Proceedings of the 3rd International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/22/20230984

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

Huyang Zhang

Beijing National Day School, 100039, China

zhanghuyang119@gmail.com

Abstract. Iron homeostasis is important in the human health system. Abnormal iron homeostasis, such as iron deficiency (ID), may cause diseases such as anemia. Iron status is a complex phenomenon that is influenced by a variety of factors, including nutrient intake. This research is needed to understand better the different nutrients which influence iron status and to develop effective strategies for improving iron status in at-risk populations. NHANES 2017-2020 data was used to simulate the linear regression model to predict the relationship between different nutrients the iron status. Results revealed that the positive association was not strong between selected nutrients (Magnesium, Protein.1, Sodium) and ferritin level (r^2=0.09723); the positive relationship was not strong as well between selected nutrients (Protein.1, Cholesterol, Magnesium) and transferrin. (r^2=0.2202).

Keywords: NHANES, linear regression, iron deficiency, nutrients

1. Introduction

Iron is required for the survival of most organisms, including bacteria, plants, and humans [1]. The abnormality of iron homeostasis may cause pathophysiological conditions: the defection of iron may lead to anemia, and iron overload may lead to genetic disorders, inflammation and infection, cardiovascular diseases, cancer, and neurodegeneration diseases [1]. 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 clinical studies [2]. 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 [3]. Iron absorption can vary from 1 to 40% [3] due to the different components of the meal. Humans generally consume micronutrients and 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 [4], which can improve iron status in individuals with iron stores. Zinc and copper are other nutrients that have been studied about iron status. Zinc deficiency has been associated with impaired iron absorption and decreased iron status [5], while copper deficiency has been linked to reduced iron utilization and increased risk of iron deficiency anemia [6]. This research

DOI: 10.54254/2753-8818/22/20230984

is needed to understand better other nutrients which influence iron status and to develop effective strategies for improving iron status in at-risk populations.

2. Materials and Methods

2.1. Data collection and analysis

NHANES [7] 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 was combined with data from the NHANES 2017-2018 cycle to form a nationally representative sample of NHANES 2017-March 2020 pre-pandemic data.) The process of data collection and data cleaning is shown in Figure 1.

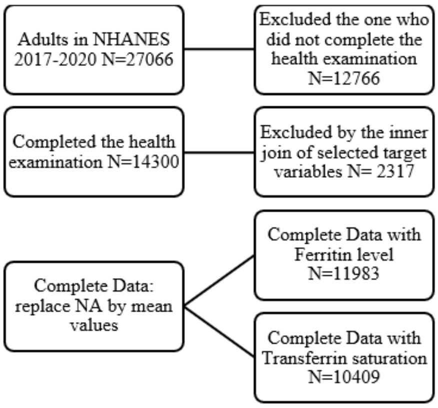

Figure 1. The flow chart of data collection and data cleaning.

2.2. Iron status

Ferritin level and transferrin saturation were 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 from the NHANES Iron status-serum data file from the laboratory data.

The logistic transformation was used to transform one of the response variables (ferritin) better to meet the assumptions of statistical models [8]. Before the logistic transformation, the distribution of ferritin level was extremely right-skewed (Figure 2. a); after the logistic transformation, the distribution of ferritin level was approximately normal (Figure 2. b).

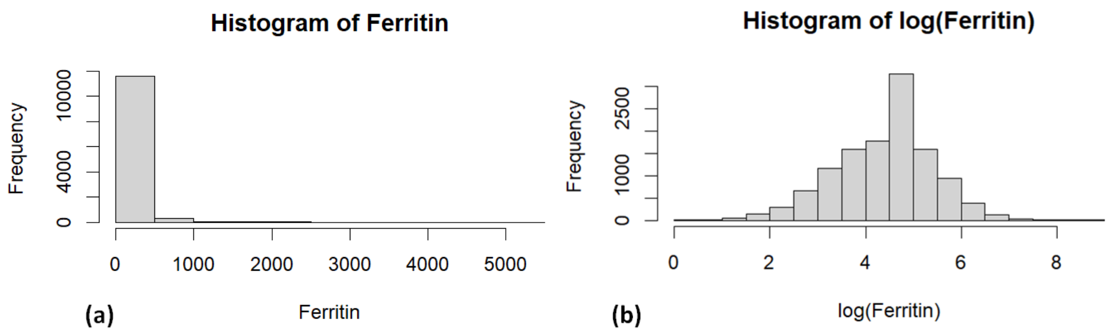

Figure 2. The comparison between the distribution of ferritin level before and after logistic transformation. (a) The distribution of ferritin level before logistic transformation (b) The distribution of ferritin level after logistic transformation.

DOI: 10.54254/2753-8818/22/20230984

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).

2.3. Dietary variables

Multiple dietary variables were included in the 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 chosen as the potential 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 3). The median and range of ferritin or transferrin saturation, both numerical data, might differ across different categories of gender, race, and country of birth. In the figure, the median of the ferritin (after logistic transformation) for males was higher than the median of ferritin (after logistic change) for females. This suggested that males generally had a higher ferritin level than females. Additionally, the range of ferritin (after logistic transformation) for females might be more comprehensive than the range of ferritin (after logistic change) for males, indicating more variability in ferritin level for females than for males. Examining the differences in the median and range of the ferritin level (after logistic transformation) across different gender indicated there might be a relationship between gender and ferritin level (after logistic change). Regarding transferrin saturation, the medians of the transferrin saturation of the two genders were also different, indicating a possible relationship between genders and transferrin saturation. However, the race and country of birth medians were approximately the same for the two response variables, indicating few possibilities that race and nation correlated to transferrin saturation and ferritin. (Figure 3. (a2), Figure 3. (a3), Figure 3. (b2), Figure 3. (b3)). Therefore, only the factor of gender was considered in the following linear model.

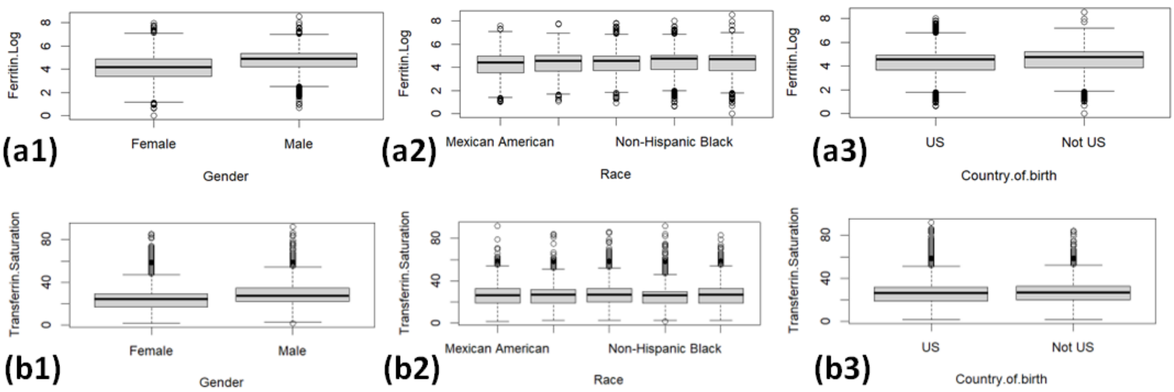

Figure 3. Boxplots were used to examine the relationship between three categorical variables and two numerical response variables.

Besides the categorical variables, other numerical explanatory variables were further selected based on the roughly predicted correlation regarding the two response variables. By arranging the correlation coefficients in descending order, the first 8 variables of each response variable were selected as the final chosen explanatory variables (Table 1).

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

Ferritin.log		Transferrin Saturation	
Variables	Correlation Coefficient	Variables	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	0.071642184
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 [9]. The heat map displays a matrix of correlation coefficients between each pair of nutritional variables, with the strength and direction of the relationship represented by a gradient of colors.

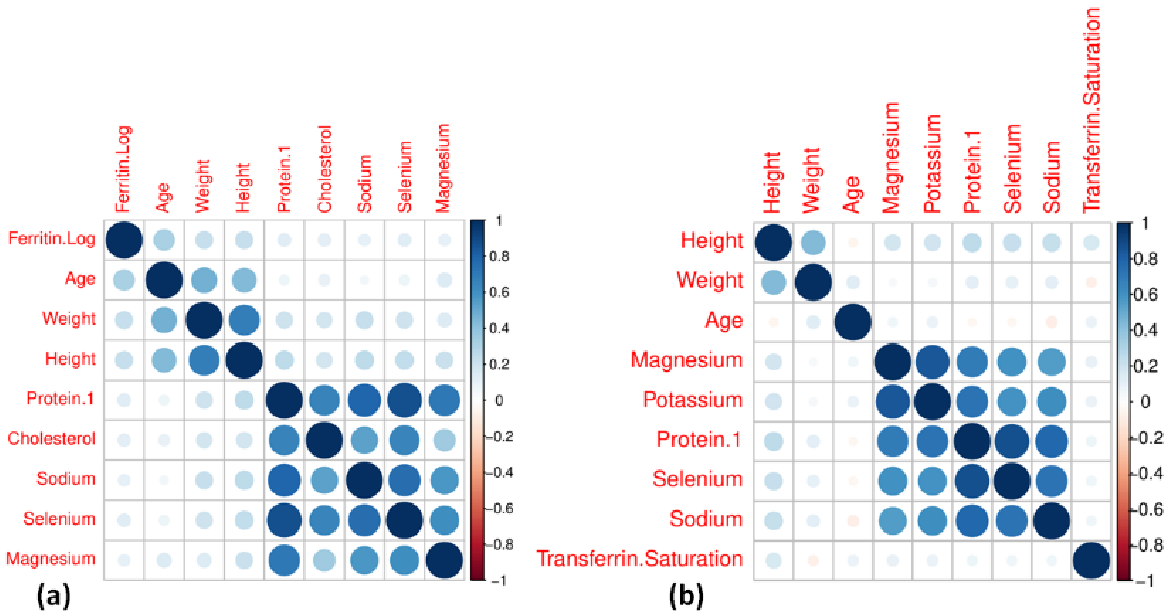

Figure 4. Correlation Coefficient Heat Map of Dietary Variables (a) Dietary Variables with ferritin data (after logistic transformation) (b) Dietary Variables with transferrin saturation.

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 blue represents a positive correlation, and 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, and sodium and selenium are relatively large and shown in relatively dark blue color. These indicate that the increase in protein may lead to an increase in sodium and selenium, and the increase in sodium may lead to a rise in selenium. Except for the relationship between these dietary variables and the response variables (ferritin), there is possibly a relationship between nutritional 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, and sodium and protein are relatively large and shown in a relatively dark blue color. These indicate that the increase in protein may lead to an increase in sodium, potassium, magnesium, and selenium, and the increase in magnesium may lead to an increase in potassium. Except for the relationship between these dietary variables and the response variables (transferrin saturation), there is possibly a relationship between nutritional variables. Therefore, there is a possibility of the existence of multilinear relationships.

2.4. Statistical analysis

To do the linear regression model, conditions were checked. First, there were genuine linear relationships between explanatory variables and response variables. Second, each data point represented one individual, so they were independent. Third, the responses were generally distributed for a particular value of the explanatory variable. 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 variable, 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 multilinear relationships, the variance inflation factor was used to guarantee whether there was necessary to exclude any of the explanatory variables or conduct an orthogonalization. VIF measures whether there was collinearity. Based on the commonly used threshold (VIF>10) for detecting collinearity [10], none of the explanatory variables in this study indicated the presence of collinearity after the optimization of the linear model (Table 2).

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

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

3. 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 and their standard errors. All predictor variables except Protein and Sodium had statistically significant coefficients, p-values less than 0.05 (Table 3). This suggested that Gender, Height, Weight, Age, and Magnesium were substantial 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 of less than 2.2e-16 suggests that the general regression model was statistically significant. The residuals figure showed the distribution of the residuals with no evidence of non-normality or heteroscedasticity (Figure 5).

Table 3. The coefficient table of the linear model of ferritin level

	Estimate	Std.Error	t value	Pr(>| t |)
(Intercept)	3.71E+00	7.04E-02	52.686	<2e-16***
Age	1.36E-02	4.03E-04	33.711	<2e-16***
Weight	1.83E-03	3.83E-04	4.778	1.8e-06***
Height	-1.89E-03	5.56E-04	-3.409	0.000655***

Table 3. (continued).

Protein.1	7.87E-04	3.93E-04	2.005	0.045023*
Cholesterol	1.26E-04	4.89E-05	2.567	0.010274*
Magnesium	-3.06E-04	8.82E-05	-3.47	0.000522***
GenderMale	6.75E-01	1.75E-02	38.498	<2e-16***

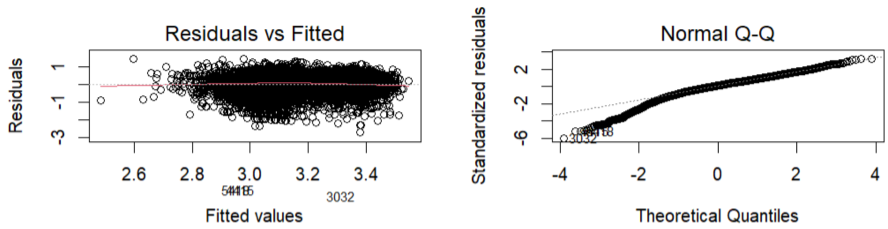

Figure 5. Coefficient table. The residual plots and the Q-Q plot of the linear model of ferritin level indicated the fulfillment of the assumptions of linear regression.

The results showed the investigation of the relationship between log-transformed Ferritin levels and several predictor variables, including Age, Weight, Height, and Protein.1, Cholesterol, Magnesium, and Gender. The coefficients table showed the estimated regression coefficients for each predictor variable and their standard errors. All predictor variables except Height had statistically significant coefficients, with p-values less than 0.05 (Table 4). This suggested Age, Weight, and 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 variation in Ferritin levels. The overall p-value of less than 2.2e-16 told that the general regression model was statistically significant. The residuals figure showed the distribution of the residuals with no evidence of non-normality or heteroscedasticity (Figure 6).

Table 4. The coefficient table of the linear model of transferrin saturation

	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

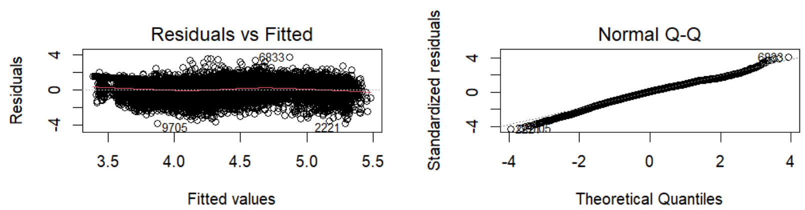

Figure 6. The residual plots and the Q-Q plot of the linear model of transferrin saturation indicated the fulfillment of the assumptions of linear regression.

The variance inflation factor (VIF) values are below 10 for all predictor variables, indicating no significant collinearity among the predictor variables [10].

4. Discussion

Our study first 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, 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 association was also not strong between selected nutrients and transferrin. (r^2=0.2202). Based on these results, people with the risk of iron deficiency can slightly increase their nutrient (Magnesium, Protein.1, Sodium; Protein.1, Cholesterol, Magnesium) intake. Different models, except linear models, can be used to fit this data in the future, 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 on different nutrient intakes was obtained from reliable NHANES data. 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, we cannot generate the cause-and-effect relationship between different nutrients to iron status because of the observational study type.

5. Conclusion

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

References

[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] 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

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

[4] Li, N., Zhao, G., Wu, W., Zhang, M., Liu, W., Chen, Q., & Wang, X. (2020). The Efficacy and Safety of Vitamin C for Iron Supplementation in Adult Patients With Iron Deficiency Anemia. *JAMA Network Open*, 3(11), e2023644. https://doi.org/10.1001/jamanetworkopen.2020.23644

[5] Kondaiah, P., Yaduvanshi, P. S., Sharp, P. A., & Pullakhandam, R. (2019). Iron and Zinc Homeostasis and Interactions: Does Enteric Zinc Excretion Cross-Talk with Intestinal Iron Absorption? *Nutrients*, 11(8), 1885. https://doi.org/10.3390/nu11081885

[6] Wazir, S. M., & Ghobrial, I. (2017). Copper deficiency, a new triad: anemia, leucopenia, and myeloneuropathy. *Journal of Community Hospital Internal Medicine Perspectives,* 7(4), 265–268. https://doi.org/10.1080/20009666.2017.1351289

[7] *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

[8] García-Carretero, R., Holgado-Cuadrado, R., & Barquero-Pérez, Ó. (2021). Assessment of Classification Models and Relevant Features on Nonalcoholic Steatohepatitis Using Random Forest. *Entropy*, 23(6), 763. https://doi.org/10.3390/e23060763

DOI: 10.54254/2753-8818/22/20230984

[9] Haarman, B. C. M. (Benno), Riemersma-Van der Lek, R. F., Nolen, W. A., Mendes, R., Drexhage, H. A., & Burger, H. (2015). Feature-expression heat maps – A new visual method to explore complex associations between two variable sets. Journal of Biomedical Informatics, 53, 156–161. https://doi.org/10.1016/j.jbi.2014.10.003

[10] Variance Inflation Factor - an overview | ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/mathematics/variance-inflation-factor