both micronutrients and macronutrients as explanatory variables. Previous studies have investigated the relationship between various nutrients and iron status, with mixed results. Several studies have shown that vitamin C enhances non-heme iron absorption, which can improve iron status in individuals with marginal iron stores. Zinc and copper are other nutrients that have been studied in relation to iron status. Zinc deficiency has been associated with impaired iron absorption and decreased iron status, while copper deficiency has been linked to decreased iron utilization and increased risk of iron deficiency anemia. This research is needed to better understand other nutrients which influence iron status and to develop effective strategies for improving iron status in at-risk populations.

Materials and Methods

Data collection and analysis

NHANES [6] is an ongoing survey conducted by the National Center for Health Statistics at the Centers for Disease Control and Prevention (CDC) that collects data on the health and nutritional status of U.S. adults and children. The research obtained all the data from NHANES 2017-2020 database (Data in this database was collected from 2019 to March 2020 and were combined with data from the NHANES 2017-2018 cycle to form a nationally representative sample of NHANES 2017-March 2020 pre-pandemic data.)

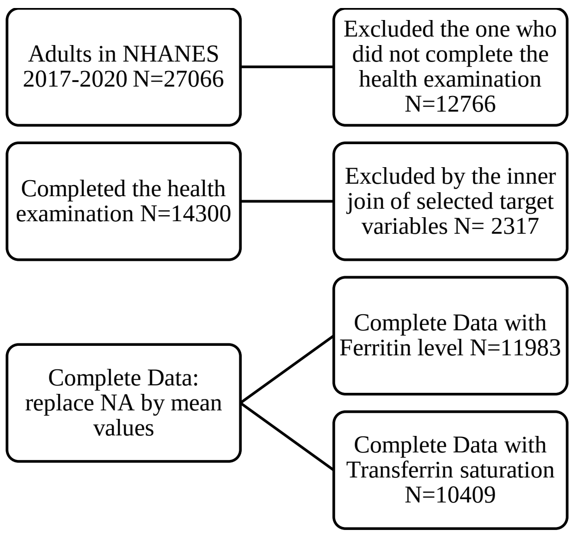

Figure 1. The flow chart of data collection and data analysis

Iron status

Ferritin level and transferrin saturation were chosen as two response variables to reflect the iron status. Ferritin level (ng/ml) was obtained from the NHANES ferritin data file from the laboratory data. Transferrin saturation was obtained form the NHANES Iron status-serum data file from the laboratory data.

The logistic transformation was used to transform one of the response variables (ferritin) to better meet the assumptions of statistical models. Before the logistic transformation, the distribution of ferritin level was extremely right skewed (Figure 1.a); after the logistic transformation, the distribution of ferritin level was approximately normal. (Figure 1.b)

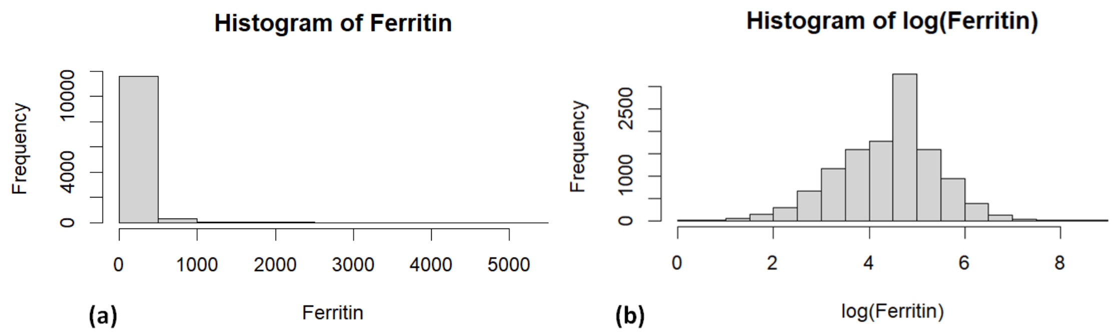

Figure 2. The comparison between the distribution of ferritin level before and after logistic transformation. (a) The distribution of ferritin level before logistic transformation (b) The distribution of ferritin level after logistic transformation

Since the original distribution of transferrin saturation was approximately normal, the logistic transformation was not required to fulfill the assumption. (But the logistic transformation of transferrin saturation was conducted to simulate the optimized linear model)

Dietary variables

Multiple dietary variables were included in NHANES dataset. Energy, protein, carbohydrate, total sugar consumed, dietary fiber, total fat consumed, total monounsaturated fat consumed, total polyunsaturated fat consumed, cholesterol, iron, vitamin C, vitamin D, vitamin K, calcium, phosphorus, magnesium, zinc, copper, sodium, potassium, selenium, caffeine and total length of food fast in minutes were selected as the possible explanatory variables. Age, gender, country of birth, race, weight and height were selected as the possible confounding variables.

Among these variables, gender, country of birth, and race were three categorical variables. Three boxplots were generated to predict whether there were relationships between these three factors to the response variables (ferritin and transferrin saturation) (Figure). The median and range of ferritin or transferrin saturation which were both numerical data might differ across different categories of gender, race, and country of birth. In figure, the median of the ferritin (after logistic transformation) for males was higher than the median of ferritin (after logistic transformation) for females, this suggested that males generally had a higher ferritin level than females. Additionally, the range of ferritin (after logistic transformation) for female might be wider than the range of ferritin (after logistic transformation) for males, indicating that there was more variability in ferritin level for females than for males. By examining the differences in the median and range of the ferritin level (after logistic transformation) across different gender, it indicated there might be a relationship between the gender and ferritin level (after logistic transformation). In terms of transferrin saturation, the medians of the transferrin saturation of two genders were different as well, indicating a possible relationship between genders and transferrin saturation. However, for two response variables, the medians of both race and country of birth were approximately the same, indicating few possibilities that race and country correlated to transferrin saturation and ferritin. (Figure 3. (a2), Figure 3. (a3), Figure 3. (b2), Figure 3. (b3)). Therefore, only the factor gender was considered in the following linear model.

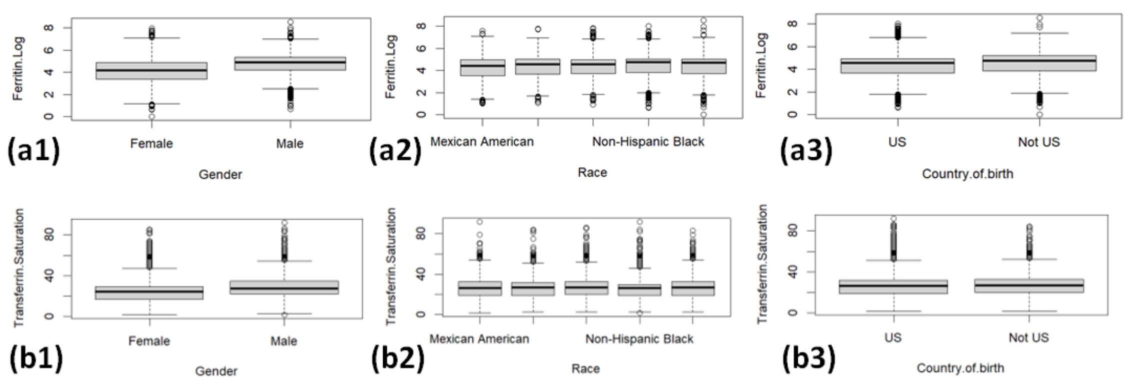

Figure 3. Boxplots that were used to examine the relationship between three categorical variables and two numerical response variables respectively.

Besides the categorical variables, other numerical explanatory variables were further selected based on the roughly predicted correlation in regard to the two response variables. By arranging the correlation coefficients in descending order, the first 8 variables of each response variables were selected as the final chosen explanatory variables.

Table 1. The explanatory variables selected for each response variables based on the descending arrangement of correlation coefficients.

Ferritin.log		Transferrin Saturation	
Varibles	**Correlation Coefficient**	**Varibles**	**Correlation Coefficient**
Age	0.326859105	Height	0.167030220
Height	0.227864961	Magnesium	0.092035271
Weight	0.225488564	Potassium	0.087240568
Protein.1	0.127957264	Age	0.082091200
Selenium	0.124902496	Weight	0.071878500
Cholesterol	0.119205759	Protein.1	0.071632184
Sodium	0.103353455	Selenium	0.069909180
Magnesium	0.101952567	Sodium	0.067882565

A correlation coefficient heat map was used to visualize the relationship between different dietary variables. The heat map displays a matrix of correlation coefficients between each pair of dietary variables, with the strength and direction of the relationship represented by a gradient of colors.

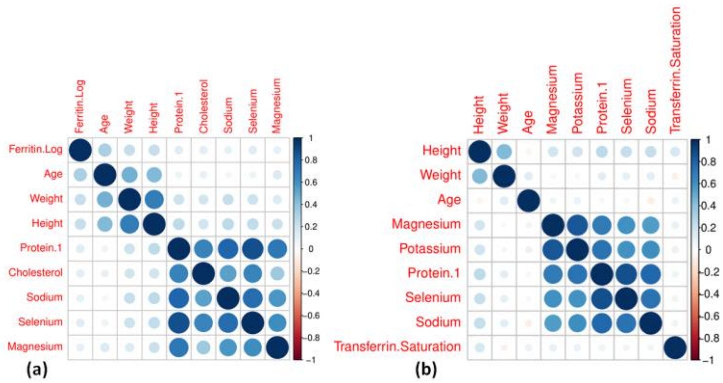

Figure 4. Correlation Coefficient Heat Map of Dietary Variables (a) Dietary Variables with ferritin data (after logistic transformation) (b) Dietary Variables with transferrin saturation

The color key on the right-hand side of the heat map shows the range of correlation coefficients, with values ranging from -1 to 1. The color white represents a correlation coefficient of 0, indicating no correlation, while the color blue represents a positive correlation, and the color red represents a negative correlation. The correlation coefficients between each pair of variables are displayed in the matrix. In Figure 4. (a), the correlation coefficients between protein and selenium, protein and sodium, sodium and selenium are relatively large and shown in relatively dark blue color. These indicate that the increase in protein may lead to the increase in sodium and selenium, and the increase in sodium may lead to the increase in selenium. Except the relationship between these dietary variables and the response variables (ferritin), there are possibly relationship between dietary variables. Therefore, there is a possibility of the existence of multilinear relationships. In Figure 4. (b), the correlation coefficients between potassium and magnesium, magnesium and protein, protein and potassium, protein and selenium, sodium and protein are relatively large and shown in relatively dark blue color. These indicate that the increase in protein may lead to the increase in sodium, potassium, magnesium and selenium, and the increase in magnesium may lead to the increase in potassium. Except the relationship between these dietary variables and the response variables (transferrin saturation), there are possibly relationship between dietary variables. Therefore, there is a possibility of the existence of multilinear relationships.

Statistical Analysis

In order to do the linear regression model, conditions were checked. First, there were true linear relationships between explanatory variables and response variables. Second, each data point represented one individual, so they were independent. Third, for a particular value of the explanatory variable, the responses were normally distributed. Fourth, the residual was roughly normally distributed. The standard deviation of errors was constant across values of x. The initial distribution of residuals was not normally distributed, but through the logistic transformation of each response variables, the histogram of each residual plot showed a roughly symmetric shape. Lastly, this observation study used the random sampling method.

Since there was a possibility of the existence of multilinear relationships, variance inflation factor was used to guarantee whether there was a necessity to exclude any of the explanatory variables or to conduct an orthogonalization. VIF measures whether there was a collinearity. Based on the commonly used threshold (VIF>5) for detecting collinearity, none of the explanatory variables in this study indicated the presence of collinearity after the optimization of linear model.

Table 2. The variance inflation factor of the selected explanatory variables in both datasets.

Ferritin.log		Transferrin Saturation	
Varibles	**VIF**	**Varibles**	**VIF**
Age	1.396328	Gender	1.736058
Weight	2.070538	Height	2.060055
Height	2.114193	Weight	1.300108
Protein.1	3.531363	Age	1.068505
Cholesterol	1.924036	Magnesium	2.013412
Magnesium	2.209379	Protein.1	3.537615
Gender	1.139076	Sodium	2.115620

Result

The results showed the investigation of the relationship between log-transformed transferrin Saturation levels and several predictor variables including Gender, Height, Weight, Age, Magnesium, Protein, and Sodium. The coefficients table showed the estimated regression coefficients for each predictor variable, along with their standard errors. All predictor variables except Protein and Sodium had statistically significant coefficients, with p-values less than 0.05. This suggested that Gender, Height, Weight, Age, and Magnesium were significant predictors of Transferrin Saturation levels. The adjusted R-squared value of 0.09723 indicated that these predictor variables explained only a small proportion (0.09723) of the total variation in Transferrin Saturation levels. The overall p-value less than 2.2e-16 indicates that the overall regression model was statistically significant. The residuals figure showed the distribution of the residuals, with no evidence of non-normality or heteroscedasticity. (Figure 5.)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.709e+00	7.040e-02	52.686	< 2e-16	***
Age	1.357e-02	4.025e-04	33.711	< 2e-16	***
Weight	1.829e-03	3.829e-04	4.778	1.8e-06	***
Height	-1.894e-03	5.556e-04	-3.409	0.000655	***
Protein.1	7.874e-04	3.928e-04	2.005	0.045023	*
Cholesterol	1.256e-04	4.893e-05	2.567	0.010274	*
Magnesium	-3.059e-04	8.815e-05	-3.470	0.000522	***
GenderMale	6.752e-01	1.754e-02	38.498	< 2e-16	***

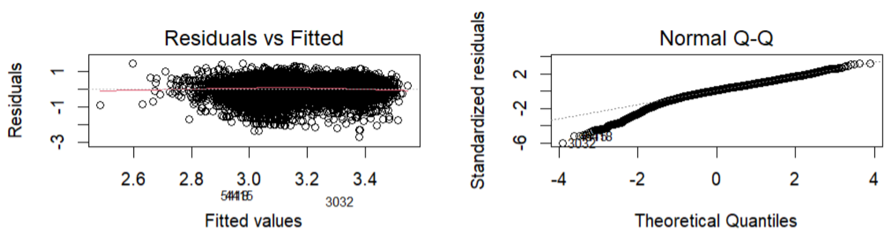

Figure 5. Coefficient table. The residual plots and the Q-Q plot of the linear model of ferritin level indicated the fulfillment of the assumptions of linear regression.

The results showed the investigation of the relationship between log-transformed Ferritin levels and several predictor variables, including Age, Weight, Height, Protein.1, Cholesterol, Magnesium, and Gender. The coefficients table showed the estimated regression coefficients for each predictor variable, along with their standard errors. All predictor variables except Height had statistically significant coefficients, with p-values less than 0.05. This suggested that Age, Weight, Protein.1, Cholesterol, Magnesium, and Gender were significant predictors of Ferritin levels. The adjusted R-squared value of 0.2202 indicated that these predictor variables explained about 22% of the total variation in Ferritin levels. The overall p-value less than 2.2e-16 indicated that the overall regression model was statistically significant. The residuals figure showed the distribution of the residuals, with no evidence of non-normality or heteroscedasticity. (Figure 6.)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.312e+00	9.575e-02	24.144	< 2e-16	***
GenderMale	1.971e-01	1.160e-02	16.988	< 2e-16	***
Height	5.304e-03	6.263e-04	8.469	< 2e-16	***
Weight	-3.258e-03	2.135e-04	-15.265	< 2e-16	***
Age	2.878e-03	2.169e-04	13.265	< 2e-16	***
Magnesium	1.216e-04	4.359e-05	2.790	0.00528	**
Protein.1	-3.871e-04	2.040e-04	-1.898	0.05779	.
Sodium	6.848e-06	4.018e-06	1.705	0.08830	.

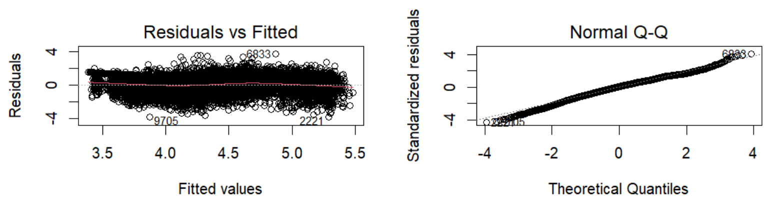

Figure 6. Coefficient table. The residual plots and the Q-Q plot of the linear model of transferrin saturation indicated the fulfillment of the assumptions of linear regression.

The variance inflation factor (VIF) values are below 5 for all predictor variables, indicating that there is no significant collinearity among the predictor variables.

Discussion

Our study firstly demonstrated the relationship between different nutrients and the risk of iron deficiency. The data was entirely obtained from the NHANES database which was reliable. The final linear models were significant because the p-value of each regression model was extremely small which was even smaller than 2.2e-16. However, the positive relationship was not strong between selected nutrients and ferritin level (r^2 =0.09723); the positive relationship was not strong as well between selected nutrients and transferrin. (r^2 =0.2202). Based on these results, people with the risk of iron deficiency can slightly increase their nutrients (Magnesium, Protein.1, Sodium; Protein.1, Cholesterol, Magnesium) intake. In the future, different models except linear model can be used to fit this data, so a more appropriate model may be generated. In addition, the investigation of different populations could be done to investigate whether the relationship was validated for a vast population.

The data of different nutrients intake was obtained from NHANES data which was reliable. However, the data was limited due to the retrospective and self-reported nature of NHANES data, so some records of individual daily intake might involve some bias. Also, because of the type of study was the observational study, we cannot generate the cause-and-effect relationship between different nutrients to iron status.

Conclusion

Nutrients (Magnesium, Protein.1, Sodium) are weakly positively associated with the ferritin level, nutrients (Protein.1, Cholesterol, Magnesium) are weakly positively associated with transferrin saturation. Therefore, the regulation of nutrients intake could possibly be slightly helpful for the recovery of iron deficiency, but the nutrients were not the dominate factor in regard to the iron status. Based on these results, people with the risk of iron deficiency can slightly increase their nutrients, but they also need to go to hospital and ask for medicine help.

Reference

[1] Gozzelino, R., & Arosio, P. (2016). Iron Homeostasis in Health and Disease. *International Journal of Molecular Sciences*, *17*(1), 130. https://doi.org/10.3390/ijms17010130

[2] Gozzelino, R., & Arosio, P. (2016). Iron Homeostasis in Health and Disease. *International Journal of Molecular Sciences*, *17*(1), 130. https://doi.org/10.3390/ijms17010130

[3] Wang, H.-H., Liao, L.-N., Chang, C.-W., Chang, Y.-C., Wu, K.-H., & Ko, J.-L. (2019). The alteration of ferritin and transferrin saturation under body mass index and blood pressure in first-time and regular male blood donors in Taiwan. *Medicine*, *98*(22), e15854. https://doi.org/10.1097/MD.0000000000015854

[4] Kouser Fathima Firdose, & Nikhath Firdose. (2021). *Dietary Iron*. https://doi.org/10.5772/intechopen.101265

[5] Kouser Fathima Firdose, & Nikhath Firdose. (2021). *Dietary Iron*. https://doi.org/10.5772/intechopen.101265

[6] *National Health and Nutrition Examination Survey (NHANES) - Health, United States*. (2022, August 8). Www.cdc.gov. https://www.cdc.gov/nchs/hus/sources-definitions/nhanes.htm