

Association of Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) with Anthropometric, Metabolic, and Lifestyle Risk Factors: A Cross-Sectional Analysis of NHANES 2017-2018 Dataset

Wei Chieh Chen

2025-02-20

Abstract

Background

Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) is a newly proposed diagnostic term that replaces Nonalcoholic Fatty Liver Disease (NAFLD) to better reflect its metabolic underpinnings. MAFLD is characterized by hepatic steatosis in the presence of metabolic abnormalities such as obesity, diabetes, and dyslipidemia. Given the increasing prevalence of metabolic syndrome, understanding the demographic, anthropometric, and metabolic determinants of MAFLD is crucial for developing targeted prevention and treatment strategies.

Methods

This study utilized data from the 2017-2018 National Health and Nutrition Examination Survey (NHANES) to examine associations between MAFLD and demographic, anthropometric, metabolic, and lifestyle factors. MAFLD was diagnosed based on hepatic steatosis identified by Controlled Attenuation Parameter (CAP) measurements and the presence of metabolic risk factors. Logistic regression models were used to identify predictors of MAFLD, while multiple linear regression analysis was conducted to examine determinants of CAP scores.

Results

A total of **4,477** participants were included in the final analysis, with **1,962 (43.8%)** classified as having MAFLD. Logistic regression analysis identified **Age** as a predictor of MAFLD ($OR = 1.02, 95\%CI : 1.00 \sim 1.04, p = 0.014$); **Waist circumference** showed a strong association ($OR = 1.09, 95\%CI : 1.08 \sim 1.09, p < 0.001$); and **Triglycerides** were significantly related to MAFLD ($OR = 1.01, 95\%CI : 1.01 \sim 1.01, p < 0.001$). CAP scores were significantly higher in individuals with MAFLD (323.9 ± 35.8 dB/m) compared to those without (220.8 ± 38.2 dB/m, $p < 0.001$).

Multiple linear regression analysis revealed:

$$CAP \sim 1.673 \times \text{Waist Circumference} + 0.091 \times \text{Triglycerides} + 0.584 \times \text{ALT}$$

where waist circumference ($\beta = 1.673, p < 0.001$), triglycerides ($\beta = 0.091, p < 0.001$), and ALT ($\beta = 0.584, p < 0.001$) were the strongest determinants of CAP scores. The overall model explained **42.4% of CAP variance** (Adjusted $R^2 = 0.4236, p < 0.001$).

A Welch Two-Sample t -test indicated that males exhibited significantly higher CAP scores than females ($t = -8.5372, p < 2.2 \times 10^{-16}$). These results highlight the **importance of central adiposity, lipid metabolism, and metabolic dysfunction** in MAFLD development. Given the strong association between

waist circumference, triglycerides, and CAP scores, these factors may serve as better screening indicators than BMI.

Conclusion

Our findings highlight the role of central adiposity, lipid metabolism, and metabolic dysfunction in MAFLD development. The strong association between waist circumference, triglycerides, and CAP scores suggests that these factors may serve as better screening indicators than BMI. The study underscores the importance of early metabolic interventions to mitigate the growing burden of MAFLD and its associated complications, including cardiovascular disease and liver fibrosis.

Introduction

Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) is a newly proposed diagnostic term for fatty liver disease related to metabolic dysregulation, replacing the traditional classification of Nonalcoholic Fatty Liver Disease (NAFLD). The shift to MAFLD reflects a broader understanding of the metabolic underpinnings of hepatic steatosis and allows for a more inclusive and precise diagnosis without requiring the exclusion of other liver diseases. This reclassification has significant implications for epidemiological research, clinical practice, and public health policies, as it enables the identification of high-risk individuals who may benefit from early interventions.

MAFLD is characterized by hepatic steatosis in the presence of at least one metabolic abnormality, such as obesity, type 2 diabetes mellitus, insulin resistance, or dyslipidemia. The condition is associated with an increased risk of adverse hepatic outcomes, including liver fibrosis, cirrhosis, and hepatocellular carcinoma (HCC), as well as extrahepatic complications such as cardiovascular disease and chronic kidney disease. Given the rising global prevalence of metabolic syndrome and obesity, understanding the demographic, anthropometric, and metabolic determinants of MAFLD is critical for developing targeted prevention and treatment strategies.

This study utilizes data from the **National Health and Nutrition Examination Survey (NHANES) 2017-2018**, a nationally representative dataset, to investigate the association between MAFLD and a range of demographic, anthropometric, metabolic, and lifestyle factors. Specifically, we aim to:

1. Identify key demographic and metabolic predictors of MAFLD using logistic regression models.
2. Examine the determinants of hepatic fat accumulation, as measured by **Controlled Attenuation Parameter (CAP) scores**, through multiple linear regression analysis.
3. Explore associations between MAFLD and clinical factors such as hypertension, lipid abnormalities, and smoking status.

By utilizing NHANES data, this study provides a comprehensive analysis of MAFLD risk factors in a diverse population, offering insights that may inform clinical screening protocols and public health interventions aimed at reducing the burden of metabolic liver disease.

Methods

Study Population

The National Health and Nutrition Examination Survey (NHANES) is an ongoing cross-sectional study conducted by the National Center for Health Statistics (NCHS) under the Centers for Disease Control and Prevention (CDC). The NHANES program has been approved by the NCHS Ethics Review Board, and all participants provided written informed consent. The dataset and study details are publicly available at <https://www.cdc.gov/nchs/nhanes/>. This analysis utilized data from the 2017-2018 NHANES cycle.

The study included individuals ≥ 18 years of age with complete liver elastography measurements. Participants with incomplete test results or those deemed ineligible for hepatic elastography were excluded. Additionally, individuals with missing data on body mass index (BMI), red blood cell (RBC) count, or critical biochemical

parameters were removed from the analysis. After applying these criteria, the final analytical sample consisted of 4,477 participants.

As this study was based on publicly available de-identified data, additional ethical review and approval were not required, in accordance with local regulations and institutional guidelines.

Definition of Metabolic dysfunction-associated fatty liver disease (MAFLD)

In 2020, the term Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) was introduced to describe fatty liver disease linked to metabolic issues. This replaced the previous term, Nonalcoholic Fatty Liver Disease (NAFLD), and came with straightforward diagnostic criteria, making it easier for doctors, including primary care physicians, to diagnose. Since its introduction, MAFLD has proven to be more effective than NAFLD in assessing risks related to liver and other health complications, identifying high-risk individuals, and understanding disease associations. Many national and international medical organizations have adopted MAFLD due to its clear diagnostic guidelines, elimination of the need to rule out other liver diseases, and reduction of stigma surrounding the condition.

Response Variable & Explanatory Variable

The **response variable** in this study was the presence or absence of **Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD)**, defined according to established clinical criteria. MAFLD was categorized as a **binary outcome variable** (0 = absence, 1 = presence), determined based on hepatic steatosis identified by controlled attenuation parameter (CAP) measurements and the presence of metabolic risk factors, including obesity, type 2 diabetes, and metabolic dysregulation.

The **explanatory variables** included a range of demographic, anthropometric, metabolic, and lifestyle factors. **Demographic variables** comprised age (years), gender (male or female), and race/ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, and Other Race). **Anthropometric measures** included body mass index (BMI, kg/m²), waist circumference (cm), weight (kg), and height (cm). **Liver-related biomarkers** included alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), and CAP score (dB/m). **Metabolic parameters** consisted of total cholesterol (mg/dL), triglycerides (mg/dL), serum uric acid (SUA, mg/dL), systolic blood pressure (SBP, mmHg), and hemoglobin levels (g/dL). **Lifestyle factors** included smoking status (never, past, current) and the presence of comorbidities such as hypertension and diabetes.

These variables were selected based on their established relevance in prior research on metabolic dysfunction and fatty liver disease. They were analyzed to assess their association with MAFLD presence and CAP scores using logistic and linear regression models.

Statistical analysis

All statistical analyses were conducted using **R version 4.3.3** (<http://www.R-project.org>). Continuous variables were expressed as **mean ± standard deviation (SD)**, while categorical variables were summarized as **counts (n) and weighted proportions (%)**. The normality of continuous variables was assessed using the **Shapiro-Wilk test**, and **Levene's test** was employed to evaluate the homogeneity of variance. Comparisons between participants with and without **Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD)** were performed using **one-way analysis of variance (ANOVA)** for normally distributed continuous variables and **Welch's t-test** for comparing Controlled Attenuation Parameter (CAP) scores across gender groups, accounting for unequal variances. **Chi-square (χ^2) tests** were applied to assess differences in categorical variables.

To investigate factors associated with MAFLD, **logistic regression analysis** was conducted, with both univariable and multivariable models fitted to estimate the association between demographic, anthropometric, and metabolic variables and MAFLD status. Results were reported as **odds ratios (ORs) with 95% confidence intervals (CIs)**. Additionally, a **multiple linear regression analysis** was performed to evaluate the determinants of CAP scores, incorporating age, weight, height, body mass index (BMI), waist circumference, liver enzyme markers (alanine aminotransferase [ALT], aspartate aminotransferase [AST],

gamma-glutamyl transferase [GGT]), lipid markers (total cholesterol, triglycerides), systolic blood pressure (SBP), red blood cell (RBC) count, and hemoglobin levels as independent predictors. Model fit was assessed using the **adjusted R^2** and **residual standard error** to determine the proportion of variance explained by the predictors.

The effect of **smoking status** on CAP scores was analyzed using **ANOVA**, and a **post hoc Tukey’s honest significant difference (HSD) test** was conducted to determine pairwise group differences. Statistical significance was defined as $p < 0.05$ for all analyses.

Results & Discussions

Association Between Metabolic, Demographic, and Clinical Characteristics and MAFLD Presence

The final analysis comprised 4,477 patients in total, including 2,282 females and 2,195 males. Table 1 presents the descriptive statistics of study participants stratified by the presence or absence of Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD). The mean age of individuals with MAFLD was significantly higher (53.0 ± 16.7 years) compared to those without MAFLD (47.3 ± 19.2 years, $p < 0.001$). The distribution of gender also differed significantly between groups, with a higher proportion of males in the MAFLD group (54.6%) compared to the non-MAFLD group (44.7%), ($p < 0.001$).

Race was significantly associated with MAFLD ($p < 0.001$), with a notably higher prevalence observed in Mexican American individuals (18.9%) compared to their non-MAFLD counterparts (10.6%). Other Hispanic, Non-Hispanic White, Non-Hispanic Black, and Other Race categories also exhibited significant differences in distribution.

Anthropometric measurements and metabolic parameters demonstrated substantial differences between groups. Individuals with MAFLD exhibited higher weight (94.2 ± 23.3 kg vs. 74.0 ± 17.9 kg, $p < 0.001$), BMI (33.6 ± 7.3 vs. 26.8 ± 5.9 , $p < 0.001$), and waist circumference (110.4 ± 15.2 cm vs. 92.5 ± 14.4 cm, $p < 0.001$). Liver-related biomarkers, including CAP, ALT, AST, and GGT, were significantly elevated in the MAFLD group compared to the non-MAFLD group ($p < 0.001$ for all comparisons).

Lipid profile markers, such as total cholesterol (TC) and triglycerides (TG), were significantly higher in individuals with MAFLD ($p < 0.001$). Similarly, serum uric acid (SUA) levels and systolic blood pressure (SBP) were elevated among those with MAFLD ($p < 0.001$). Red blood cell (RBC) counts and hemoglobin levels were also significantly higher in the MAFLD group ($p < 0.001$).

Smoking history varied significantly between groups, with a lower proportion of current smokers in the MAFLD group (15.7%) compared to the non-MAFLD group (18.7%), ($p < 0.001$). The prevalence of hypertension was significantly higher in individuals with MAFLD (43.3%) compared to those without (25.4%), ($p < 0.001$). Additionally, the presence of diabetes was markedly higher among individuals with MAFLD (29.1%) compared to the non-MAFLD group (10.8%), ($p < 0.001$).

These findings indicate that individuals with MAFLD exhibit significantly different demographic, anthropometric, metabolic, and clinical characteristics compared to those without MAFLD, suggesting a strong association between these factors and the presence of the disease.

Table 1: Descriptive Statistics

Dependent: MAFLD		Absence	Presence	p
Age	Mean (SD)	47.3 (19.2)	53.0 (16.7)	<0.001
Gender.factor	Male	1124 (44.7)	1071 (54.6)	<0.001
	Female	1391 (55.3)	891 (45.4)	
Race.factor	Mexican American	266 (10.6)	371 (18.9)	<0.001
	Other Hispanic	236 (9.4)	189 (9.6)	
	Non-Hispanic White	865 (34.4)	689 (35.1)	
	Non-Hispanic Black	642 (25.5)	363 (18.5)	

Dependent: MAFLD		Absence	Presence	p
Weight	Other Race	506 (20.1)	350 (17.8)	<0.001
	Mean (SD)	74.0 (17.9)	94.2 (23.3)	
Height	Mean (SD)	166.0 (9.8)	167.1 (10.2)	<0.001
	Mean (SD)	26.8 (5.9)	33.6 (7.3)	
BMI	Mean (SD)	26.8 (5.9)	33.6 (7.3)	<0.001
	Mean (SD)	92.5 (14.4)	110.4 (15.2)	
Waist	Mean (SD)	92.5 (14.4)	110.4 (15.2)	<0.001
	Mean (SD)	220.8 (38.2)	323.9 (35.8)	
CAP	Mean (SD)	220.8 (38.2)	323.9 (35.8)	<0.001
	Mean (SD)	19.2 (15.9)	26.8 (19.2)	
ALT	Mean (SD)	19.2 (15.9)	26.8 (19.2)	<0.001
	Mean (SD)	21.1 (12.3)	23.4 (14.2)	
AST	Mean (SD)	21.1 (12.3)	23.4 (14.2)	<0.001
	Mean (SD)	26.5 (41.9)	39.1 (48.6)	
GGT	Mean (SD)	26.5 (41.9)	39.1 (48.6)	<0.001
	Mean (SD)	184.4 (39.9)	190.3 (41.5)	
TC	Mean (SD)	184.4 (39.9)	190.3 (41.5)	<0.001
	Mean (SD)	116.5 (76.4)	177.7 (128.2)	
TG	Mean (SD)	116.5 (76.4)	177.7 (128.2)	<0.001
	Mean (SD)	5.2 (1.4)	5.9 (1.5)	
SUA	Mean (SD)	5.2 (1.4)	5.9 (1.5)	<0.001
	Mean (SD)	123.6 (20.2)	130.0 (18.5)	
SBP	Mean (SD)	123.6 (20.2)	130.0 (18.5)	<0.001
	Mean (SD)	4.7 (0.5)	4.8 (0.5)	
RBC	Mean (SD)	4.7 (0.5)	4.8 (0.5)	<0.001
	Mean (SD)	13.9 (1.5)	14.2 (1.6)	
Hemoglobin	Mean (SD)	13.9 (1.5)	14.2 (1.6)	<0.001
	Never	1544 (61.4)	1116 (56.9)	
Smoking	Past Smoker	500 (19.9)	537 (27.4)	<0.001
	Current Smoker	471 (18.7)	309 (15.7)	
	No	1876 (74.6)	1113 (56.7)	
Hypertension	Yes	639 (25.4)	849 (43.3)	<0.001
	No	2244 (89.2)	1391 (70.9)	
Diabetes	No	2244 (89.2)	1391 (70.9)	<0.001
	Yes	271 (10.8)	571 (29.1)	

Association of Controlled Attenuation Parameter (CAP) Scores by MAFLD Status and Gender

Figure 1 presents the distribution of Controlled Attenuation Parameter (CAP) scores stratified by the presence or absence of Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) and further categorized by gender. The box plots depict the median, interquartile range (IQR), and variability of CAP values for males and females across the two MAFLD status groups.

A notable trend observed in the figure is that males consistently exhibit higher CAP scores than females across both MAFLD statuses. In individuals without MAFLD, the median CAP values for males and females are comparable, though males display a slightly broader range of values. However, among those with MAFLD, CAP values increase significantly for both genders, with a more pronounced shift in males. This suggests a potential sex-based difference in hepatic fat accumulation patterns associated with MAFLD.

The presence of outliers, particularly in the absence of MAFLD, indicates variability in CAP scores among individuals. Furthermore, the IQR is wider for individuals with MAFLD, highlighting increased heterogeneity in CAP scores within this group.

A Welch Two Sample t-test was conducted to compare the mean CAP scores between female and male groups. The test indicates a statistically significant difference in the average CAP values between females (mean = 258.13) and males (mean = 274.16), with a t-statistic of -8.5372 and approximately 4433 degrees of freedom (df = 4433.1). Given the extremely small p-value ($< 2.2 \times 10^{-16}$), the null hypothesis of no difference in CAP means between genders is strongly rejected. The 95% confidence interval for the difference in means is estimated between -19.71 and -12.35, indicating that, on average, females have significantly lower CAP scores than males.

These findings align with the statistical analysis, supporting the hypothesis that gender influences CAP levels and suggesting that males are more likely to exhibit higher hepatic fat accumulation. This warrants further investigation into sex-specific patho-physiological mechanisms contributing to MAFLD and hepatic steatosis progression.

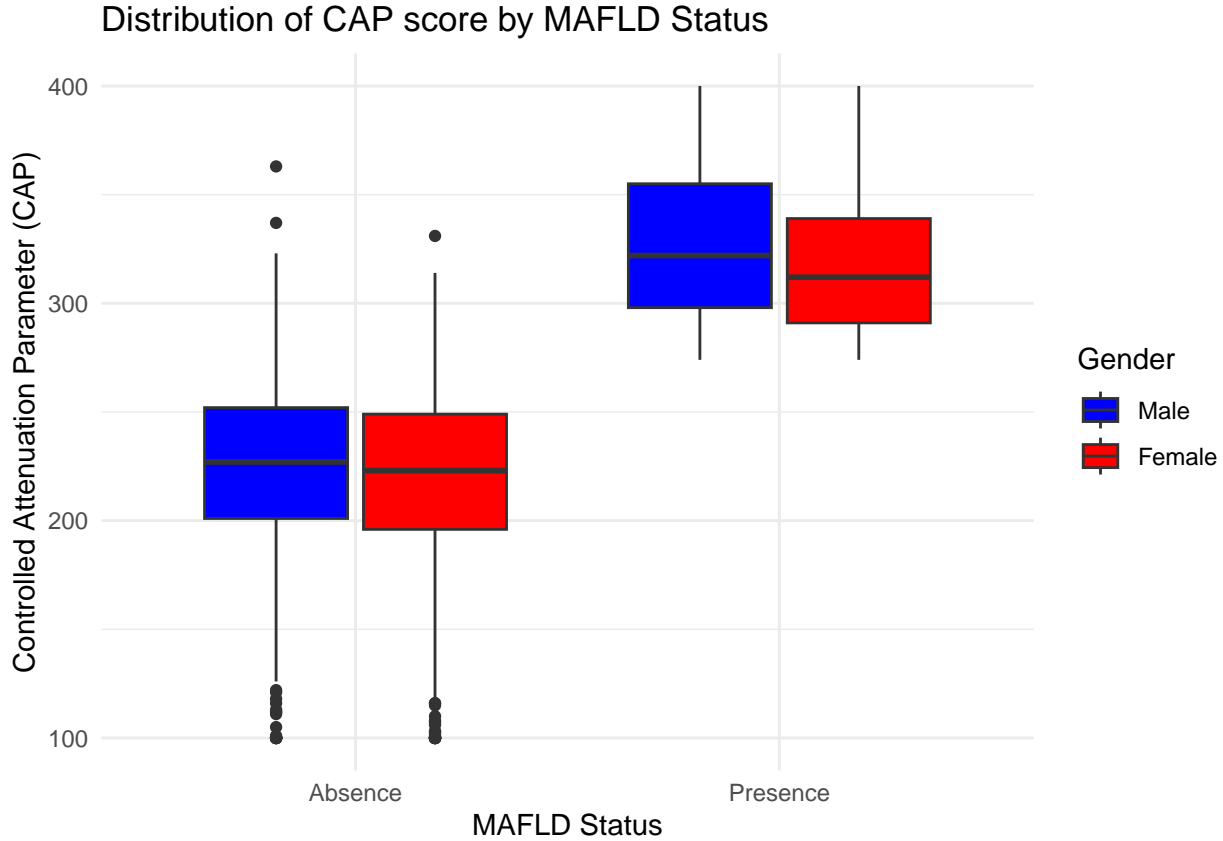


Figure 1: Distribution of Controlled Attenuation Parameter (CAP) score by MAFLD Status

Table 2: Welch Two Sample t-test Results

Statistic	Value
t-value	-8.5372
Degrees of Freedom	4433.1
p-value	< 2.2e-16
95% CI Lower	-19.71247
95% CI Upper	-12.34968
Mean of X	258.1288
Mean of Y	274.1599

Association of Controlled Attenuation Parameter (CAP) Scores and Smoking Status

An analysis of variance (ANOVA) was conducted to evaluate the effect of smoking status on Controlled Attenuation Parameter (CAP) scores. The ANOVA results (Table 1) indicate a statistically significant effect of smoking status on CAP levels,

$$F(2, 4474) = 31.14, \quad p < 0.001.$$

The smoking factor accounts for a sum of squares of 245,796, with a mean square of 122,898, while the residual variance is 17,655,870 with a mean square of 3,946.

The extremely small p-value

$$p = 3.7 \times 10^{-14}$$

suggests strong evidence against the null hypothesis, indicating that CAP levels significantly differ across smoking categories. The large

$$F\text{-statistic} = 31.14$$

further supports the conclusion that smoking status has a substantial impact on hepatic fat accumulation, as measured by CAP.

Given the statistical significance observed, a post hoc analysis (e.g., Tukey's HSD) would be necessary to determine which specific smoking groups exhibit significant differences in CAP scores. These findings suggest a potential relationship between smoking behavior and hepatic steatosis, warranting further investigation into underlying physiological mechanisms.

To further investigate these differences, *Tukey's HSD test* and *Bonferroni-adjusted pairwise t-tests* were conducted. The results indicate that:

- **Past smokers** had significantly higher CAP scores than both never smokers ($p < 0.001$) and current smokers ($p < 0.001$), suggesting that smoking cessation may be linked to increased hepatic fat accumulation.
- **Current smokers** did not differ significantly from never smokers ($p = 0.996$), indicating that active smoking might not strongly influence hepatic steatosis.

These findings suggest that **post-cessation metabolic changes**, such as weight gain and altered lipid metabolism, may contribute to the increased CAP scores observed in past smokers. While smoking has been associated with oxidative stress and insulin resistance, its cessation may trigger metabolic shifts that promote hepatic fat accumulation. The lack of a significant difference between current and never smokers challenges previous research linking smoking to liver fat accumulation and underscores the need for further exploration.

Table 3: Tukey HSD Pairwise Comparisons for CAP Scores by Smoking Status

	Comparison	Difference	Lower CI	Upper CI	p-value
Past Smoker-Never	Past Smoker-Never	17.5121	12.1202	22.9039	0.0000
Current Smoker-Never	Current Smoker-Never	-0.2223	-6.2193	5.7747	0.9958
Current Smoker-Past Smoker	Current Smoker-Past Smoker	-17.7344	-24.7149	-10.7539	0.0000

Table 4: Bonferroni-Adjusted Pairwise t-Tests for CAP Scores by Smoking Status

Comparison	Compared_to	p-value
Past Smoker	Never	0
Current Smoker	Never	1
Current Smoker	Past Smoker	0

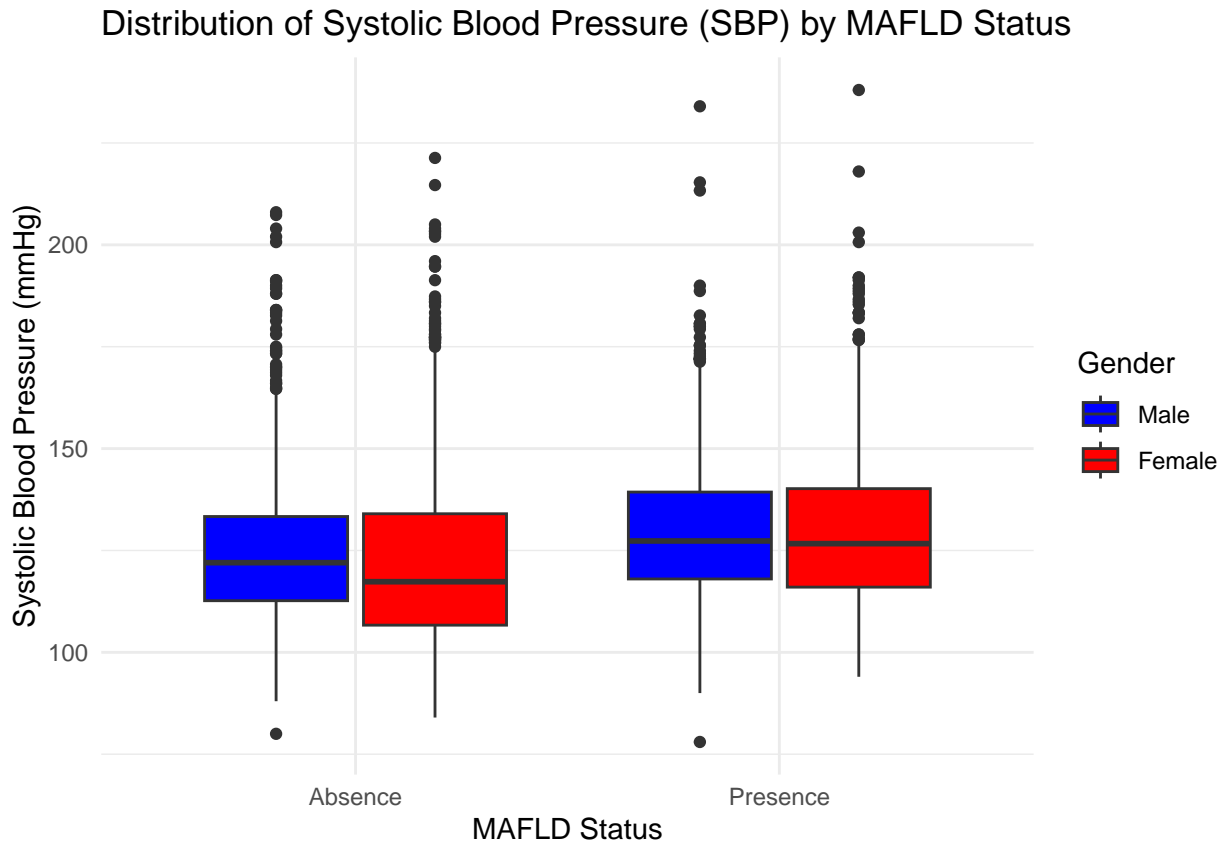


Figure 2: Distribution of Systolic Blood Pressure (SBP) by MAFLD Status

Association of Metabolic Dysfunction-Associated Fatty Liver Disease with Systolic Blood Pressure, Total Cholesterol, and Triglycerides

Figure 2 presents the distribution of systolic blood pressure (SBP) levels stratified by Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) status and further categorized by gender. The boxplots illustrate the median, interquartile range (IQR), and overall variability of SBP values in both males and females across the two MAFLD status groups. This visualization provides insights into potential differences in blood pressure levels associated with MAFLD and gender-based variations in its distribution.

Individuals diagnosed with MAFLD exhibit slightly higher median SBP values compared to those without MAFLD. This trend suggests a potential association between MAFLD and elevated blood pressure, aligning with previous research that links metabolic dysfunction with an increased risk of hypertension. However, despite this observable increase, the degree of separation between the two groups is relatively small, indicating that additional factors beyond MAFLD status may contribute to SBP variability.

When considering gender-based differences, the distribution of SBP levels among males and females appears largely similar across both MAFLD status groups. Males tend to display a slightly broader interquartile range and greater variability in SBP values compared to females, but the median values remain relatively close. This observation suggests that while gender may influence the dispersion of SBP, it does not substantially impact the central tendency of blood pressure levels within each MAFLD status category.

The presence of outliers is another notable feature of the data distribution. Across all subgroups, a substantial number of individuals exhibit markedly high SBP values, particularly above 150 mmHg. These outliers suggest that certain individuals within the dataset may be experiencing hypertension or other underlying cardiovascular conditions. The similar distribution of outliers across both MAFLD status groups further supports the need to consider additional covariates, such as age, body mass index (BMI), and lifestyle factors,

in assessing the relationship between MAFLD and blood pressure.

Overall, the visualization suggests a modest increase in SBP levels among individuals with MAFLD, though the degree of variation remains relatively small. Gender-related differences in SBP appear minimal, with males exhibiting slightly greater variability but no substantial difference in median values compared to females. Given the observed variability and the presence of numerous outliers, further statistical analyses, such as analysis of variance (ANOVA) or multivariate regression, are necessary to determine the significance of these differences and to control for potential confounders. Future studies should also explore longitudinal trends to assess the causal relationship between MAFLD and hypertension, considering the broader metabolic implications of both conditions.

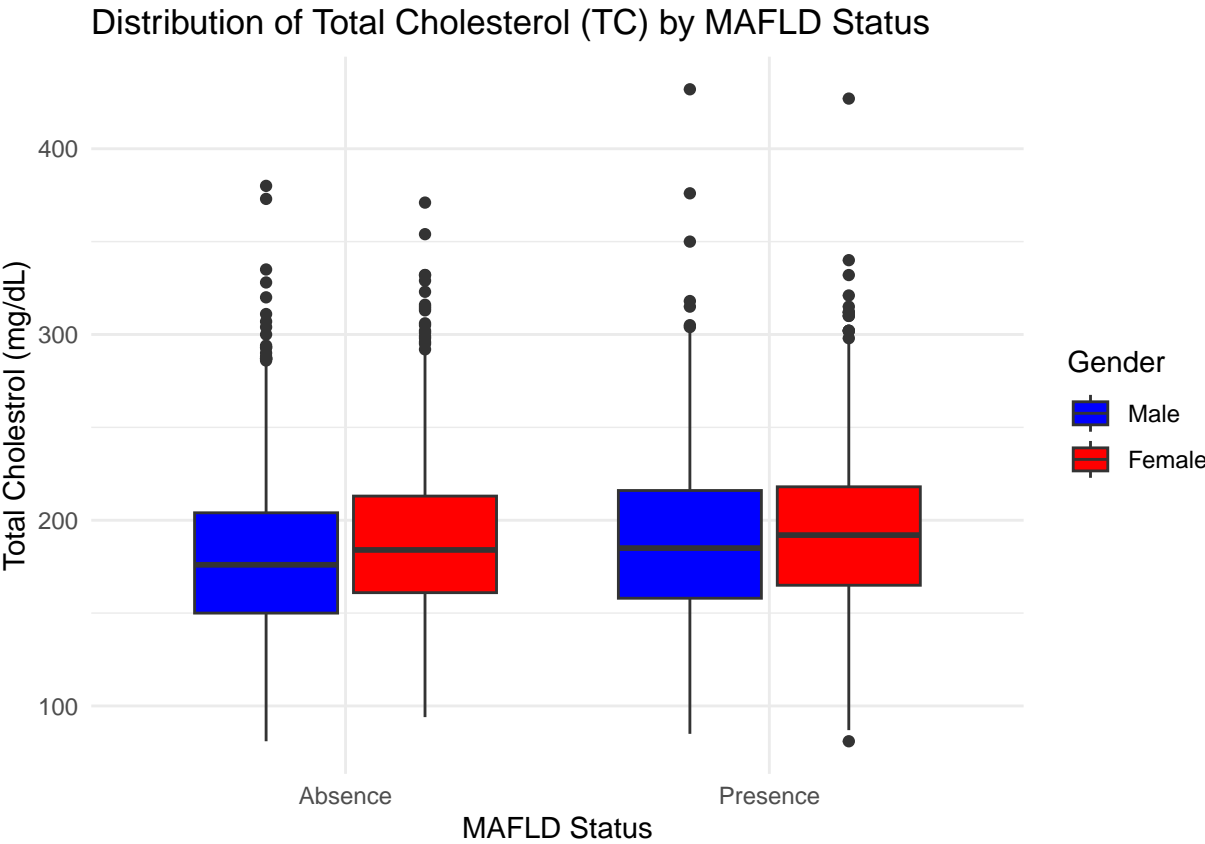


Figure 3: Distribution of Total Cholesterol (TC) by MAFLD Status

Figure 3 presents the distribution of Total Cholesterol (TC) levels stratified by Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) status and further categorized by gender. The boxplots illustrate the median, interquartile range (IQR), and overall variability of TC levels among males and females in both MAFLD absence and presence groups. This visualization provides insights into potential variations in cholesterol levels associated with MAFLD and gender-related differences in lipid metabolism.

Individuals diagnosed with MAFLD exhibit slightly higher median TC levels compared to those without MAFLD. This observation suggests a potential association between MAFLD and increased cholesterol levels, consistent with previous research indicating a link between metabolic dysfunction and lipid abnormalities. However, the overlap in distributions between the two MAFLD groups indicates that the effect of MAFLD on TC levels may not be substantial, and additional factors could contribute to cholesterol variability.

When analyzing gender-based differences, the distribution of TC levels among males and females appears relatively similar in both MAFLD status groups. While females tend to exhibit a marginally higher median TC value, the overall spread of cholesterol levels remains comparable across genders. This suggests that gender may not play a significant role in determining total cholesterol levels within each MAFLD category.

A notable number of outliers are present across all subgroups, particularly in the higher TC range (> 300 mg/dL), suggesting that certain individuals have markedly elevated cholesterol levels. The distribution of these outliers appears relatively balanced across MAFLD status and gender, indicating that other metabolic or lifestyle-related factors may contribute to extreme cholesterol levels.

Overall, the visualization suggests a slight increase in TC levels among individuals with MAFLD, but the degree of separation between groups remains relatively small. Gender differences in cholesterol distribution appear minimal, with slightly higher median values observed in females. Given the observed variability and the presence of numerous outliers, further statistical analyses, such as ANOVA or multivariate regression modeling, are necessary to determine the significance of these differences while controlling for potential confounders, such as diet, physical activity, and genetic predisposition. Future research should explore the metabolic pathways linking MAFLD to dyslipidemia and investigate longitudinal changes in cholesterol levels to better understand the progression of metabolic dysfunction.

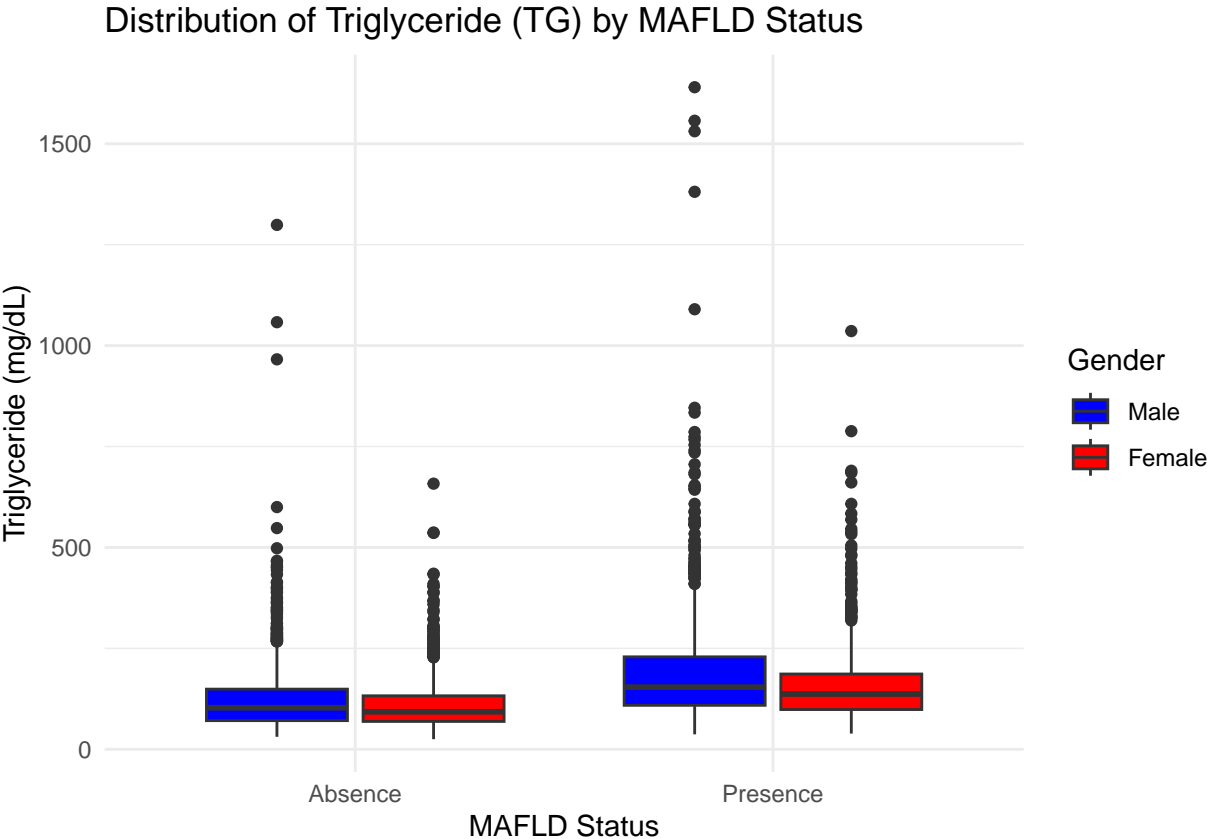


Figure 4: Distribution of Triglyceride (TG) by MAFLD Status

Figure 4 presents the distribution of triglyceride (TG) levels stratified by Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) status and further categorized by gender. The boxplots illustrate the median, interquartile range (IQR), and overall variability of TG levels in both males and females across the two MAFLD status groups. This visualization provides insights into potential differences in triglyceride levels associated with MAFLD and gender-related variations in lipid metabolism.

Individuals with MAFLD exhibit higher median triglyceride levels compared to those without the condition. This trend aligns with existing literature linking MAFLD to dyslipidemia, particularly hypertriglyceridemia, as a key metabolic abnormality associated with hepatic fat accumulation. The rightward shift in TG distribution among individuals with MAFLD supports the hypothesis that this condition is associated with impaired lipid metabolism. However, the degree of overlap between MAFLD presence and absence groups suggests that additional metabolic factors, such as insulin resistance or dietary influences, may contribute to triglyceride

variability.

When considering gender-based differences, males and females exhibit relatively similar distributions of triglyceride levels within each MAFLD status group. Although the median TG levels are slightly higher in males, the overall spread of values remains comparable across genders. This observation suggests that gender alone may not be a primary determinant of triglyceride levels in individuals with or without MAFLD.

A substantial number of outliers are observed in all subgroups, particularly in the higher TG range (> 500 mg/dL). The presence of extreme outliers, some exceeding 1,000 mg/dL, indicates that a subset of individuals may have severe hypertriglyceridemia, which is a known risk factor for cardiovascular disease and pancreatitis. The frequency of these outliers appears greater in individuals with MAFLD, reinforcing the association between this condition and elevated triglyceride levels.

Overall, the visualization suggests that triglyceride levels tend to be higher in individuals with MAFLD, though the degree of variation within groups is substantial. Gender-related differences in triglyceride levels appear minimal, with a slightly broader distribution in males. The presence of numerous extreme outliers highlights the need for further statistical analyses, such as ANOVA or regression modeling, to quantify the significance of these differences while controlling for potential confounders, including diet, body mass index (BMI), and metabolic disorders such as diabetes. Future research should explore longitudinal changes in triglyceride levels to assess their role in MAFLD progression and associated cardiovascular risks.

Logistic Regression for MAFLD Status

Table 5 presents the results of a logistic regression analysis examining factors associated with Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) status. The table reports descriptive statistics, including mean and standard deviation values for individuals with and without MAFLD, along with odds ratios (ORs) from both univariable and multivariable logistic regression models. The univariable ORs assess the crude association between each predictor and MAFLD, while the multivariable ORs account for potential confounders.

Age was significantly higher among individuals with MAFLD (mean = 53.0 years) compared to those without MAFLD (mean = 47.3 years). The univariable analysis indicated a positive association between age and MAFLD risk (OR = 1.02, $p < 0.001$), suggesting that each additional year of age increased the odds of having MAFLD by approximately 2%. This association remained statistically significant in the multivariable model, though with a slightly weaker effect size (OR = 1.02, $p = 0.014$), confirming age as an independent risk factor for MAFLD.

Gender differences were also examined, with males comprising 48.8% of the MAFLD group compared to 51.2% of the non-MAFLD group. In the univariable model, females had significantly lower odds of MAFLD (OR = 0.67, $p < 0.001$). However, after adjusting for confounders, this association was no longer significant (OR = 0.90, $p = 0.763$), indicating that gender differences in MAFLD prevalence may be mediated by other metabolic factors rather than being an independent determinant.

Race and ethnicity were strongly associated with MAFLD in the unadjusted analysis. Mexican Americans had the highest prevalence of MAFLD (58.2%), while Non-Hispanic Black individuals had the lowest (36.1%). Compared to Mexican Americans, Non-Hispanic White individuals (OR = 0.57, $p < 0.001$), Non-Hispanic Black individuals (OR = 0.41, $p < 0.001$), and Other Hispanic individuals (OR = 0.57, $p < 0.001$) all had significantly lower odds of MAFLD. However, in the multivariable model, only the association for Non-Hispanic White individuals remained statistically significant (OR = 0.46, $p = 0.023$), suggesting that race-specific differences in MAFLD risk are largely influenced by other metabolic factors.

Anthropometric and metabolic measures, including BMI, waist circumference, and body weight, were significantly associated with MAFLD in the univariable model. BMI had a particularly strong association (OR = 1.19, $p < 0.001$), indicating that each unit increase in BMI increased the odds of MAFLD by 19%. Waist circumference (OR = 1.09, $p < 0.001$) and weight (OR = 1.05, $p < 0.001$) also showed significant positive associations. However, in the multivariable model, only controlled attenuation parameter (CAP) remained a strong independent predictor of MAFLD (OR = 1.26, $p < 0.001$), reinforcing the role of hepatic fat accumulation in disease development.

Biochemical markers, including ALT, AST, triglycerides (TG), and total cholesterol (TC), were all significantly associated with MAFLD in the unadjusted model. However, after adjustment, most associations diminished, except for CAP and ALT, which remained borderline significant (OR = 1.02, $p = 0.055$). This suggests that while these markers reflect metabolic dysfunction, their impact on MAFLD risk is mediated through other factors, such as obesity and hepatic fat accumulation.

Blood pressure and hematological markers were also analyzed. Serum uric acid (SUA, OR = 1.38, $p < 0.001$) was significantly associated with MAFLD in the univariable model, but this association lost significance after adjustment (OR = 0.90, $p = 0.217$), suggesting that SUA is more likely a consequence rather than a direct cause of MAFLD. Similarly, systolic blood pressure (SBP) and red blood cell (RBC) count were significant in univariable models. While RBC count remained borderline significant in the multivariable model (OR = 1.15, $p = 0.659$), SBP did not (OR = 0.99, $p = 0.306$), implying that hypertension may not be an independent risk factor for MAFLD after adjusting for other variables.

Overall, the analysis identifies age, BMI, waist circumference, and hepatic fat accumulation (CAP) as the strongest independent predictors of MAFLD. While unadjusted analyses suggested significant associations for gender, race, and several metabolic markers, multivariable adjustments indicate that many of these factors are confounded by obesity-related measures. The strong association between CAP and MAFLD highlights its clinical relevance as a diagnostic marker for hepatic steatosis. Future research should focus on longitudinal studies to assess causal relationships and investigate targeted interventions addressing obesity and metabolic dysfunction to reduce MAFLD risk.

Table 5: Logistic Regression for MAFLD Status

	MAFLD		Absence	Presence	OR (univariable)	OR (multivariable)
1	Age	Mean (SD)	47.3 (19.2)	53.0 (16.7)	1.02 (1.01-1.02, $p < 0.001$)	1.02 (1.00-1.04, $p = 0.014$)
9	Gender.factor	Male	1124 (51.2)	1071 (48.8)	-	-
8		Female	1391 (61.0)	891 (39.0)	0.67 (0.60-0.76, $p < 0.001$)	0.90 (0.46-1.77, $p = 0.763$)
15	Race.factor	Mexican American	266 (41.8)	371 (58.2)	-	-
18		Other Hispanic	236 (55.5)	189 (44.5)	0.57 (0.45-0.74, $p < 0.001$)	0.71 (0.31-1.63, $p = 0.417$)
17		Non-Hispanic White	865 (55.7)	689 (44.3)	0.57 (0.47-0.69, $p < 0.001$)	0.46 (0.23-0.89, $p = 0.023$)
16		Non-Hispanic Black	642 (63.9)	363 (36.1)	0.41 (0.33-0.50, $p < 0.001$)	0.52 (0.24-1.09, $p = 0.084$)
19		Other Race	506 (59.1)	350 (40.9)	0.50 (0.40-0.61, $p < 0.001$)	0.64 (0.32-1.29, $p = 0.215$)
29	Weight	Mean (SD)	74.0 (17.9)	94.2 (23.3)	1.05 (1.05-1.06, $p < 0.001$)	1.08 (0.97-1.22, $p = 0.169$)
11	Height	Mean (SD)	166.0 (9.8)	167.1 (10.2)	1.01 (1.01-1.02, $p < 0.001$)	0.93 (0.82-1.05, $p = 0.244$)
4	BMI	Mean (SD)	26.8 (5.9)	33.6 (7.3)	1.19 (1.17-1.20, $p < 0.001$)	0.90 (0.66-1.23, $p = 0.509$)
28	Waist	Mean (SD)	92.5 (14.4)	110.4 (15.2)	1.09 (1.08-1.09, $p < 0.001$)	0.99 (0.96-1.03, $p = 0.737$)
5	CAP	Mean (SD)	220.8 (38.2)	323.9 (35.8)	1.26 (1.23-1.29, $p < 0.001$)	1.26 (1.23-1.30, $p < 0.001$)
2	ALT	Mean (SD)	19.2 (15.9)	26.8 (19.2)	1.03 (1.03-1.04, $p < 0.001$)	1.02 (1.00-1.05, $p = 0.055$)
3	AST	Mean (SD)	21.1 (12.3)	23.4 (14.2)	1.01 (1.01-1.02, $p < 0.001$)	0.98 (0.95-1.00, $p = 0.074$)

	MAFLD		Absence	Presence	OR (univariable)	OR (multivariable)
10	GGT	Mean (SD)	26.5 (41.9)	39.1 (48.6)	1.01 (1.01-1.01, p<0.001)	1.00 (1.00-1.01, p=0.474)
26	TC	Mean (SD)	184.4 (39.9)	190.3 (41.5)	1.00 (1.00-1.01, p<0.001)	1.00 (1.00-1.01, p=0.399)
27	TG	Mean (SD)	116.5 (76.4)	177.7 (128.2)	1.01 (1.01-1.01, p<0.001)	1.00 (1.00-1.00, p=0.188)
25	SUA	Mean (SD)	5.2 (1.4)	5.9 (1.5)	1.38 (1.32-1.44, p<0.001)	0.90 (0.77-1.06, p=0.217)
21	SBP	Mean (SD)	123.6 (20.2)	130.0 (18.5)	1.02 (1.01-1.02, p<0.001)	0.99 (0.98-1.01, p=0.306)
20	RBC	Mean (SD)	4.7 (0.5)	4.8 (0.5)	1.80 (1.60-2.04, p<0.001)	1.15 (0.62-2.14, p=0.659)
12	Hemoglobin	Mean (SD)	13.9 (1.5)	14.2 (1.6)	1.17 (1.12-1.21, p<0.001)	0.91 (0.74-1.11, p=0.364)
23	Smoking	Never	1544 (58.0)	1116 (42.0)	-	-
24		Past Smoker	500 (48.2)	537 (51.8)	1.49 (1.29-1.72, p<0.001)	0.65 (0.39-1.08, p=0.099)
22		Current Smoker	471 (60.4)	309 (39.6)	0.91 (0.77-1.07, p=0.244)	0.84 (0.47-1.51, p=0.563)
13	Hypertension	No	1876 (62.8)	1113 (37.2)	-	-
14		Yes	639 (42.9)	849 (57.1)	2.24 (1.97-2.54, p<0.001)	1.58 (0.96-2.59, p=0.070)
6	Diabetes	No	2244 (61.7)	1391 (38.3)	-	-
7		Yes	271 (32.2)	571 (67.8)	3.40 (2.90-3.99, p<0.001)	0.86 (0.49-1.50, p=0.602)

Regression Analysis of Determinants of Controlled Attenuation Parameter (CAP)

A multiple linear regression analysis was conducted to examine the association between **Controlled Attenuation Parameter (CAP)** and a set of demographic, anthropometric, and metabolic variables. The model included **age, weight, height, BMI, waist circumference, liver enzymes (ALT, AST, GGT), lipid markers (total cholesterol [TC], triglycerides [TG]), serum uric acid (SUA), systolic blood pressure (SBP), red blood cell count (RBC), and hemoglobin levels** as independent predictors.

The regression model demonstrated a statistically significant overall fit ($F_{14,4184} = 221.3$, $p < 2.2 \times 10^{-16}$), indicating that the included predictors explain a significant portion of the variance in Controlled Attenuation Parameter (CAP).

The adjusted R-squared value (0.4236) accounts for the number of predictors in the model, providing a more accurate estimate of the model's generalizability. The minimal difference between R^2 and adjusted R^2 indicates that most predictors contribute meaningful explanatory power and that the model does not suffer significantly from overfitting.

The residual standard error (47.94 dB/m) represents the average deviation of observed CAP values from the predicted values. Given the scale of CAP measurements, this suggests moderate dispersion in residuals, indicating that while the model provides reasonable predictions, unexplained variability remains.

The dataset originally contained more observations, but 278 cases were removed due to missing values, potentially influencing the model's estimates. While deletion due to missingness is a common approach, further sensitivity analysis or imputation techniques could be explored to assess the impact of missing data on the results.

Table 6: Regression Summary: Controlled Attenuation Parameter (CAP) score Model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	119.3781	54.8640	2.1759	0.0296
Age	0.3160	0.0558	5.6659	0.0000
Weight	0.4836	0.3216	1.5036	0.1328
Height	-0.8450	0.3249	-2.6007	0.0093
BMI	-0.8247	0.8756	-0.9419	0.3463
Waist	1.6730	0.1403	11.9236	0.0000
ALT	0.5837	0.0732	7.9789	0.0000
AST	-0.2663	0.0979	-2.7215	0.0065
GGT	0.0072	0.0193	0.3735	0.7088
TC	-0.0236	0.0195	-1.2109	0.2260
TG	0.0905	0.0078	11.6583	0.0000
SUA	0.8015	0.5667	1.4145	0.1573
SBP	0.0938	0.0444	2.1105	0.0349
RBC	6.4695	2.1240	3.0459	0.0023
Hemoglobin	1.7943	0.7040	2.5486	0.0109

Key findings Age was a **significant positive predictor of CAP** ($\beta = 0.316, p < 0.001$), indicating that for each additional year of age, CAP increased by approximately **0.32 dB/m**. This finding suggests that hepatic fat accumulation, as measured by CAP, tends to increase with aging.

Among anthropometric measures, **waist circumference demonstrated the strongest positive association with CAP** ($\beta = 1.673, p < 0.001$), highlighting the role of central adiposity in hepatic fat accumulation. Weight was not a significant predictor ($p = 0.133$), while BMI, despite its known association with obesity, was also not significantly related to CAP ($p = 0.346$). Interestingly, **height was inversely associated with CAP** ($\beta = -0.845, p = 0.009$), suggesting a potential negative correlation between stature and liver fat accumulation.

Liver enzymes showed mixed associations with CAP. **ALT was significantly positively associated with CAP** ($\beta = 0.584, p < 0.001$), while **AST demonstrated a significant inverse association** ($\beta = -0.266, p = 0.007$). GGT, another marker of liver function, was not significantly associated with CAP ($p = 0.709$). These results indicate that while ALT is strongly correlated with hepatic fat accumulation, AST may reflect distinct hepatic processes or injury mechanisms.

Regarding lipid markers, **triglycerides (TG) exhibited a strong positive association with CAP** ($\beta = 0.091, p < 0.001$), reinforcing the well-established link between hypertriglyceridemia and hepatic fat accumulation. However, **total cholesterol (TC) was not significantly associated with CAP** ($p = 0.226$), suggesting that triglyceride levels may be a more relevant lipid marker for predicting hepatic steatosis.

Among other metabolic markers, **systolic blood pressure (SBP) was positively associated with CAP** ($\beta = 0.094, p = 0.035$), indicating a modest but significant relationship between blood pressure and hepatic fat accumulation. **Red blood cell (RBC) count was also a significant predictor** ($\beta = 6.47, p = 0.002$), as was **hemoglobin** ($\beta = 1.79, p = 0.011$), suggesting potential hematological contributions to CAP variation. Serum uric acid (SUA), despite being elevated in metabolic dysfunction, was not significantly associated with CAP in this model ($p = 0.157$).

Conclusion

This study provides a comprehensive analysis of the demographic, anthropometric, and metabolic determinants of Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) using data from NHANES 2017-2018. Through logistic regression analysis, we identified age, waist circumference, triglycerides, and hepatic fat accumulation (CAP scores) as the strongest independent predictors of MAFLD. The significant association

between metabolic dysfunction—particularly elevated triglycerides, hypertension, and liver enzyme markers—and MAFLD highlights the systemic nature of the disease. Additionally, race/ethnicity remained a significant factor in MAFLD risk, with Mexican Americans exhibiting the highest prevalence, even after adjusting for metabolic covariates.

The multiple linear regression analysis further reinforced these findings by identifying waist circumference, triglycerides, ALT levels, and systolic blood pressure (SBP) as key determinants of hepatic fat accumulation, as measured by CAP scores. Interestingly, while BMI was strongly associated with MAFLD in unadjusted analyses, it did not remain a significant independent predictor in the multivariable models, suggesting that central adiposity (waist circumference) is a more precise indicator of hepatic steatosis risk than BMI alone. The strong association between triglycerides and CAP scores underscores the critical role of lipid metabolism in liver fat accumulation and suggests that triglyceride levels could serve as a non-invasive biomarker for early MAFLD detection.

Our findings emphasize the importance of a risk-based screening approach for MAFLD, shifting away from reliance on BMI alone and incorporating waist circumference, lipid abnormalities, and CAP measurements for more accurate disease identification. The results also support the clinical utility of the MAFLD classification over NAFLD, as it enables a more inclusive and targeted assessment of metabolic liver disease without the requirement to exclude other liver conditions.

Given the increasing global burden of metabolic syndrome and obesity, these findings highlight the urgent need for early interventions, lifestyle modifications, and targeted metabolic therapies to reduce MAFLD prevalence and its associated complications, including cardiovascular disease and liver fibrosis.

Limitation

Despite the strengths of this study, including the use of a large, nationally representative dataset (NHANES 2017-2018) and the application of robust statistical analyses, several limitations should be acknowledged.

First, residual confounding may exist, as NHANES does not capture certain lifestyle factors that could influence MAFLD risk, such as dietary composition, physical activity intensity, or genetic predisposition. While we adjusted for key metabolic and anthropometric variables, the potential for unmeasured confounders remains. Additionally, violations of regression assumptions, such as multicollinearity among predictors, could affect the reliability of our findings. Although we assessed Variance Inflation Factors (VIFs) to detect collinearity, subtle dependencies among metabolic markers may still introduce bias in the estimated associations.

Second, the cross-sectional design of this study limits causal inference, as the definition of MAFLD used did not account for longitudinal metabolic changes. The reliance on a single time-point assessment of metabolic parameters may not fully capture the dynamic progression or regression of the disease over time. Future studies should incorporate longitudinal data to better evaluate temporal trends in MAFLD development and metabolic risk factors.

Despite these limitations, this study provides important insights into the metabolic and anthropometric predictors of MAFLD, reinforcing the need for early screening and targeted interventions to mitigate the growing burden of metabolic liver disease. Future research should explore the integration of longitudinal metabolic profiling, causal inference methods, and machine learning approaches to improve risk stratification and intervention strategies.

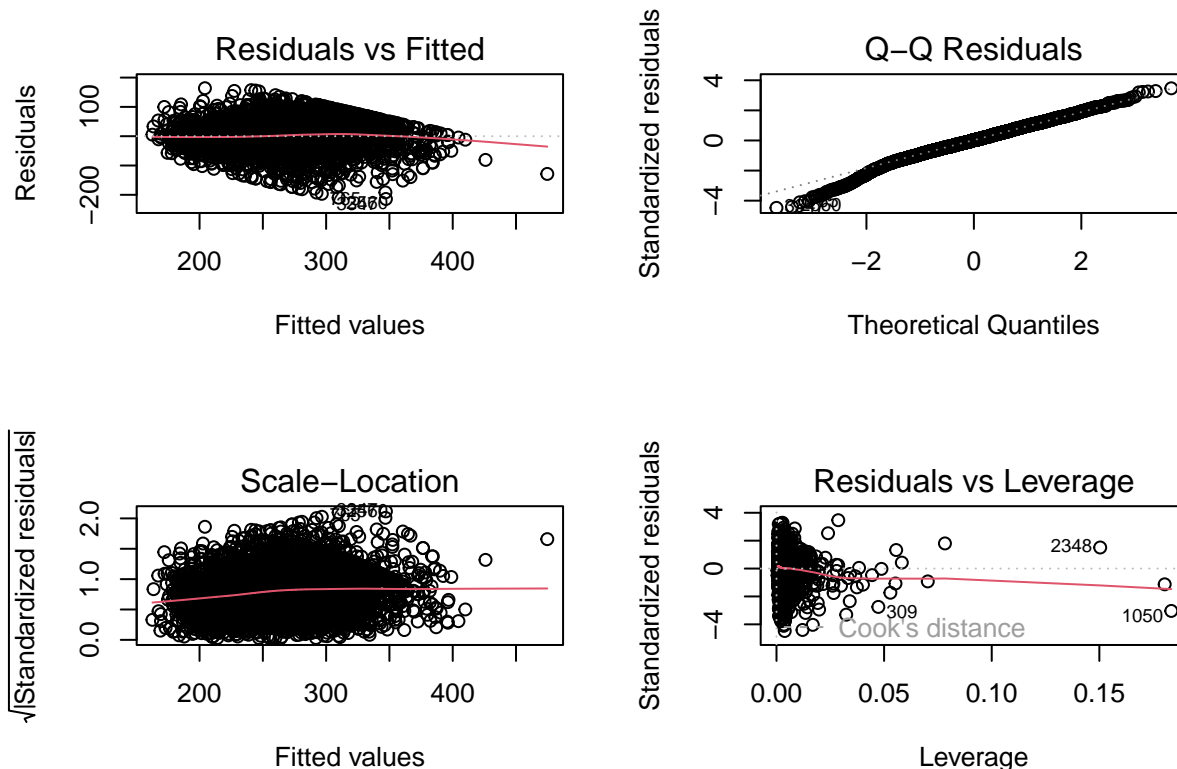
Reference

1. Dai, X., Zhou, G., & Xu, L. (2022). Associations between red blood cell count and metabolic dysfunction-associated fatty liver disease(MAFLD). *PloS one*, 17(12), e0279274. <https://doi.org/10.1371/journal.pone.0279274>
2. Gofton, C., Upendran, Y., Zheng, M. H., & George, J. (2023). MAFLD: How is it different from NAFLD?. *Clinical and molecular hepatology*, 29(Suppl), S17–S31. <https://doi.org/10.3350/cmh.2022.0367>

3. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R (1st ed., Vol. 103). Springer Nature. <https://doi.org/10.1007/978-1-4614-7138-7>

Supplement

The diagnostic tests for model assumptions indicate several key findings. The Durbin-Watson test ($DW = 1.977$, $p = 0.2283$) suggests no significant autocorrelation, meaning the residuals are independent, satisfying this assumption. However, the Kolmogorov-Smirnov (KS) test ($D = 0.027$, $p = 0.0041$) indicates a significant deviation from normality, which is also reflected in the Q-Q plot, where slight deviations appear at the tails. The Breusch-Pagan test ($BP = 66.085$, $p < 0.001$) detects heteroscedasticity, implying that residual variance is not constant across fitted values, a finding supported by the Scale-Location plot, which shows increasing residual spread at higher fitted values. Multicollinearity is another concern, with Weight ($VIF = 94.07$), BMI ($VIF = 71.88$), and Height ($VIF = 19.21$) exhibiting extremely high collinearity, suggesting redundancy among these predictors. The Residuals vs. Fitted plot shows some curvature, hinting at potential non-linearity, while the Residuals vs. Leverage plot identifies influential points (Observations 2348, 309, and 1050) that may affect model stability. While the assumption of independent errors holds, corrections are needed for normality (e.g., log transformation), variance consistency (e.g., weighted least squares or robust standard errors), and multicollinearity (e.g., removing or combining highly correlated predictors like Weight and BMI) to improve model validity.



```
##
## Durbin-Watson test
##
## data: CAP_model
## DW = 1.977, p-value = 0.2283
## alternative hypothesis: true autocorrelation is greater than 0
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
```



```
## data: residuals_std
## D = 0.027166, p-value = 0.004068
## alternative hypothesis: two-sided
```

```
##
## studentized Breusch-Pagan test
##
```

```
## data: CAP_model
## BP = 66.085, df = 14, p-value = 9.787e-09
```

##	Age	Weight	Height	BMI	Waist	ALT	AST
##	1.905855	94.072454	19.205627	71.876510	10.612434	3.117873	3.060650
##	GGT	TC	TG	SUA	SBP	RBC	Hemoglobin
##	1.435682	1.166302	1.256598	1.282933	1.377562	2.077860	2.146937