# The relationship between HDL cholesterol and BMI in US adults: a cross-sectional study based on the NHANES database


## Abstract

This study investigates the relationship between cholesterol level and a variety of demographic, health and lifestyle factors, focusing mainly on BMI effect and interaction between BMI and race. Using data from 2,145 participants aged 20-60 in the NHANES dataset, multiple linear regression analyses were conducted to identify the best-fitted model and significant interaction terms. Based on prior knowledge of factors associated with cholesterol, several predicted have been selected from the dataset. After model selection, the final model retained these predictors included age groups, gender, race, BMI, pulse pressure, diabetes, testosterone levels, depression status, physical activity, alcohol use, smoking, and the interaction between race and BMI. The model was adjusted for heteroscedasticity by transforming the outcome variable ($\sqrt{DirectChol}$). The final model explained 29.8% of the variability in direct cholesterol levels (Adjusted $R^2$=0.2975). BMI shows a negative association with cholesterol levels (β=-0.009, p<0.001), while the interaction term (β=0.003, p=0.003) indicated that the effect of BMI varied by race, with stronger negative associations in non-back participants compared to Black participants. The findings underscore the need for conducting public health interventions that address these factors to improve cholesterol management and reduce health disparities in diverse populations.

## 1. Introduction

The prevalence of obesity is becoming a threat to global public health. Societal changes such as fast food and physical activity limitations are some of the factoring issues. Obesity is associated with various chronic diseases such as cardiovascular diseases, diabetes, osteoarthritis, sleep apnea, and cancer (Wang & Peng, 2011) (1). Understanding the potential factors contributing to high cholesterol is important for comprehending the etiology of cardiovascular risk and the development of effective medicine.

High body mass index (BMI) is a well-known indicator of obesity, and it has been well-studied and linked to lower HDL cholesterol levels. Direct HDL cholesterol, or high-density lipoprotein (HDL) is the measurement of the amount of good cholesterol in blood. Lower levels of HDL are associated with an increase in cardiovascular disease. The study "Cardiovascular risk factors and HDL- cholesterol levels in obesity," P et al. (2), found that the negative correlation between BMI and HDL is only significant in women but not men.

Certain racial groups as well as pre-existing diabetes conditions are often at higher health risk, meaning these demographics can influence HDL levels. A study by Xepapadaki et al. (3) proved that reduced levels of HDL-C and impairment of HDL impact the organs that regulate glucose levels crucial to the development of diabetes.

This study will explore the association between HDL and BMI in adult Americans 20 years and older using the data from NHANES. The rationale behind our study is the significance of scientific evidence finding suggests the association between obesity and cholesterol (2).

## 2. Methodology

Our study population was drawn from a large cross-sectional survey using the NHANES dataset, which represents the non-institutionalized population of the United States. To ensure the completeness and reliability of statistical analyses, the study included participants aged 20 years and older with complete data on Direct HDL cholesterol levels and selected predictors. Participants with missing or invalid values for key variables (Direct HDL levels, predictors, and covariates) were excluded from the study. The NHANES dataset employs a complex, multistage probability sampling design, ensuring representativeness while oversampling low-income populations and minority groups to enhance the precision of subgroup analyses. The resulting study population guarantees high data quality, enabling a comprehensive evaluation of the relationships between health and lifestyle factors and Direct HDL cholesterol levels.

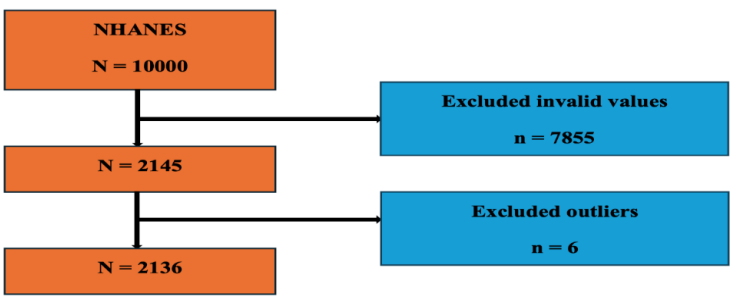


**Figure 1** Flowchart of Study Population Selection

The predictors in this project are categorized into health factors and lifestyle factors. **Health factors** include body mass index (BMI) (1,4,5,6,7,8), pulse pressure difference (diastolic-systolic), diabetes status (3,11,12), and testosterone levels (13,14). **Lifestyle factors** include sleep trouble (15,16), physical activity (17,18), marijuana use (19), hard drug use (20,21), alcohol use (22), and smoking (23).

The outcome variable is **Direct HDL cholesterol levels** (2), measured in mg/dL and treated as a continuous variable in all analyses. HDL levels were obtained from the laboratory test data in the NHANES dataset, providing precise and reliable measurements. HDL cholesterol is a critical biomarker for cardiovascular health, playing a protective role against coronary heart disease. This project explores the relationship between health and lifestyle factors and Direct HDL levels, aiming to better understand how these predictors influence to variations in Direct HDL cholesterol levels.

The confounders in this project included age, gender, and race. **Age** was categorized into three groups: 30-40 years, 40-50 years, and 50-60 years, with individuals under 30 years old serving as the reference group. **Gender** was coded as female and male. **Race** was primarily categorized into Black and others (Mexican, Hispanic, White, etc). These covariates were included in the analysis to adjust for potential confounding factors and to capture demographic variations in the relationship between health and lifestyle predictors and Direct HDL cholesterol levels.

**Descriptive analysis**

According to the National Heart, Lung, and Blood Institute (NHLBI), participants were categorized into two groups based on direct HDL cholesterol: those with HDL cholesterol levels below 1.0 mmol/L were classified as the low HDL cholesterol group, while those with levels ≥1.0 mmol/L were classified as the normal HDL cholesterol group (22). Then a descriptive analysis was used to summarize the characteristics of the study population, with continuous variables expressed as means ± standard deviations and compared using Student's t-test, while categorical variables were presented as sample size (percentages) and compared using Chi-square tests.

To assess whether specific factors might influence the relationship between direct HDL cholesterol and BMI, interaction plots were implemented for factors including age group, gender, race, diabetes. The slope for different age group and gender are the same. It means that there is no interaction effect between BMI and age group as well as BMI and gender. However, the slope for race and diabetes are different. It means that these variables might have an interaction effect with BMI. So, we added two interaction terms in the model (Race*BMI_Center, Diabetes*BMI_Center).

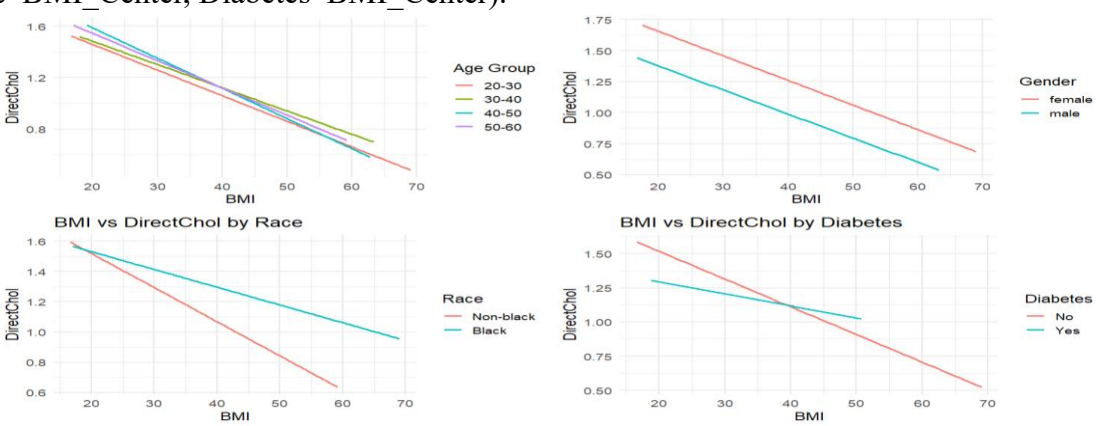


**Figure 2** Interaction Plots of BMI and Direct Cholesterol Levels by Subgroups

**Model selection**

Backward model selection method with p value equals 0.1 was used to prevent overfitting of the model. Age, gender and race were kept in the model regardless of whether they are significant or not.

In this project, we check linearity, independence, constant variance, and normality, respectively to ensure the assumptions of the regression model were satisfied.

To check **Linearity**, we used Partial Regression Plots to visually assess the independent linear relationship between each predictor and the outcome variable (Direct HDL cholesterol levels). The plots demonstrated clear linear relationships for the continue predictors.

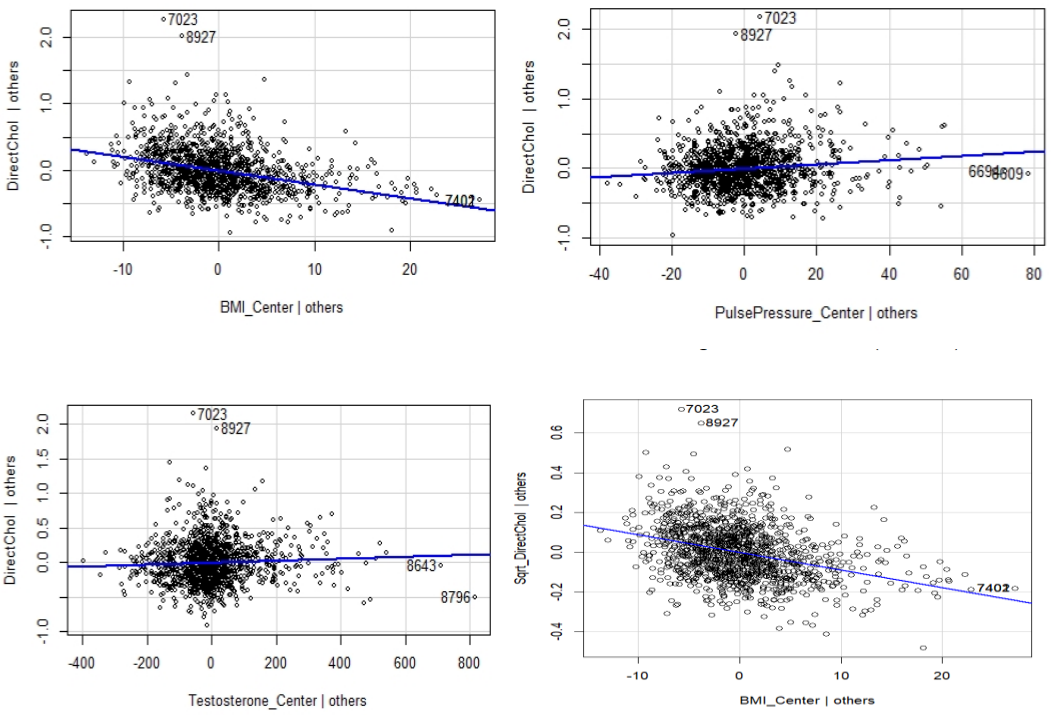


**Figure 3** Partial Regression Plots for Key Predictors of Direct Cholesterol Level

**Independence** was found to be violated using Durbin-Watson Test ($p < 0.001$). However, we can't correct it using current knowledge.

We plotted residuals versus fitted values to check **Constant variance**. The initial results indicated violated. Hence, we transformed the response variable by using the square root (sqrt(Y)) and fitted the model. After transformation, the residuals displayed a relatively uniform spread in the new plot, satisfying the constant variance assumption.

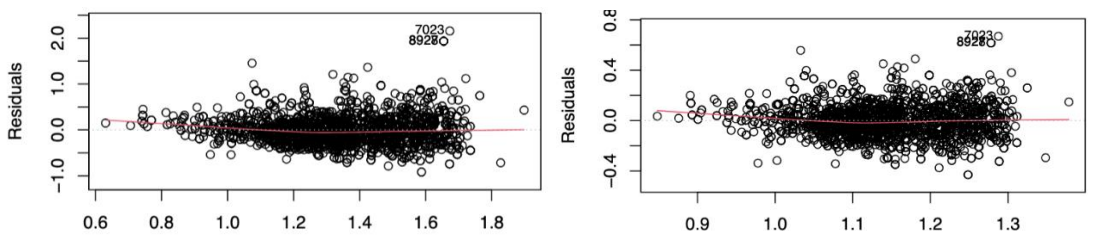


**Figure 4** Residuals vs Fitted Values Before and After Transformation

Finally, to evaluate **Normality**, we plotted the histogram of residuals and Q-Q plot. The histogram showed a symmetric, bell-shaped distribution, while the Q-Q Plot revealed that the residuals aligned closely with the diagonal line, with only minimal deviations. These results confirmed that the normality assumption was satisfied.

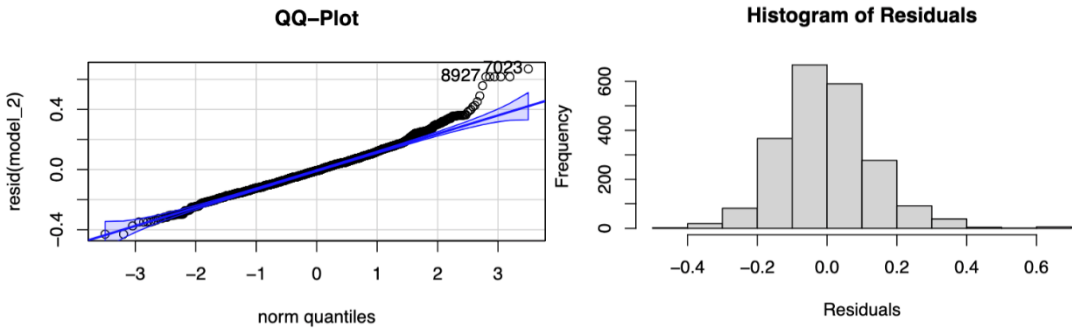


**Figure 5** Diagnostic Plots for Normality of Residuals

In this part, we first plotted Cook's distance to identify influential points in the model that could significantly affect the regression results. The plot revealed that points 7040 and 9169 had notably high Cook's distances, indicating that these were likely high-leverage points with a strong influence on the model. Most of the remaining points had Cook's distances below the threshold, suggesting a minimal impact on the model. Next, we created the influential points analysis table to further investigate and locate the potential outliers. Points 7040, 9169, 9170, 6661, 8517, and 8796 were identified as having high Cook's distances. Finally, we deleted these points to reduce their impact on model fitting and inference. This step improved the stability of accuracy of the model, ensuring more reliable results.

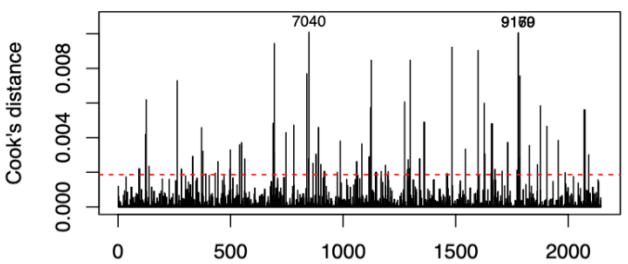


**Figure 6** Cook's Distance Plot for Identifying Influential Points

All statistical analyses were conducted using R (http://www.R-project.org, The R Foundation), with all statistical significance set at a two-sided p-value of <0.05.

**Final model**

The final model is specified as a multiple linear regression predicting the square root of direct cholesterol (Sqrt_DirectChol) as the outcome variable. The formula is expressed as:

$$Sqrt_{DirectChol} = \beta 0 + \beta 1(Age\_Group) + \beta 2(Gender) + \beta 3(Race) + \beta 4(BMI\_Center) + \beta 5(PulsePressure\_Center) + \beta 6(Diabetes) + \beta 7(Testosterone\_Center) + \beta 8(Depressed) + \beta 9(PhysActive) + \beta 10(Alcohol12PlusYr) + \beta 11(Smoke100) + \beta 12(Race \times BMI\_Center) + \epsilon$$

This model includes main effects for demographic, health-related, and lifestyle predictors, as well as an interaction term (Race × BMI_Center) to investigate whether BMI's relationship with cholesterol varies by race. Continuous variables are centered to improve interpretability, and categorical variables are factorized to enable group comparisons.

## 3. Results

**Results of descriptive analysis**

The overall mean direct HDL cholesterol was 1.34 ± 0.38 mmol/L, the mean BMI was 28.66 ± 6.38 kg/m2, and the mean age was 39.69 ± 11.67 years. Participants with low HDL

cholesterol had significantly higher BMI (31.62 ± 6.47 vs. 28.05 ± 6.19 kg/m2, $p < 0.001$) and higher testosterone (285.47 ± 201.68 vs. 220.71 ± 232.59 ng/dL, $p < 0.001$). However, those with low HDL cholesterol had lower pulse pressure (43.99 ± 12.14 vs. 46.04 ± 12.70 mm Hg, $p = 0.004$). Low HDL cholesterol was observed at a significantly higher rate in male (77.17% vs. 48.51%, $p < 0.001$) and diabetes patients (8.97% vs. 5.68%, $p = 0.024$). Low HDL cholesterol was observed at a significantly lower rate in black populations (6.79% vs. 11.71%, $p = 0.008$) and participant who regularly engage in physical exercise (52.99% vs. 60.95%, $p = 0.006$). More detailed comparison results can be found in Table 1.

**Table 1** Comparison of Demographic, Clinical, and Lifestyle Characteristics Between Participants with Low and Normal Direct HDL Cholesterol Levels

| Characteristics | N | Low, N=368 | Normal, N=1777 | P-value |
|---|---|---|---|---|
| Age Group | 2145 | | | 0.585 |
| 20-30 | | 106(28.80%) | 502(28.25%) | |
| 30-40 | | 75(20.38%) | 408(22.96%) | |
| 40-50 | | 108(29.35%) | 471(26.51%) | |
| 50-60 | | 79(21.47%) | 396(22.28%) | |
| Gender | 2145 | 284(77.17%) | 862(48.51%) | <0.001 |
| Race | 2145 | 25(6.79%) | 208(11.71%) | 0.008 |
| BMI | 2145 | 31.62±6.47 | 28.05±6.19 | <0.001 |
| Pulse Pressure | 2145 | 43.99±12.14 | 46.04±12.70 | 0.004 |
| Diabetes | 2145 | 33(8.97%) | 101(5.68%) | 0.024 |
| Sleep Trouble | 2145 | 97(26.36%) | 471(26.51%) | 1 |
| Testosterone | 2145 | 285.47±201.68 | 220.71±232.59 | <0.001 |
| Depressed | 2145 | | | 0.284 |
| None | | 301(81.79%) | 1403(78.95%) | |
| Several | | 49(13.32%) | 249(14.01%) | |
| Most | | 18(4.89%) | 125(7.03%) | |
| Physical Activity | 2145 | 195(52.99%) | 1083(60.95%) | 0.006 |
| Marijuana | 2145 | 214(58.15%) | 1096(61.68%) | 0.229 |
| Hard Drugs | 2145 | 79(21.47%) | 367(20.65%) | 0.780 |
| Alcohol | 2145 | 299(81.25%) | 1488(83.74%) | 0.277 |
| Smoke | 2145 | 160(43.48%) | 748(42.09%) | 0.666 |

**Results of final model**

The final regression model examined the associations between direct cholesterol levels and a range of demographic, health, and lifestyle predictors using data from 2,139 participants after excluding missing values. Predictors included age groups, gender, race, BMI, pulse pressure, diabetes, testosterone levels, depression status, physical activity, alcohol use, smoking, and an interaction term between race and BMI. The model had a residual standard error of 0.1312 with 2,123 degrees of freedom, and an adjusted R-squared of 0.2975, explaining approximately 29.8% of the variability in cholesterol levels. Table 2 details the regression coefficients, and Figure 7 illustrates standardized coefficients with confidence intervals.

BMI showed a significant negative association with cholesterol levels ($\beta = -0.009$, $p < 0.001$), with a narrow confidence interval indicating its robust effect (Figure 7). The interaction term between race and BMI ($\beta = 0.003$, $p = 0.003$) revealed that the effect of BMI on cholesterol differed by race.

Gender had the strongest effect, with males exhibiting significantly lower cholesterol levels than females ($\beta = -0.148$, $p < 0.001$), as reflected in Figure 7, where the coefficient for gender is positioned furthest from the null line, underscoring its importance as a determinant of cholesterol levels. Age was positively associated with cholesterol, particularly in the 40-50 age group ($\beta = 0.046$, $p < 0.001$). Race was also significant, with Black participants showing higher cholesterol levels than Non-Black participants ($\beta = 0.027$, $p = 0.009$).

Among lifestyle predictors, alcohol use was positively associated with cholesterol levels ($\beta = 0.025$, $p = 0.002$), while smoking showed a significant negative association ($\beta = -0.017$, $p = 0.005$). Physical activity had a weak positive association ($\beta = 0.011$, $p = 0.071$) but was not statistically significant. These findings align with Figure 7, where alcohol and smoking coefficients have confidence intervals that exclude zero, while physical activity overlaps with the null.

**Table 2** Regression Coefficients, Standard Errors, t-values, and P-values for the Predictors of Final Model

| Predictor | Coefficient | Std. Error | t-value | P-value |
|---|---|---|---|---|
| Intercept | 1.196 | 0.010 | 114.709 | < 2e-16 |
| Age_Group30-40 | 0.028 | 0.008 | 3.416 | 0.001 |
| Age_Group40-50 | 0.046 | 0.008 | 5.873 | < 5e-09 |
| Age_Group50-60 | 0.040 | 0.008 | 4.787 | < 2e-06 |
| Gender (Male) | -0.148 | 0.011 | -13.306 | < 2e-16 |
| Race (Black) | 0.027 | 0.010 | 2.633 | 0.009 |
| BMI_Center | -0.009 | 0.001 | -16.806 | < 2e-16 |
| PulsePressure_Center | 0.001 | 0.000 | 4.976 | < 7e-07 |
| Diabetes (Yes) | -0.031 | 0.012 | -2.547 | 0.011 |
| Testosterone_Center | 0.000 | 0.000 | 3.195 | 0.001 |
| Depressed (Several) | -0.019 | 0.008 | -2.306 | 0.021 |
| Depressed (Most) | 0.028 | 0.012 | 2.396 | 0.017 |
| PhysActive (Yes) | 0.011 | 0.006 | 1.804 | 0.071 |
| Alcohol12PlusYr (Yes) | 0.025 | 0.008 | 3.120 | 0.002 |
| Smoke100 (Yes) | -0.017 | 0.006 | -2.787 | 0.005 |
| Race × BMI_Center | 0.003 | 0.001 | 2.946 | 0.003 |

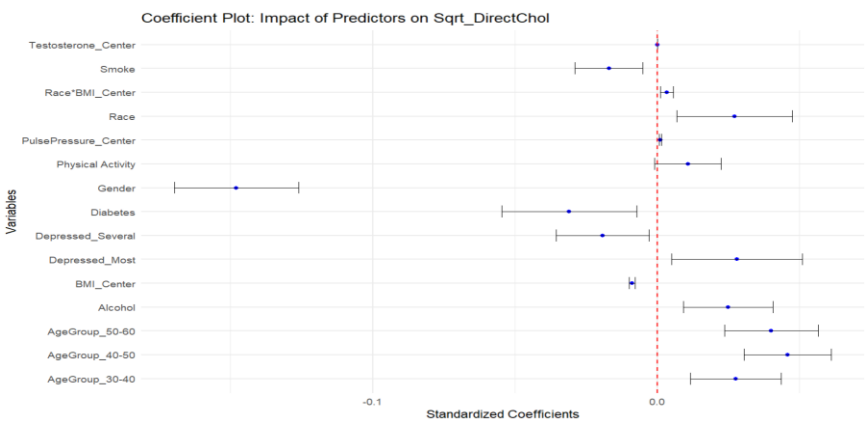


**Figure 7** Coefficient Plot about Impact of Predictors on direct cholesterol levels

## The Association of BMI and Direct Cholesterol Levels

The relationship between BMI and direct cholesterol levels was further examined using a scatter plot (Figure 8) and partial regression plots (Figure 3). Both visualizations confirm the

significant negative association observed in the regression model. Figure 8 shows a clear downward trend, where higher centered BMI values are associated with lower direct cholesterol levels. The fitted regression line (red) emphasizes this negative relationship, with the shaded confidence interval indicating the precision of the estimates. The data points demonstrate variability, but the trend remains robust across the range of BMI values, supporting the strength of the relationship.

The last plot of Figure 3 isolates the effect of centered BMI on direct cholesterol levels, adjusting for other predictors in the model. The negative slope of the partial regression line corroborates the findings from Figure 8. Additionally, the plot highlights several influential observations (e.g., points 7023, 7402, and 8927), which may have a stronger impact on the overall regression model. Despite these influential data points, the negative relationship between BMI and direct cholesterol levels remains consistent and statistically significant.

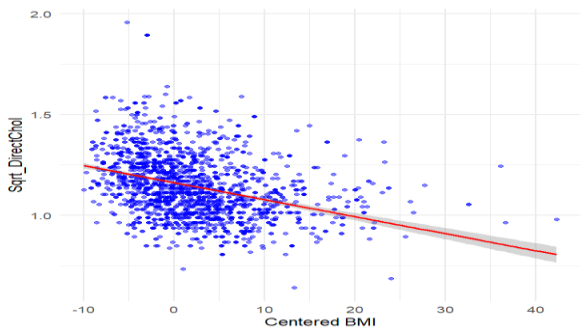


**Figure 8** Scatter Plot of Centered BMI vs. Direct Cholesterol Levels with Fitted Regression Line

## Interaction Effect of BMI and Race on Direct Cholesterol Levels

The inclusion of the interaction term (Race × BMI_Center) in the model revealed significant differences in the relationship between BMI and direct cholesterol levels across racial groups ($\beta = 0.003$, $p = 0.003$). This interaction indicates that the slope of the association between BMI and cholesterol varies between Black and Non-Black participants.

When stratifying the analysis by race, distinct patterns emerge. For Black participants, BMI had a smaller negative association with cholesterol levels ($\beta = -0.005$, $p < 0.001$), suggesting that increases in BMI correspond to a modest reduction in cholesterol. Conversely, for Non-Black participants, BMI had a stronger negative association ($\beta = -0.009$, $p < 0.001$). These findings highlight that while BMI is inversely related to cholesterol levels in both groups, the effect size is more pronounced in Non-Black participants. The interaction plot (Figure 9) further illustrates these differences. For Black participants, the slope of the regression line is less steep compared to Non-Black participants, indicating a weaker relationship between BMI and cholesterol. The shaded regions represent the 95% confidence intervals, which are wider for Black participants, reflecting greater variability in the data.

Furthermore, the output of stratified analyses revealed a slightly higher adjusted R-squared for Non-Black participants (0.3006) compared to Black participants (0.2535), suggesting that the predictors explain more variability in cholesterol levels for Non-Black individuals. Also, the residual standard error was lower for Black participants (0.1299) than for Non-Black participants (0.1313), indicating less residual variability among Black participants.

**Table 3** Regression Estimates, Standard Errors and P-values by Race

| Predictor | Black Coefficient | Black Std. Error | Black P-value | Non-Black Coefficient | Non-Black Std. Error | Non-Black P-value |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| Intercept | 1.232 | 0.028 | < 0.001 | 1.195 | 0.011 | < 0.001 |
| Age_Group30-40 | 0.040 | 0.024 | 0.099 | 0.025 | 0.009 | 0.004 |
| Age_Group40-50 | 0.073 | 0.025 | 0.004 | 0.041 | 0.008 | < 0.001 |
| Age_Group50-60 | 0.094 | 0.025 | < 0.001 | 0.031 | 0.009 | < 0.001 |
| Gender (Male) | -0.180 | 0.036 | < 0.001 | -0.145 | 0.012 | < 0.001 |
| BMI_Center | -0.005 | 0.001 | < 0.001 | -0.009 | 0.001 | < 0.001 |
| PulsePressure_Center | 0.001 | 0.001 | 0.186 | 0.001 | 0.000 | < 0.001 |
| Diabetes (Yes) | -0.039 | 0.033 | 0.234 | -0.029 | 0.013 | 0.029 |
| Testosterone_Center | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.007 |
| Depressed (Several) | -0.049 | 0.027 | 0.068 | -0.015 | 0.009 | 0.099 |
| Depressed (Most) | -0.047 | 0.041 | 0.256 | 0.037 | 0.012 | 0.003 |
| PhysActive (Yes) | 0.009 | 0.018 | 0.614 | 0.010 | 0.006 | 0.101 |
| Alcohol12PlusYr | 0.017 | 0.020 | 0.413 | 0.028 | 0.009 | 0.001 |
| Smoke100 (Yes) | -0.026 | 0.021 | 0.231 | -0.016 | 0.006 | 0.011 |

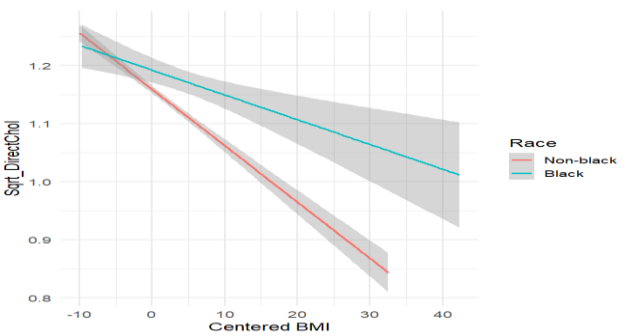


**Figure 9** Interaction Effect of Race and Centered BMI on Direct Cholesterol Levels

## 4. Discussion and Conclusion

This study examined the relationship between BMI as a predictor and direct cholesterol (DirectChol) in the NHANES dataset. After addressing potential issues such as identifying interaction term, collinearity, heteroscedasticity, backward model selection to find the best inference variables and influential points, the final model identified significant predictors, including Race, BMI_Center, PulsePressure_Center, and their interactions. The interaction between Race and BMI_Center revealed that the relationship between BMI and cholesterol differs across racial groups. In the final model, the corresponding R is 0.2975. The analysis indicates that BMI_Center has a small but negative effect on DirectChol for the general population, with a coefficient of -0.009. However, this relationship is moderated by race, as evidenced by an interaction term (+0.003) in the Black subgroup. This reduces the negative impact of BMI_Center on cholesterol levels within the Black population compared to the Non-Black group. These findings highlight that the relationship between BMI and cholesterol is not only minor but also influenced by demographic factors such as race. Other predictors, such as age, gender, and physical activity, likely play a more significant role in determining cholesterol levels.

The significant interaction between Race and BMI_Center suggests that the effect of BMI on cholesterol levels is moderated by race. Specifically, the results indicate that the association

between BMI and cholesterol is more negative and stronger among Non-Black individuals than Black individuals. This may reflect underlying physiological or genetic differences between racial groups. For instance, previous studies have shown that Black individuals tend to have more favorable profiles, such as higher levels of HDL cholesterol (the "good" cholesterol), even when BMI is elevated as observed in this study. Publication has already reported that the prevalence of hypertension is higher in blacks than in the other race (23,24). Additionally, differences in fat distribution patterns, such as higher visceral adiposity in Non-Black individuals, might contribute to the stronger relationship between BMI and cholesterol in this group. Several studies have reported this finding of less susceptibility of blacks to visceral adipose tissue deposition than the other race (25).

The adjustment by transforming the outcome variable (DirectChol) to square terms improved model fit. This improvement ensured more reliable estimates and highlighted consistent relationships, such as the negative association between BMI and cholesterol levels. The scatter plot and partial regression plot further confirmed the negative association between BMI and cholesterol, even when accounting for other predictors. This relationship aligns with studies indicating that increasing BMI can lead to decreased HDL cholesterol (26). Predictors such as gender and age also played a prominent role in determining cholesterol levels. Males exhibited significantly lower cholesterol levels compared to females, which might be explained by hormonal differences, as estrogen is known to elevate HDL cholesterol levels, leading to higher HDL cholesterol in women (27).

Lifestyle factors, including alcohol consumption and smoking, also significantly influenced cholesterol levels. Alcohol was positively associated with cholesterol, which is consistent with evidence suggesting that moderate alcohol intake can increase HDL cholesterol due to increasing HDL cholesterol (HDL-C) levels and cholesterol efflux capacity (CEC) caused by moderate alcohol consumption (28). On the other hand, smoking was negatively associated with cholesterol levels, because cigarette smoking can mediate changes in the composition of blood, altering levels of lipids (29). While physical activity showed a weak positive association with cholesterol levels in this study, its effect was not statistically significant, which could be due to insufficient data in measurement or the presence of confounding variables.

The involvement of stratified analysis provided further insights into racial disparities in cholesterol management. For Black participants, the relatively lower negative association between BMI and cholesterol suggests that interventions targeting BMI reduction may have different impacts across racial groups. This emphasizes the importance of stratified public health strategies and personalized cholesterol management plans.

The limitations of this study include several important aspects. First, while the NHANES dataset is comprehensive, it may not fully represent all demographic groups, which limits the generalizability of the findings to populations outside the U.S. Second, some relevant variables, such as diet or genetic predisposition, were not included in the dataset. Lastly, the study primarily relied on linear terms to model the relationship between predictors and cholesterol levels. The exclusion of quadratic terms, such as $BMI^2$, might have restricted the model's ability to capture potential nonlinear relationships. Future work might focus on investigating non-linear relationships between predictors such as BMI and cholesterol levels. Incorporating higher-order terms in future analyses could enhance model accuracy and predictive power. And perform detailed subgroup analyses, particularly stratifying by age and other discrete variables trying to best the best stratification.

## References


1. Wang, H., & Peng, D. (2011). New insights into the mechanism of low high-density lipoprotein cholesterol in obesity. *Lipids in Health and Disease*, *10*(1), 176. https://doi.org/10.1186/1476-511x-10-176
2. P, B., M, B., Jörgens, C, D., E, C., Fa, G., & H, Z. (1981). Cardiovascular risk factors and HDL-cholesterol levels in obesity. *PubMed*, *5*(1), 1–10. https://pubmed.ncbi.nlm.nih.gov/7012054
3. Xepapadaki, E., Nikdima, I., Sagiadinou, E. C., Zvintzou, E., & Kypreos, K. E. (2021). HDL and type 2 diabetes: the chicken or the egg? *Diabetologia*, *64*(9), 1917–1926. https://doi.org/10.1007/s00125-021-05509-0
4. McGillicuddy, F. C., Reilly, M. P., & Rader, D. J. (2011). Adipose Modulation of High-Density Lipoprotein Cholesterol. *Circulation*, *124*(15), 1602–1605. https://doi.org/10.1161/circulationaha.111.058453
5. Mooradian, A. D., Haas, M. J., Wehmeier, K. R., & Wong, N. C. (2008). Obesity-related Changes in High-density Lipoprotein Metabolism. *Obesity*, *16*(6), 1152–1160. https://doi.org/10.1038/oby.2008.202
6. Salazar, M. R., Carbajal, H. A., Espeche, W. G., Aizpurúa, M., Marillet, A. G., Sisnieguez, C. E. L., Sisnieguez, B. C. L., Stavile, R. N., March, C. E., & Reaven, G. M. (2016a). Use of the Triglyceride/High-Density Lipoprotein Cholesterol Ratio to Identify Cardiometabolic Risk: Impact of Obesity? *Journal of Investigative Medicine*, *65*(2), 323–327. https://doi.org/10.1136/jim-2016-000248
7. Salazar, M. R., Carbajal, H. A., Espeche, W. G., Aizpurúa, M., Marillet, A. G., Sisnieguez, C. E. L., Sisnieguez, B. C. L., Stavile, R. N., March, C. E., & Reaven, G. M. (2016b). Use of the Triglyceride/High-Density Lipoprotein Cholesterol Ratio to Identify Cardiometabolic Risk: Impact of Obesity? *Journal of Investigative Medicine*, *65*(2), 323–327. https://doi.org/10.1136/jim-2016-000248
8. Salazar, M. R., Carbajal, H. A., Espeche, W. G., Aizpurúa, M., Marillet, A. G., Sisnieguez, C. E. L., Sisnieguez, B. C. L., Stavile, R. N., March, C. E., & Reaven, G. M. (2016c). Use of the Triglyceride/High-Density Lipoprotein Cholesterol Ratio to Identify Cardiometabolic Risk: Impact of Obesity? *Journal of Investigative Medicine*, *65*(2), 323–327. https://doi.org/10.1136/jim-2016-000248
9. Morgantini, C., Natali, A., Boldrini, B., Imaizumi, S., Navab, M., Fogelman, A. M., Ferrannini, E., & Reddy, S. T. (2011). Anti-inflammatory and Antioxidant Properties of HDLs Are Impaired in Type 2 Diabetes. *Diabetes*, *60*(10), 2617–2623. https://doi.org/10.2337/db11-0378
10. Yan, Z., Xu, Y., Li, K., & Liu, L. (2024). Association between high-density lipoprotein cholesterol and type 2 diabetes mellitus: dual evidence from NHANES database and Mendelian randomization analysis. *Frontiers in Endocrinology*, *15*. https://doi.org/10.3389/fendo.2024.1272314
11. Gomes, G. K., De Branco, F. M. S., Santos, H. O., Pereira, J. L., Orsatti, F. L., & De Oliveira, E. P. (2023). Cholesterol intake and serum total cholesterol levels are not associated with total testosterone levels in men: a cross-sectional study from NHANES 2013–2014. *Lipids in Health and Disease*, *22*(1). https://doi.org/10.1186/s12944-023-01928-7
12. Mei, Y., Chen, Y., Wang, X., Xu, R., & Feng, X. (2024). The relationship between remnant cholesterol and the risk of testosterone deficiency in US adults: a cross-sectional study based on

the NHANES database. *Frontiers in Endocrinology*, *15*. https://doi.org/10.3389/fendo.2024.1458193

13. Cai, Y., Zhou, Z., & Zeng, Y. (2024). Association between non-high-density lipoprotein cholesterol to high-density lipoprotein cholesterol ratio (NHHR) and sleep disorders in US adults: NHANES 2005 to 2016. *Medicine*, *103*(26), e38748. https://doi.org/10.1097/md.0000000000038748
14. Pan, X., Zhang, X., Wu, X., Zhao, Y., Li, Y., Chen, Z., Hu, Y., & Cao, X. (2024). Association between non-high-density lipoprotein cholesterol to high-density lipoprotein cholesterol ratio and obstructive sleep apnea: a cross-sectional study from NHANES. *Lipids in Health and Disease*, *23*(1). https://doi.org/10.1186/s12944-024-02195-w
15. Blodgett, J. M., Ahmadi, M. N., Atkin, A. J., Chastin, S., Chan, H., Suorsa, K., Bakker, E. A., Hettiarcachchi, P., Johansson, P. J., Sherar, L. B., Rangul, V., Pulsford, R. M., Mishra, G., Eijsvogels, T. M. H., Stenholm, S., Hughes, A. D., Teixeira-Pinto, A. M., Ekelund, U., Lee, I. M., . . . Thijssen, D. H. J. (2023). Device-measured physical activity and cardiometabolic health: the Prospective Physical Activity, Sitting, and Sleep (ProPASS) consortium. *European Heart Journal*. https://doi.org/10.1093/eurheartj/ehad717
16. Chen, J., Luo, Q., Su, Y., Wang, J., Fang, Z., & Luo, F. (2023). Effects of physical activity on the levels of remnant cholesterol: A population-based study. *Journal of Cellular and Molecular Medicine*. https://doi.org/10.1111/jcmm.18062
17. Panee, J., Pomozi, V., Franke, A. A., Saux, O. L., & Gerschenson, M. (2020). Chronic marijuana use moderates the correlations of serum cholesterol with systemic mitochondrial function and fluid cognition. *Mitochondrion*, *52*, 135–143. https://doi.org/10.1016/j.mito.2020.03.006
18. Cotto, B., Natarajaseenivasan, K., Ferrero, K., Wesley, L., Sayre, M., & Langford, D. (2018). Cocaine and HIV-1 Tat disrupt cholesterol homeostasis in astrocytes: Implications for HIV-associated neurocognitive disorders in cocaine user patients. *Glia*, *66*(4), 889–902. https://doi.org/10.1002/glia.23291
19. Escobar, M., Scherer, J. N., Soares, C. M., Guimarães, L. S. P., Hagen, M. E. K., Von Diemen, L., & Pechansky, F. (2018). Active Brazilian crack cocaine users: nutritional, anthropometric, and drug use profiles. *Brazilian Journal of Psychiatry*, *40*(4), 354–360. https://doi.org/10.1590/1516-4446-2017-2409
20. Li, H., & Xia, N. (2023). Alcohol and the vasculature: a love-hate relationship? *Pflügers Archiv - European Journal of Physiology*, *475*(7), 867–875. https://doi.org/10.1007/s00424-023-02818-8
21. Ambrose, J. A., & Barua, R. S. (2004). The pathophysiology of cigarette smoking and cardiovascular disease. *Journal of the American College of Cardiology*, *43*(10), 1731–1737. https://doi.org/10.1016/j.jacc.2003.12.047
22. Grundy, S. M., Stone, N. J., Bailey, A. L., Beam, C., Birtcher, K. K., Blumenthal, R. S., Braun, L. T., De Ferranti, S., Faiella-Tommasino, J., Forman, D. E., Goldberg, R., Heidenreich, P. A., Hlatky, M. A., Jones, D. W., Lloyd-Jones, D., Lopez-Pajares, N., Ndumele, C. E., Orringer, C. E., Peralta, C. A., . . . Yeboah, J. (2019). 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*, *139*(25). https://doi.org/10.1161/cir.0000000000000625

23. Arnett, D. K., Blumenthal, R. S., Albert, M. A., et al. (2019). "2019 ACC/AHA guideline on the primary prevention of cardiovascular disease." Circulation. DOI:10.1161/CIR.0000000000000678.
24. Lovejoy, J. C., Smith, S. R., Rood, J. C. (2001). "Comparison of regional fat distribution and health risk factors in middle-aged White and African American women: the Healthy Transitions Study." Obesity Research. DOI:10.1038/oby.2001.38.
25. Després, J. P. (2012). "Body fat distribution and risk of cardiovascular disease: an update." Circulation. DOI:10.1161/CIR.0b013e3182539482.
26. Dansinger, M., Williams, P. T., Superko, H. R., Asztalos, B. F., & Schaefer, E. J. (2019). Effects of weight change on HDL-cholesterol and its subfractions in over 28,000 men and women. Journal of Clinical Lipidology, 13(2), 308–316. https://doi.org/10.1016/j.jacl.2018.12.001.
27. Krauss, R.M., Lindgren, F.T., Wingerd, J. et al. Effects of estrogens and progestins on high density lipoproteins. Lipids 14, 113–118 (1979). https://doi.org/10.1007/BF02533579
28. Badia, R. R., Pradhan, R. V., Ayers, C. R., Chandra, A., & Rohatgi, A. (2023). The Relationship of Alcohol Consumption and HDL Metabolism in the Multiethnic Dallas Heart Study. Journal of Clinical Lipidology, 17(1), 72–81.
29. Craig, W. Y., Palomaki, G. E., Haddow, J. E. (1989). "Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data." BMJ. DOI:10.1136/bmj.298.6676.784.

Xingyou Zhou - Data cleaning, Model diagnostics, descriptive analysis

Elizabeth Li - Introduction, References

Junyi Zhang - Study population, Variables, Diagnostics

Jianyi Xu – Variables of Final model, Results, Figure

Brian Ge - Abstract, Conclusion, Discussion, Limitation