Cardiovascular risk factors: Hospital Patient Study in South Asia

Daniel Tehrani and Ranko Kettunen

George Washington University / Milken Institute School of Public Health

PUBH 6867 – Health Data Visualization

Professor Nusrat Epsi

December 11, 2025

Introduction

Cardiovascular disease (CVD) is a general term for adverse health conditions that affect the heart and blood vessels. CVD is a critical global health issue, causing more deaths than any other condition and leading to significant long-term disability (World Health Organization, 2025 & Center for Disease Control and Prevention [CDC], 2024). In 2020, CVD accounted for approximately 20 million deaths worldwide, representing about one-third of all global mortality (WHO, 2025). Moreover, cardiovascular diseases have a disproportionate impact globally, with about 75% of CVD related deaths taking place in low- and middle-income countries (LMICs) (WHO, 2025). Cardiovascular disease is largely preventable through healthy lifestyle choices, and timely interventions can significantly reduce its impact (CDC, 2024). Yet managing CVD remains difficult as it results from the combined effects of biological, behavioral, environmental, and social determinants factors, which interact in a complex way (CDC, 2024). Factors that increase the likelihood of developing CVD are known as CVD risk factors, including high blood pressure, high cholesterol, diabetes, obesity, and smoking (WHO, 2025 & CDC, 2024).

Cardiovascular disease risk factors are similar worldwide, but their epidemiological prevalence and interactions vary across regions (WHO, 2019 & GBD 2023 Cardiovascular Disease Collaborators, 2025). Prior studies have adopted CVD risk prediction models which were derived based on risk factors that were considered in high-income populations, and these models may not necessarily be applicable in LMICs regions (WHO, 2019). WHO emphasized that CVD risk models require regional specific adjustment to increase accuracy of risk prediction and strengthen prevention strategies (WHO, 2019). Moreover, hospital-based CVD studies are allowed to study acute events, where short-term prognosis is critical (Yang et al., 2024). In such cases, hospital-based CVD risk assessment should focus on predicting short term risk, which

may be more relevant for clinical decision making rather than 10-year projections, which are often used in CVD risk assessment (Yan et al., 2024). Hospital-based CVD risk studies have been conducted primarily in developed countries; however, a gap in knowledge regarding hospital-based CVD risk in low- and middle-income countries remains. In this study, we sought to evaluate CVD risk among hospital patients in Bangladesh, within South Asia region, to examine patterns of risk and variability among contributing risk factors.

The data set we were dealing with was known by the acronym CAIR CVD 2025 (Nirob et al., 2025, March 3). The sample size included 1,529 patients. Collection of data took place in Bangladesh. The location was Jamalpur Medical College Hospital. The variables we tested in our data were systolic BP, total cholesterol, smoking status, diabetes status, BMI, age, sex, and the risk score and risk level of CVDs. I just love the data in this one. There's just enough data here to be scientifically valid, and just enough data to be meaningful when we talk about prevention.

Thus, in our situation, how we were able to formulate our plan was first by examining the distribution of systolic blood pressure, then, by means of steps two and three, we had to examine if people with diabetes had indeed higher heart disease risk scores.

To our knowledge, no prior study has been published in CVD prevalence using this dataset. This study aims to (i) examine the association between cardiovascular disease risk and biological risk factors (blood pressure, diabetes status, body mass index, and total cholesterol), (ii) assess the relationship between cardiovascular disease risk and smoking status, (iii) evaluate age and sex-related differences in cardiovascular disease risk.

METHODS

Study population and study design

The Cardiovascular Assessment and Intervention Research–Cardiovascular Disease 2025 (CAIR-CVD-2025) dataset was obtained from the Mendeley Data repository (Nirob et al., 2025, March 3). CAIR-CVD-2025 is a cross-sectional, representative sample of 1,026 patients from Jamalpur Medical College Hospital, Jamalpur, Bangladesh. Enrollment occurred between January 20, 2024, and January 1, 2025. All participants were hospital patients aged 18 years or older, and demographic, clinical, and lifestyle data were collected. CVD risk factors data, including blood pressure, body mass index, diabetes status, smoking status, total cholesterol, sex and age were assessed. Patients lacking data on cardiovascular risk level, blood pressure, body mass index, height, weight, total cholesterol, smoking status, diabetes status, sex or age were excluded from the analysis. The final sample included 523 female and 503 male patients, with ages of 47 and 46 years, respectively. The study applied the same age-adjustment method used by National Center for health Statistics (NCHS) for cardiovascular risk factors studies (CDC 2001, CDC 2025), ensuring comparability with national surveillance statistics. Age was grouped into three categories: 20-39 years, 40-59 years, and 60 years or older. Body mass index (BMI) was derived by dividing body weight into kilograms by the square of height in meters (American Heart Association (AHA), n.d.) Then, BMI was categorized as underweight (<18.5 kg/m^2), healthy weight (18.5–25 kg/m^2), overweight (25–30 kg/m^2), or obese (>30 kg/m^2). Body mass index categorizations were based on clinical guidelines from American Heart Association (AHA, n.d). A reaffirmed blood pressure guideline in 2025 by American College of Cardiology(ACC) /American Heart Association (AHA) were used for blood pressure classification (ACC, 2025, October 1), as normal (<120/80 mmHg), elevated (120-129/<80 mmHg), stage 1 hypertension (130-139/80-89 mmHg), stage 1 hypertension (130-139/80-89 mmHg), and stage 2 hypertension (≥140/90 mmHg). Total cholesterol was categorized as lower

risk (200 mg/dL), at-risk (200-239 mg/dL), and higher risk (240 mg/dL) (ACC, 2025, October 1). Diabetes status was defined as a binary variable, coded yes if diabetes was present and no if absent. Smoking status was defined as a binary variable, coded yes for current smokers and no for non-smokers. Age and sex were reported by the patients. The dataset contained variable names with brief description but did not include detailed information on recruitment procedures, materials, and measurement protocols.

Determination of cardiovascular disease risk scores and cardiovascular risk levels

The CAIR-CVD-2025 study reported cardiovascular risk both as a numeric score and as categorical levels, stratified into three groups: low, intermediate, and high (Nirob et al., 2025, March 3). However, detailed on the methodology used to compute the CVD risk score and the threshold guidelines for defining CVD risk level were not specified. Cardiovascular disease score and CVD risk level were defined based on observed associations with established CVD risk factors, including blood pressure, total cholesterol, BMI, diabetes status, smoking status. Based on the Framingham Heart Study and SCORE2 models, CVD risk levels were derived by converting 10-year CVD risk percentage, calculated from CVD risk scores (Framingham Heart Study, n.d., & SCORE2 Working Group & ESC Cardiovascular Risk Collaboration, 2021). Thus, to determine whether the CVD risk score was associated with CVD risk level, CVD risk score distributions across categories were assessed.

To evaluate whether higher CVD risk scores link higher cardiovascular risk, the association between CVD risk scores and established CVD risk factors known to influence CVD were examined. Based on these observed associations, higher CVD risk scores were assumed to correspond to higher relative CVD risk. The CVD risk score was treated as a relative measure. Similarly, the association between CVD risk levels and established risk

factors known to affect CVD were assessed. To evaluate the association between CVD risk levels and the risk factors, the distribution of CVD risk levels across six risk factors, blood pressure, BMI, total cholesterol, smoking status, diabetes status, and age, were evaluated. These assumptions were made by evidence from prior studies and established guidelines (CDC, 2024).

Determination of cardiovascular risk factors

In accordance with WHO laboratory-based cardiovascular risk charts for low- and middle-income countries, blood pressure, total cholesterol, smoking status, diabetes status, body mass index (BMI), age and sex as risk factors were selected (WHO, 2019) BMI was also included as a risk factor in accordance with non-laboratory CVD risk prediction list (WHO, 2019). Adding BMI allows this study to cover all relevant risk factors in both laboratory and non-laboratory CVD risk charts, accounting for their relative importance in the regional variability.

Data Cleaning and Validation

Data cleaning and validation were performed in response to inconsistency patterns observed during the pilot test. Height and weight variables were cross-checked, and originally recorded BMI values were compared against calculated BMI derived from height and weight. The calculated BMI values were not consistently aligned with the BMI measurements provided (Supplementary Figure 1). Body mass index was recalculated based on the recorded height and weight measurements and was categorized based on clinical cutoff guidelines. Blood pressure data were provided in multiple formats, including combined measures (systolic/diastolic), individual systolic and diastolic values, and categorical blood pressure classifications. Missing

values and inconsistencies between measured blood pressure values and the corresponding categorical classifications were assessed.

Ethics

The Cognitive AI & Informatics Research Lab (CLAIR Lab) claimed that the data collection complied with ethical research practices and adhered to applicable laws to safeguard patient rights and confidentiality (Nirob et al., 2025, March 3). Data quality control and validation were maintained under supervised oversight (Nirob et al., 2025, March 3).

Statistical Analysis

Summary statistics were calculated for demographic characteristics, BMI categories, blood pressure categories, total cholesterol categories, smoking status, and diabetes status, with P values computed using Pearson's Chi-squared test (Table 1).

Data was converted to clinical categories of BP, then plots were made based on each question. A pie chart was made to show distribution, a density plot to show comparisons of distributions, then a simple scatter plot with simple regression lines to display correlation between variables. A person r and p-value were added to the scatter plot.

A pie chart was used to show how normal, elevated, stage 1 HTN, and stage 2 HTN systolic blood pressure categories distribute within the study population.

Density Plots, one density plot showing density of CVD Risk Score of Diabetics and the other density plot showing density of CVD Risk Score of Non-Diabetics, will be used to answer the question of whether there is a difference in CVD risk scores between diabetics and non-diabetics.

We used clinical thresholds to represent BP values, a pie chart to represent proportions, a density plot to compare group distributions, and a stratified scatter plot with Pearson's r to

evaluate strengths and directions of associations. Our approach offers the best compromise between understandability and scientific rigor in terms of audience.

Univariate regression was performed to assess whether CVD risk score was associated with selected CVD risk factors, including total cholesterol, blood pressure, BMI and age. Independent t-test was performed to assess whether CVD risk score was associated with diabetes and smoking status. ANOVA test was performed to evaluate mean difference of CVD risk score across age groups. We generated a table and figures and conducted statistical analysis using R software, version 4.5.1 (R Core Team, 2024).

RESULTS

Demographic, Biological, and Behavioral characteristics of CAIR CVD 2025 Participants

A total of 1,529 adults (aged >18 years) enrolled in the CAIR CVD 2025 study at Jamalpur Medical College Hospital from January 20, 2024, to January 1, 2025. Of these, 1,026 (67%) had biological and behavioral data available and were included in this analysis (Table 1). The study population was approximately equally distributed by sex, with 51% female and 49% male participants (Table 1). Overall, participants showed a high cardiovascular disease risk profile: diabetes affected nearly half (49%), obesity prevalence was 41%, and hypertension at Stages 1 and 2 was present in 74% (Table 1). Analysis revealed statistically significant associations between CVD risk level and sex, age, blood pressure category, body mass index, total cholesterol group, diabetes status, and smoking status (Table 1).

Cardiovascular Risk Score and Cardiovascular Risk level Are Independent Risk Measurements

We evaluated whether cardiovascular risk score is associated with cardiovascular risk level. The intermediary risk group had the lowest median CVD risk score, suggesting potential

overlap in the classification of CVD risk level (Figure 1). To further explore this association, we used violin plots to visualize the distribution of CVD risk scores across CVD risk levels. All three CVD risk levels showed wide range of distribution that intersected between all three levels (Figure 2). Therefore, we affirmed that cardiovascular risk scores are independent indicators within the risk assessment framework.

Association Between Cardiovascular Risk Level and Cardiovascular Risk Factors

We observed the association between cardiovascular disease risk level and established six CVD risk factors. Across all six panels, patients with known risk factors showed higher proportion of high CVD risk level category (Figure 3). Patients with diabetes (a) and smoker (b) showed higher proportions in the high CVD risk level category, with 57.9% and 59.4% respectively. BMI class (c) showed obesity and overweight groups showing over 55% in the high CVD risk level category (Figure 3). Blood pressure class (d) showed higher proportion of high CVD risk level across blood pressure class, and 40-59 age group (e) had the highest proportion of high CVD risk level category (55.9%) (Figure 3). Finally, total cholesterol (f) confirmed that individuals with high risk of total cholesterol had the highest proportion of high CVD risk level (59.9%) (Figure 3). These patterns affirm that CVD risk levels reflect similar trends observed across established CVD risk factors.

Association Between Cardiovascular Risk scores and Cardiovascular Risk Factors

We evaluated the association between CVD risk score and four established risk factors, total cholesterol, systolic blood pressure, body mass index (BMI), and age. Total cholesterol ($R = 0.52$, $p < 0.001$), systolic blood pressure ($R = 0.48$, $p < 0.001$), and BMI ($R = 0.39$, $p < 0.001$), showed statistically significant association with CVD risk score (Figure 4). Age was not

associated with CVD risk score in this study (R = 0.03, P = 0.29). Cardiovascular risk score reflects overall trends observed across established CVD risk factors.

Distribution of Total Cholesterol Across CVD Risk Levels

The stacked bar chart showed that patients with high CVD risk level were disproportionately represented in the higher cholesterol ranges (Figure 5). Conversely, low risk patients clustered in the lower cholesterol ranges (Figure 5).

Sex-Based Distribution Across CVD Risk Levels

The proportion of female patients increased with higher CVD risk level, while the proportion of male patients decreased with higher CVD risk level (Figure 6). The inverse pattern suggests females were disproportionately represented in higher CVD risk level categories, consist with the distribution observed in Table 1 (Table 1).

Systolic Blood Pressure Category Distribution

When BP is disaggregated, out of every five people, two fall into the normal category, whereas almost one-fourth of the people fall into Stage 2 hypertensive level (Table 7). The remaining people fall to higher and crisis levels. All in all, this pie chart shows that high BP is a big problem.

Cardiovascular Risk Score Distribution and Diabetes Status

The Diabetes density curve reaches a greater max height compared to the non-diabetes density curve (Figure 8). The mean differences of diabetes status groups are statistically significant (P = 0.001). The fact that diabetes amplifies cardiovascular risk score is proven when comparing these two density plots.

Cardiovascular Risk Score by Smoking Status

We compare CVD risk scores between smoking and non-smoking patients. The mean risk score among non-smoker was 16.95, while smokers had a mean score of 17.11(Figure 9). The mean difference between the two groups was not statistically significant (P = 0.296) (Figure 9). Thus, smoking status was not associated with CVD risk score within this study.

Age Is Not Correlated with Cardiovascular Risk Score

Cardiovascular risk scores across three age groups were assessed. Boxplots showed overlapping distribution CVD risk score across age groups (Figure 10). ANOVA test for mean differences was not statistically significant (P = 0.082), suggesting age group alone did not explain variability of cardiovascular risk scores (Figure 10).

Interaction Analysis of BMI and Sex in Relation to Cardiovascular Risk score

We assessed the association between BMI and CVD risk score, stratified by sex. The regression lines showed a positive relationship between BMI and CVD risk score for both sex groups (Figure 11). The interaction term between BMI and sex was not statistically significant (P = 0.642), suggesting BMI has same effect on CVD risk score in both male and female patients (Figure 11).

Conclusion

There are two themes. One, BP is doing heavy lifting when it comes to predictive factors in this group of patients. Two, the risk factor of Diabetes increases risk notably. The overall message is to first control BP, then prevent diabetes, then focus on getting sufficient exercise after all that.

Several notable limitations of the dataset included the fact that the data was collected only from individuals in only one setting. Also, there were not any variables in the data related to

medication and diet. A final limitation could have been potential self-report bias. We were dealing with cross-sectional data, and therefore we cannot conclude anything significant with respect to causation. Next time, what we would do would be to gather more data with respect to their diet and drugs, and then longitudinal data with respect to the effect of treatment and lifestyle change on risk.

Thus, in conclusion, the higher the person's blood pressure is, the higher his or her chances of cardiovascular diseases even if he or she is active. But lifestyle also plays an important part, and if one actually does desire to work on what can be altered, then one must work on lifestyle change and then screening.

Taken together, the results strongly emphasize the need for BP management, with added benefit from controlling diabetes. The primary limitations involve the fact that data comes from one center, data does not include specifics on medications and diets, and data was obtained in a cross-sectional manner. Future trials should involve data on treatment and diets, especially prospective studies tracking patients longitudinally.

The findings of our study are consistent with previous studies, confirming that elevated cholesterol levels, high blood pressure, diabetes, and high BMI values are significantly associated with increased risk of cardiovascular disease. Our finding also suggests that cardiovascular risk is not affected by age group or modified by sex in relation to BMI. These findings are concerned in hospital-based patient populations, with the high prevalence of obesity and hypertension observed in this study. The implications extend beyond the hospital-based study, as similar patterns may be expected in other populations with high obesity and hypertension. Moreover, our hospital-based sample was heterogeneous with respect to age, with a lower proportion of patients aged 65 years and older. This pattern differs from hospitals, where

older adults represent the largest age group among patients in United States (CDC, 2020). As a result, our findings may underestimate the contribution of advanced age to cardiovascular risk, and the age distribution in our study sample could limit the generalizability of these results to hospital populations that include a higher proportion of older patients. Another factor to acknowledge is the diversity within hospital-based patient populations. Cardiovascular risk may differ depending on whether patients are inpatients, outpatient, or emergency department patients. Lastly, the presence of existing comorbidities of patients could influence CVD risk.

This study is subject to several limitations that should be acknowledged. The absence of study design, materials, detailed procedures may limit the validation of our findings. The definitions of cardiovascular risk score and CVD risk level were not fully validated, and our assumptions may not have been aligned with those used in the original studies from which the risk measures were derived. Such discrepancies may bias the interpretation of our findings. Finally, issues of data consistency may have affected our analyses, potentially introducing bias and reduced accuracy.

Future studies on hospital-based cardiovascular disease risk assessment are crucial in low- and middle-income countries, as those studies can help understand unique variation in risk factors in specific regions and hospital settings. Hospital patients are especially vulnerable to cardiovascular risk and require short-term evaluations to predict their health outcomes. Therefore, future research could focus on developing regional specific and applicable risk models that allow flexible adjustment to diverse patient populations.

References

American Heart Association. (n.d.). *Body mass index in adults*. Retrieved December 11, 2025, from https://www.heart.org/en/healthy-living/healthy-eating/losing-weight/bmi-in-adults

American College of Cardiology. (2025, October 1). *New in clinical guidance: High blood pressure*. Retrieved December 11, 2025, from https://www.acc.org/latest-in-cardiology/articles/2025/10/01/01/new-in-clinical-guidance-hbp

Centers for Disease Control and Prevention, National Center for Health Statistics. (2025, September). *Prevalence of cardiovascular disease risk factors in adults: United States, August 2021–August 2023* (NCHS Data Brief). U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention. (2024, June 3). *Cardiovascular disease*. Retrieved December 11, 2025, from https://www.cdc.gov/cdi/indicator-definitions/cardiovascular-disease.html#cdc_data_surveillance_section_2-definition-details

Centers for Disease Control and Prevention. (2020). *Hospitalization*. National Center for Health Statistics. U.S. Department of Health & Human Services. https://www.cdc.gov/nchs/hus/topics/hospitalization.htm

Centers for Disease Control and Prevention. (2001). *Age adjustment using the 2000 projected U.S. population* (Statistical Notes, No. 20). National Center for Health Statistics. Retrieved December 11, 2025, from https://www.cdc.gov/nchs/data/statnt/statnt20.pdf

Epsi, N. (2025). Session 3-6: Health Data Visualization. Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, The Geroge Washington University.

Framingham Heart Study. (n.d.). *Cardiovascular disease (10-year risk)*. Retrieved December 12, 2025, from https://www.framinghamheartstudy.org/fhs-risk-functions/cardiovascular-disease-10-year-risk/

GBD 2023 Cardiovascular Disease Collaborators. (2025). Global, regional, and national burden of cardiovascular diseases and risk factors in 204 countries and territories, 1990–2023. *Journal of the American College of Cardiology.* Advance online publication. https://doi.org/10.1016/j.jacc.2025.09.540

Nirob, M. A. S., Bishshash, P., SIAM, A. K. M. F. K., et al. (2025, March 3). *CAIR-CVD-2025: An extensive cardiovascular disease risk assessment dataset from Bangladesh* (Version 1). Mendeley Data. https://doi.org/10.17632/d9scg7j8fp.1

R Core Team. R: A language environment for statistical computing. Vienna Austria: R Foundation for Statistical Computing, 2024.

SCORE2 Working Group & ESC Cardiovascular Risk Collaboration. (2021). SCORE2 risk prediction algorithms: New models to estimate 10-year risk of cardiovascular disease in Europe. *European Heart Journal, 42*(25), 2439–2454. https://doi.org/10.1093/eurheartj/ehab309

WHO CVD Risk Chart Working Group. (2019). World Health Organization cardiovascular disease risk charts: Revised models to estimate risk in 21 global regions. *The Lancet Global Health, 7*(10), e1332–e1345. https://doi.org/10.1016/S2214-109X(19)30318-3

World Health Organization. (2025, July 31). *Cardiovascular diseases (CVDs)*. Retrieved December 11, 2025, from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29

Yang, W.-Y., Lip, G. Y. H., Sun, Z.-J., Peng, H., Fawzy, A. M., Li, H.-W., & The Cardiovascular Center Beijing Friendship Hospital Database Bank (CBD Bank) Study Group. (2022). *Implications of new-onset atrial fibrillation on in-hospital and long-term prognosis of patients with acute myocardial infarction: A report from the CBD bank study. Frontiers in Cardiovascular Medicine, 9,* 979546. https://doi.org/10.3389/fcvm.2022.979546

Characteristic	N	Overall N = 1,026[1]	LOW N = 145[1]	INTERMEDIARY N = 380[1]	HIGH N = 501[1]	p-value[2]
Sex	1,026					0.11
Female		523 (51%)	63 (43%)	193 (51%)	267 (53%)	
Male		503 (49%)	82 (57%)	187 (49%)	234 (47%)	
Age	1,026					<0.001
20-39		330 (32%)	49 (34%)	145 (38%)	136 (27%)	
40-59		546 (53%)	58 (40%)	183 (48%)	305 (61%)	
60+		150 (15%)	38 (26%)	52 (14%)	60 (12%)	
Blood Pressure Category	1,026					<0.001
Normal		199 (19%)	19 (13%)	79 (21%)	101 (20%)	
Elevated		65 (6.3%)	4 (2.8%)	28 (7.4%)	33 (6.6%)	
Hypertension Stage 1		322 (31%)	30 (21%)	123 (32%)	169 (34%)	
Hypertension Stage 2		440 (43%)	92 (63%)	150 (39%)	198 (40%)	
Body Mass Index	1,026					<0.001
Underweight		109 (11%)	17 (12%)	46 (12%)	46 (9.2%)	
Healthy Weight		287 (28%)	51 (35%)	132 (35%)	104 (21%)	
Overweight		208 (20%)	26 (18%)	66 (17%)	116 (23%)	
Obesity		422 (41%)	51 (35%)	136 (36%)	235 (47%)	
Total Cholesterol	1,026					<0.001
Lower Risk		506 (49%)	69 (48%)	236 (62%)	201 (40%)	
At-Risk		201 (20%)	33 (23%)	59 (16%)	109 (22%)	
Higher Risk		319 (31%)	43 (30%)	85 (22%)	191 (38%)	
Diabetes Status	1,026	525 (51%)	67 (46%)	154 (41%)	304 (61%)	<0.001
Smoking Status	1,026	535 (52%)	69 (48%)	148 (39%)	318 (63%)	<0.001

[1] n (%)

[2] Pearson's Chi-squared test

Table 1. Demographic and behavioral characteristics of 1,026 patients by CVD risk level.

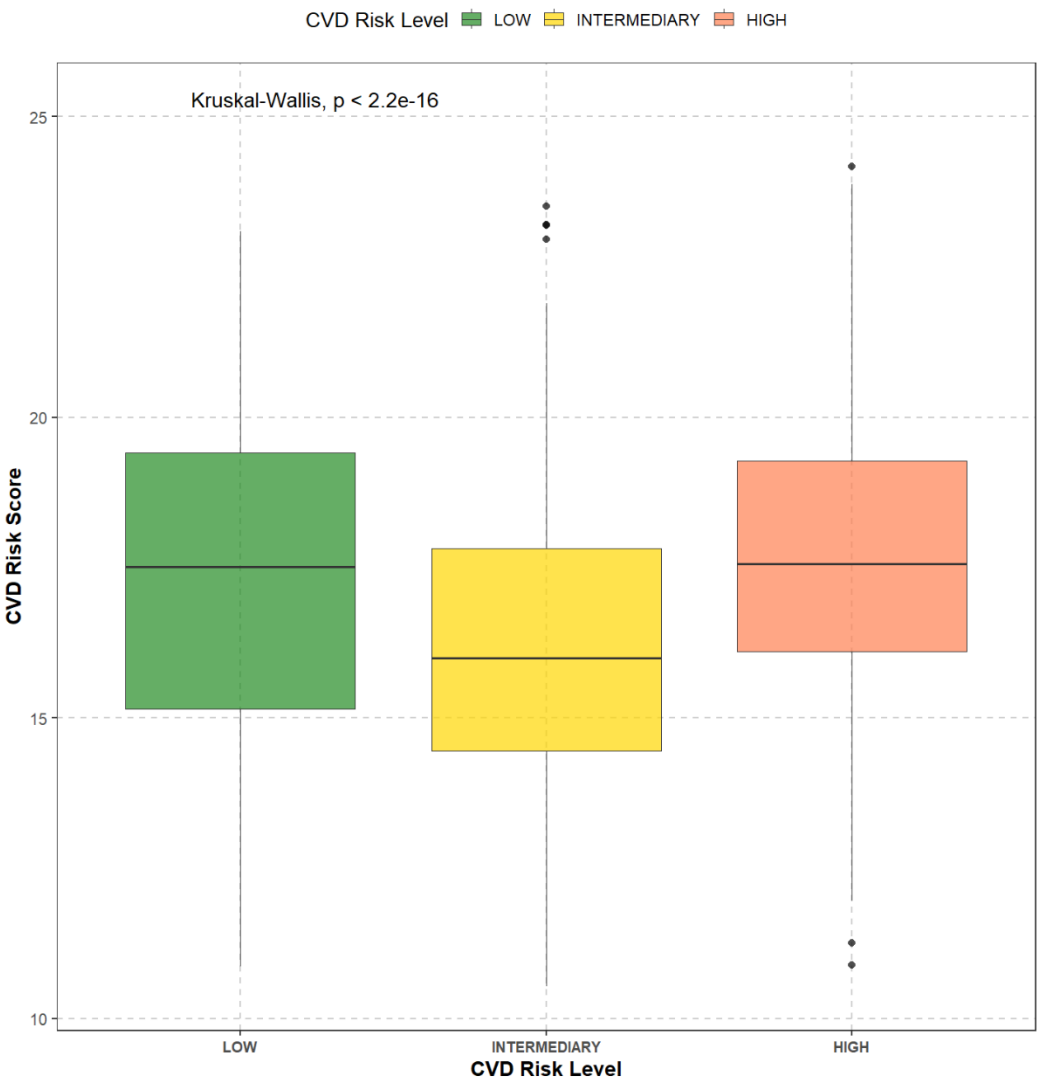


Figure 1. Association between CVD Risk Score and CVD Risk Level. Each dot indicates an outlier. The boxplot displays the median and 25th and 75th percentiles. Statistical significance was assessed with using Kruskal-Wallis test.

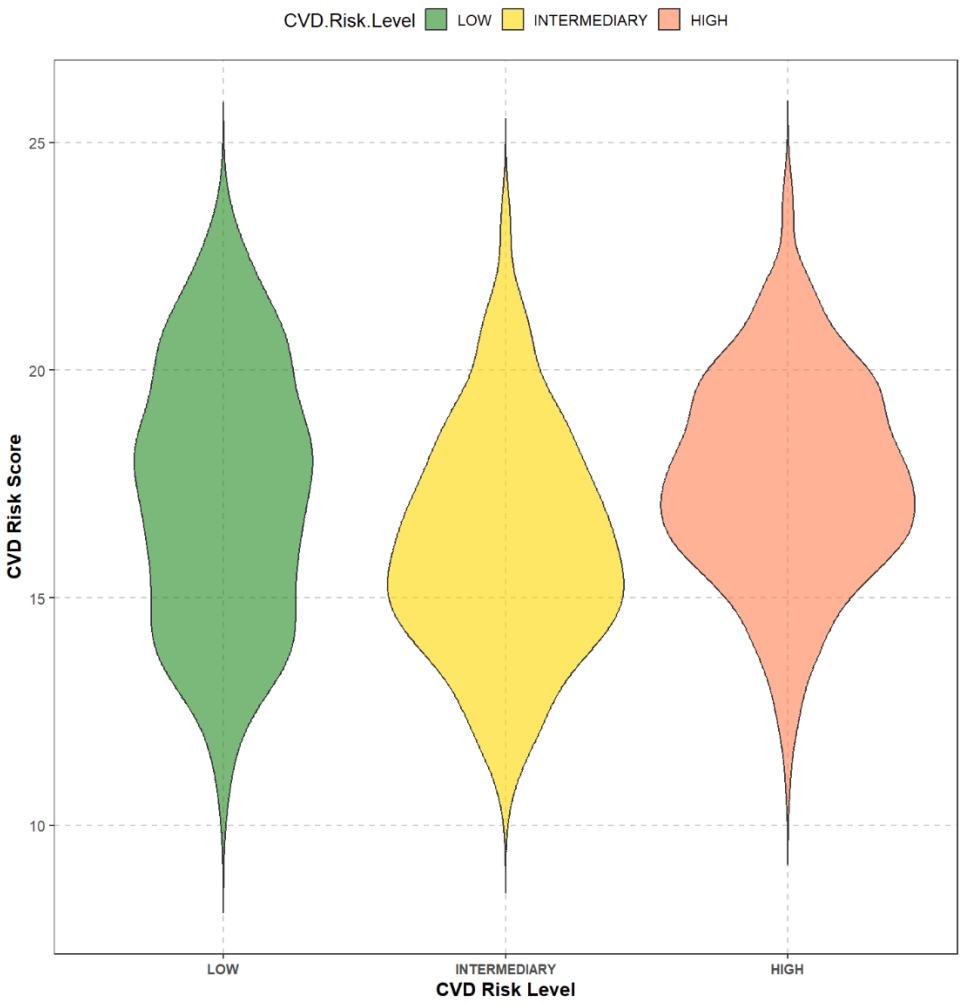


Figure 2. Distribution of CVD score across CVD risk level. Violin plot displays the distribution of CVD risk score across CVD risk level.

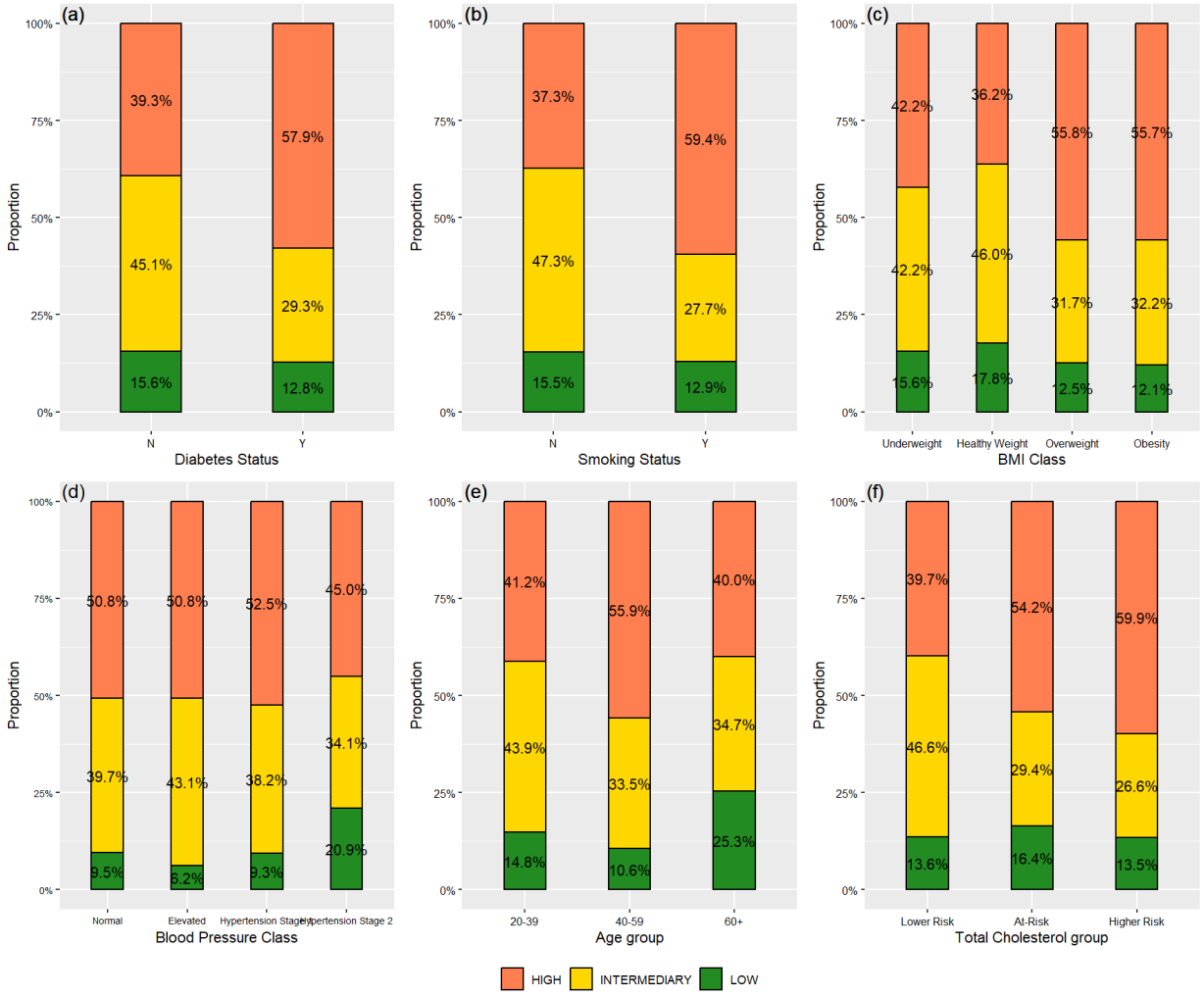


Figure 3. Distribution of CVD risk level across six CVD risk factors, diabetes status (a), smoking status (b), BMI (c), blood pressure (d), age (e), and total cholesterol (f).

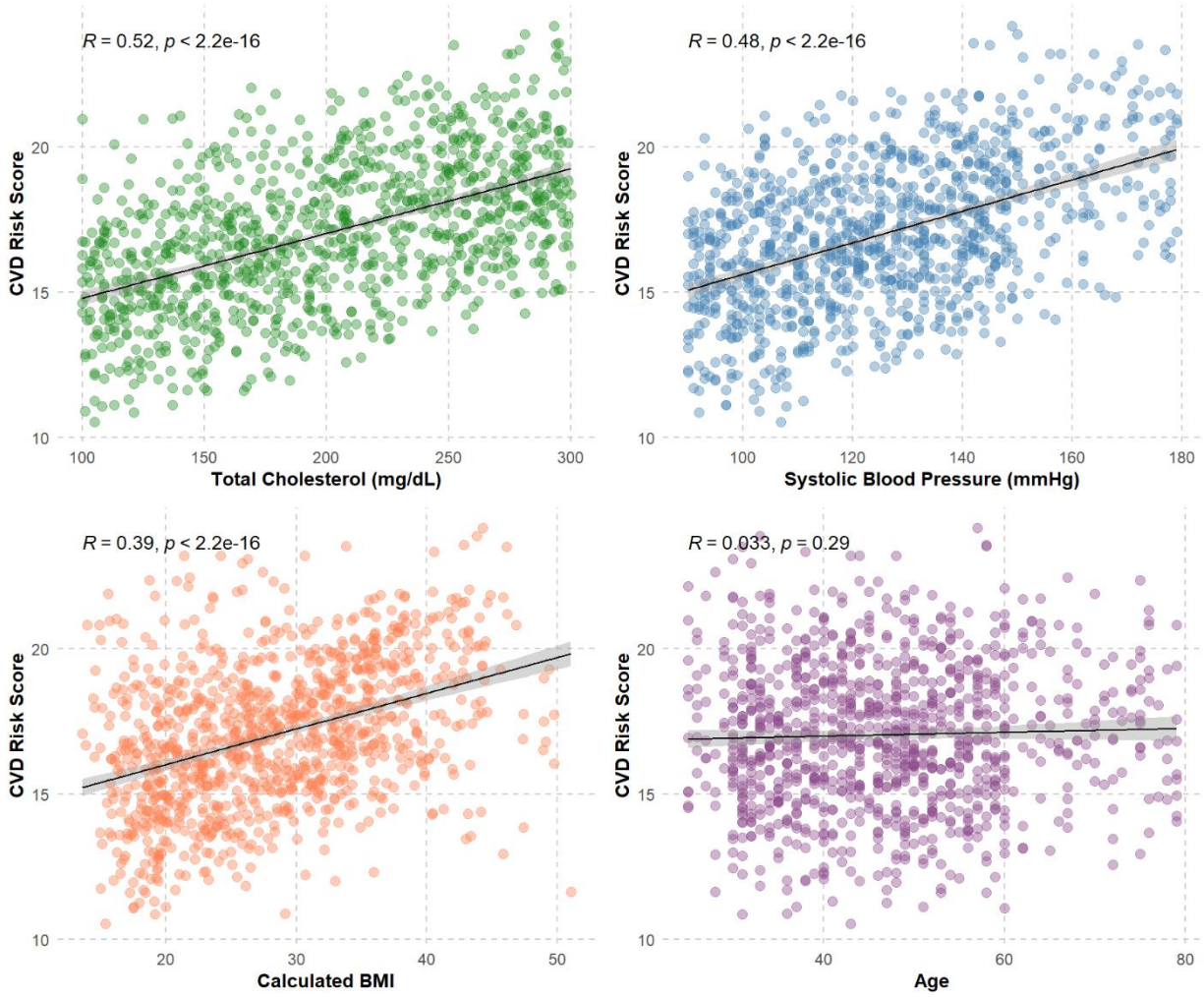


Figure 4. Cardiovascular disease (CVD) risk scores plotted by total cholesterol, systolic blood pressure, BMI, and age. Linear regressions were fit with 95% confidence intervals. Statistical significance was assessed by using t-test for regression coefficients, with Pearson's correlation (r) reported.

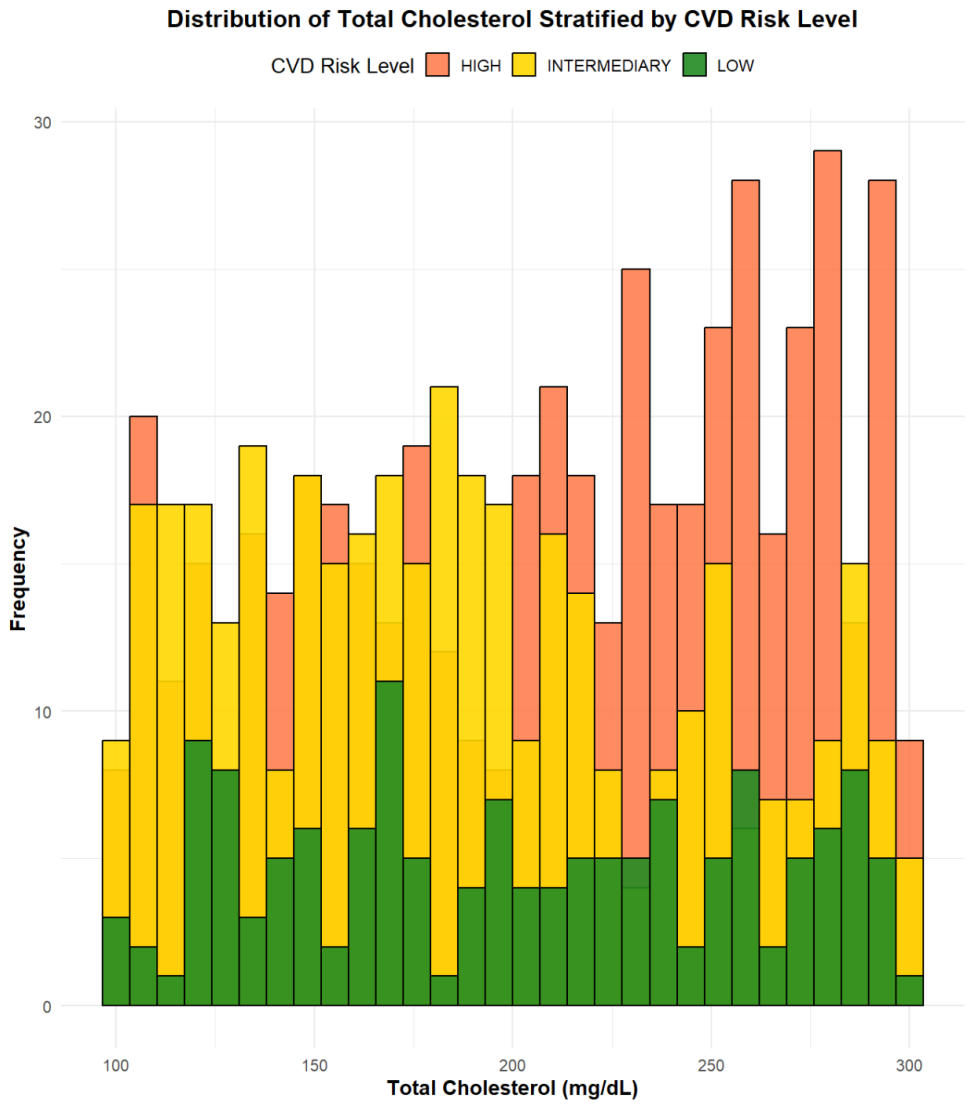


Figure 5. The distribution of total cholesterol levels stratified by cardiovascular disease risk level categories. The y-axis represents frequency, and each bar is segmented according to CVD risk level. The relative frequency of CVD risk level across total cholesterol is illustrated.

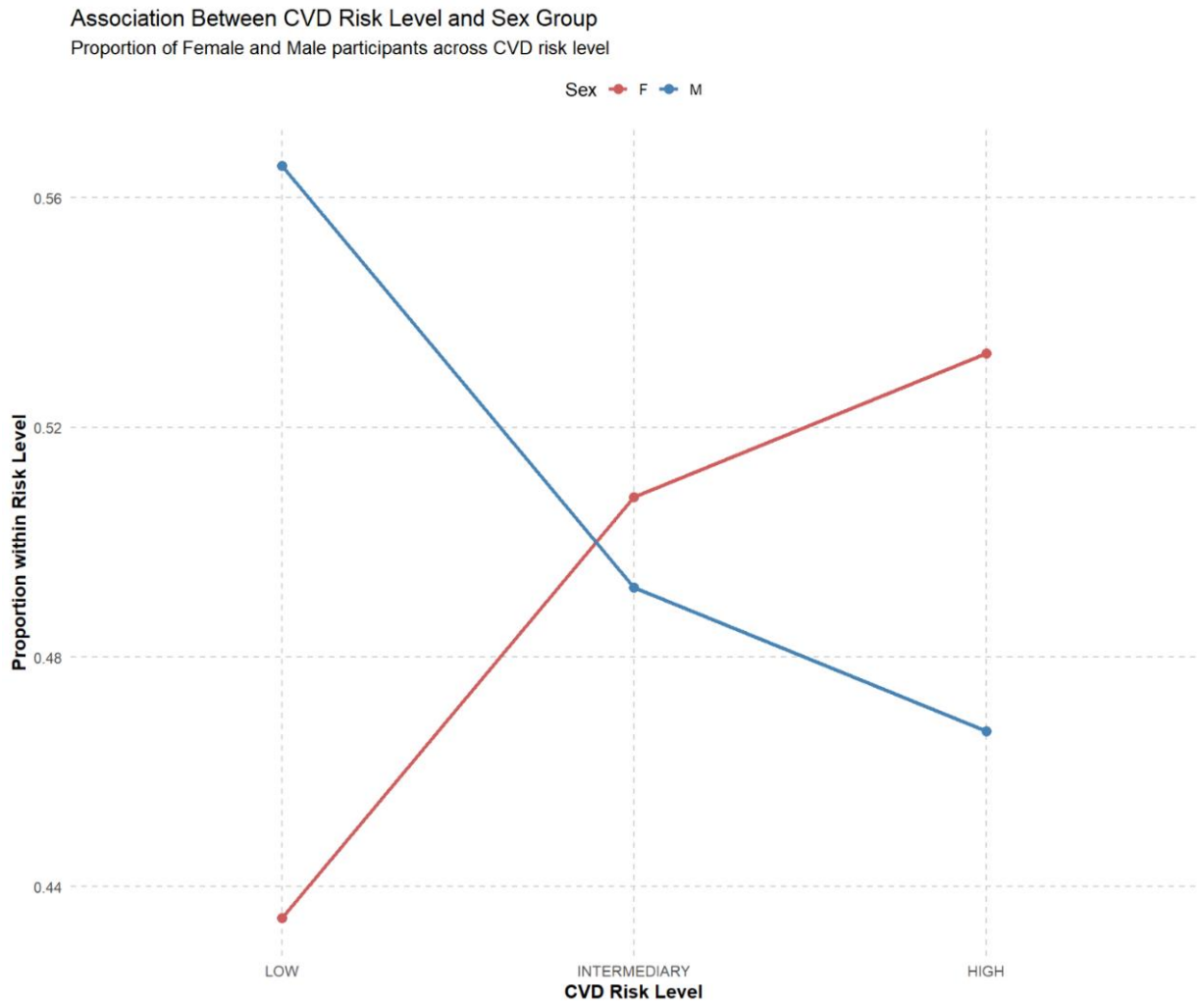


Figure 6. Line plot showing the proportion of patients across cardiovascular disease risk level. The plot is stratified by sex, with blue line represent male and red line for female. The visualization showed differences in the proportion of CVD risk level between sexes.

Pie Chart of Systolic Blood Pressure Category Distribution

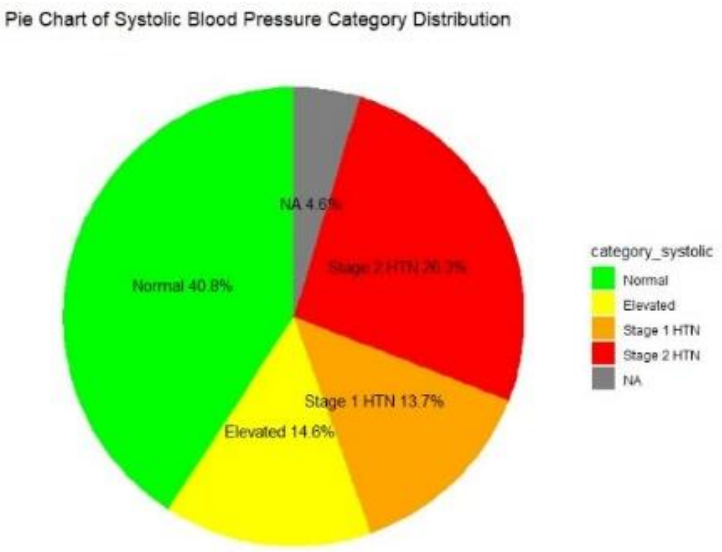


Figure 7. Distribution of patients across blood pressure categories, normal, elevated, stage 1 HTN, stage 2 HTN. The relative frequency of each category was visualized.

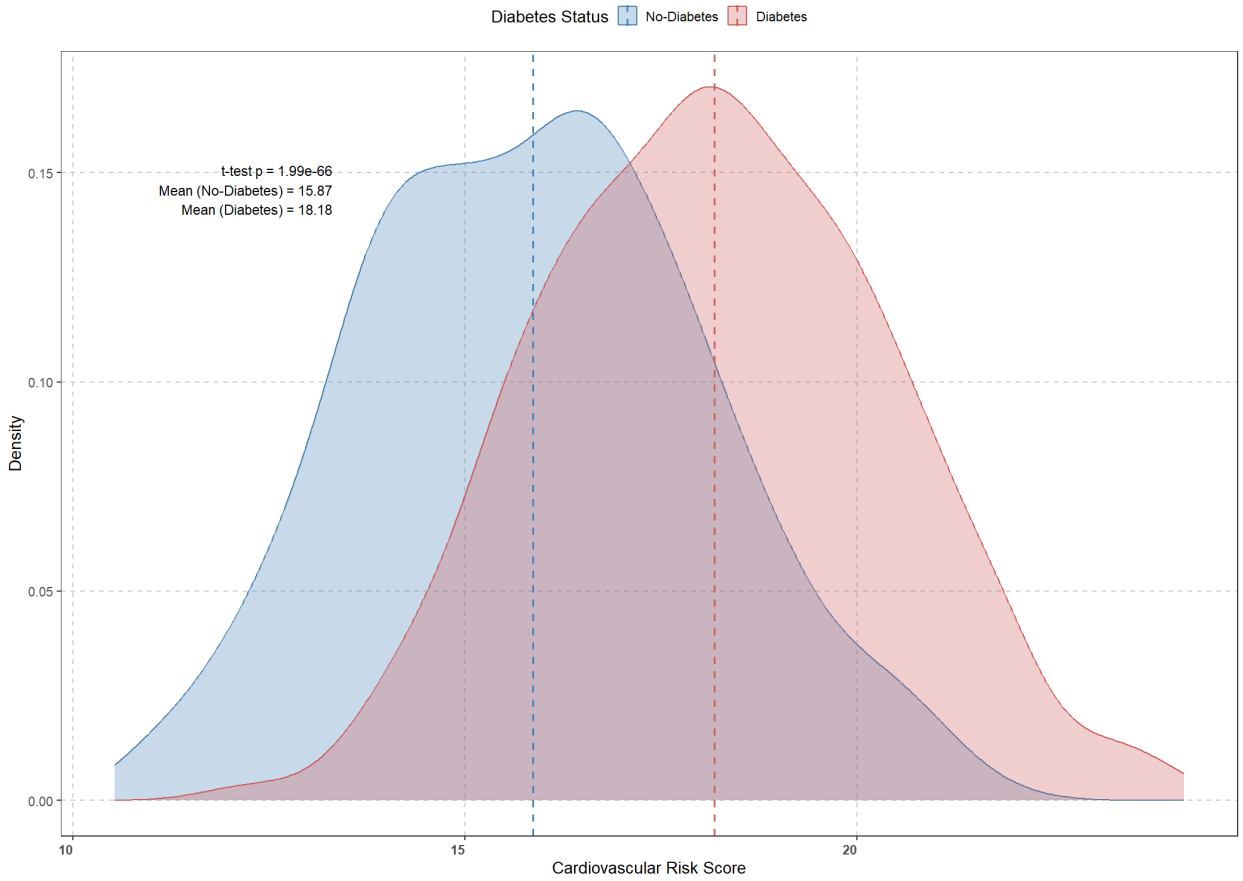


Figure 8. Density plot comparing the distribution of cardiovascular risk score between non-diabetes and diabetes. Vertical dashed lines indicating the group means, and P-value from the independent sample t-test were reported for the mean difference between these two groups.

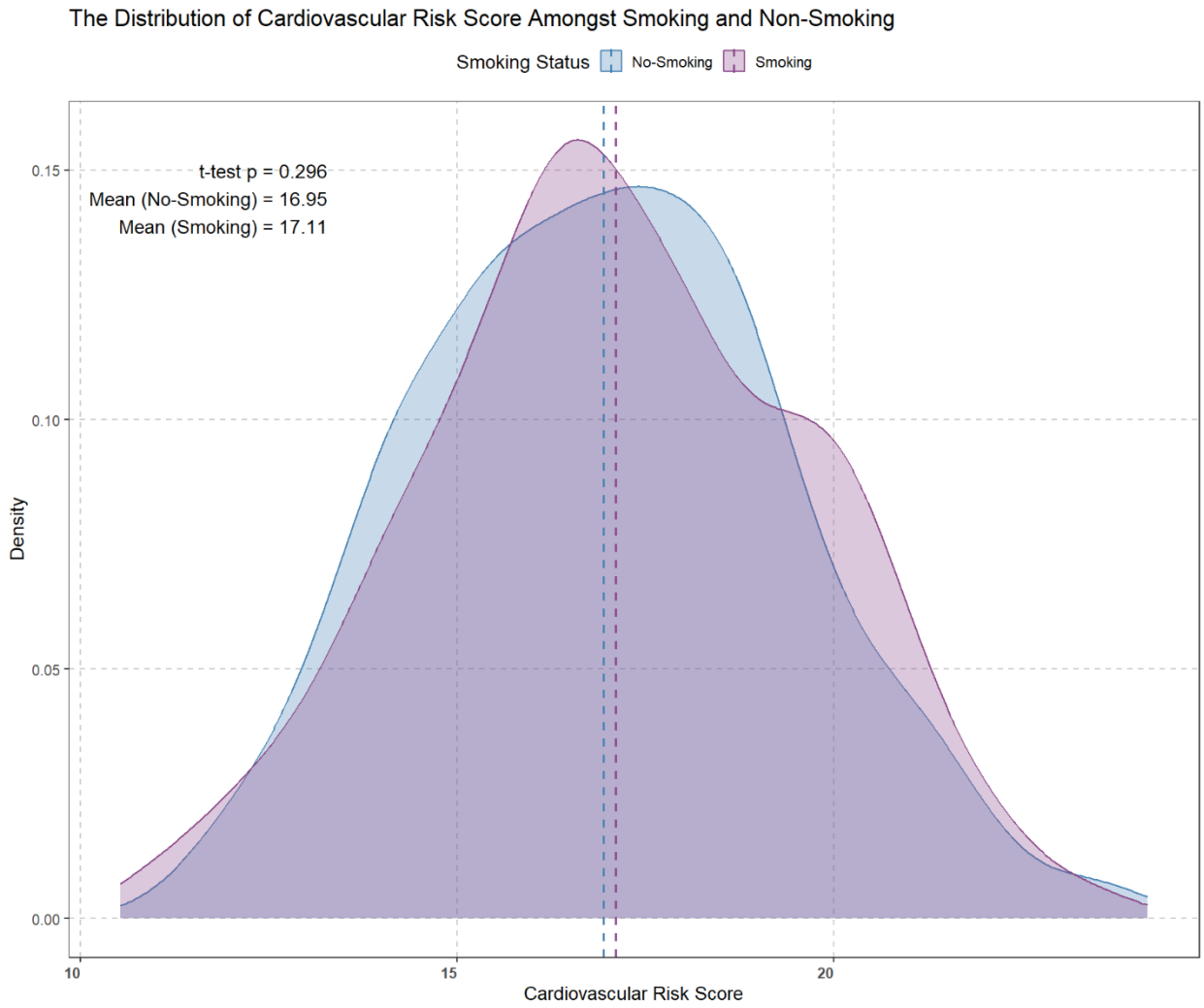


Figure 9. Density plot comparing the distribution of cardiovascular risk score between non-smokers and smokers. Vertical dashed lines indicating the group means, and P-value from the independent sample t-test were reported for the mean difference between these two groups.

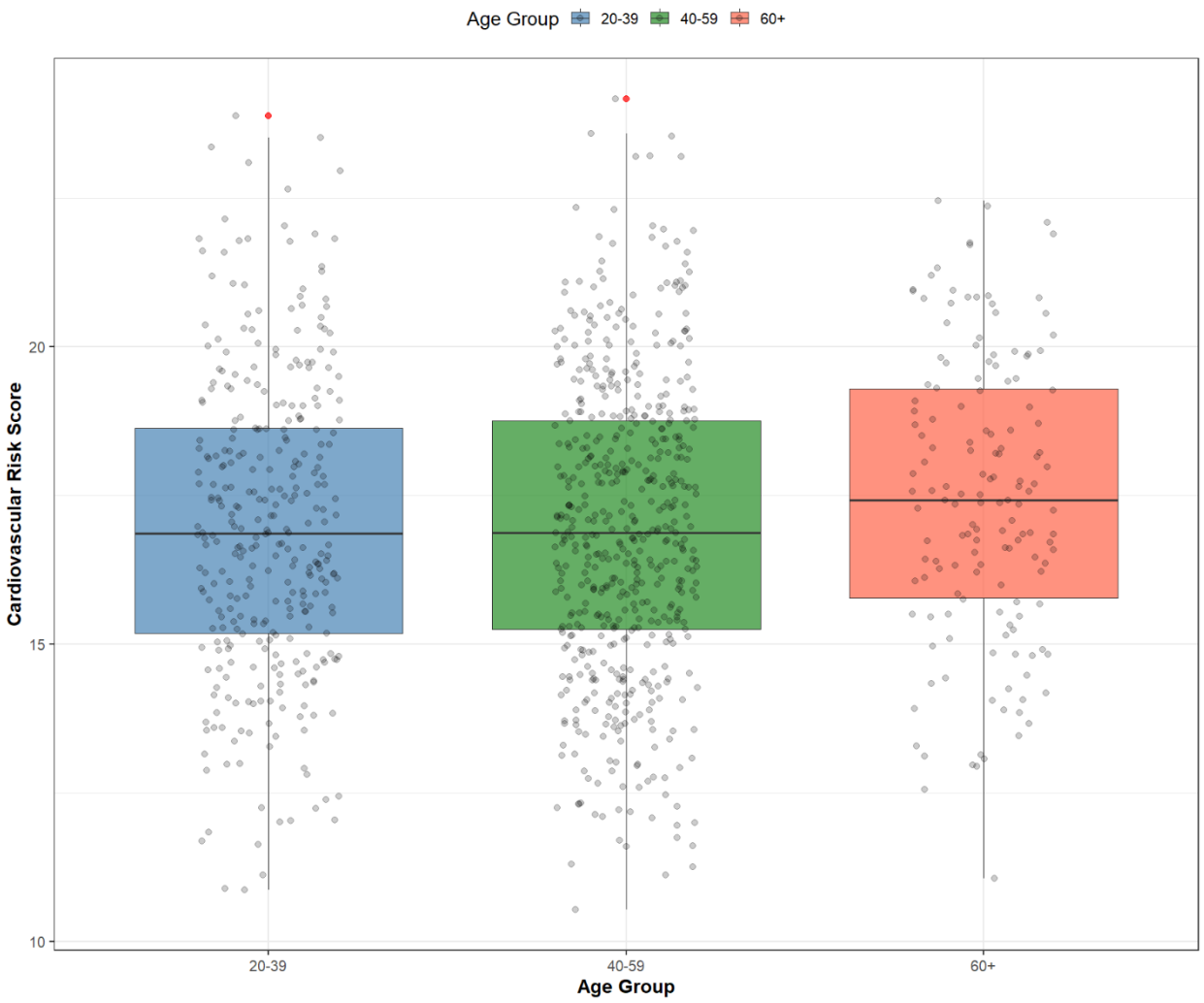


Figure 10. Boxplots display the distribution of cardiovascular disease risk scores across age groups. The boxes represent the means, the first quantile and the third quantile, and individual red dots indicate subjects identified as outliners. An AOVA test was performed to evaluate mean differences in CVD risk scores between age groups, with P-value is displayed.

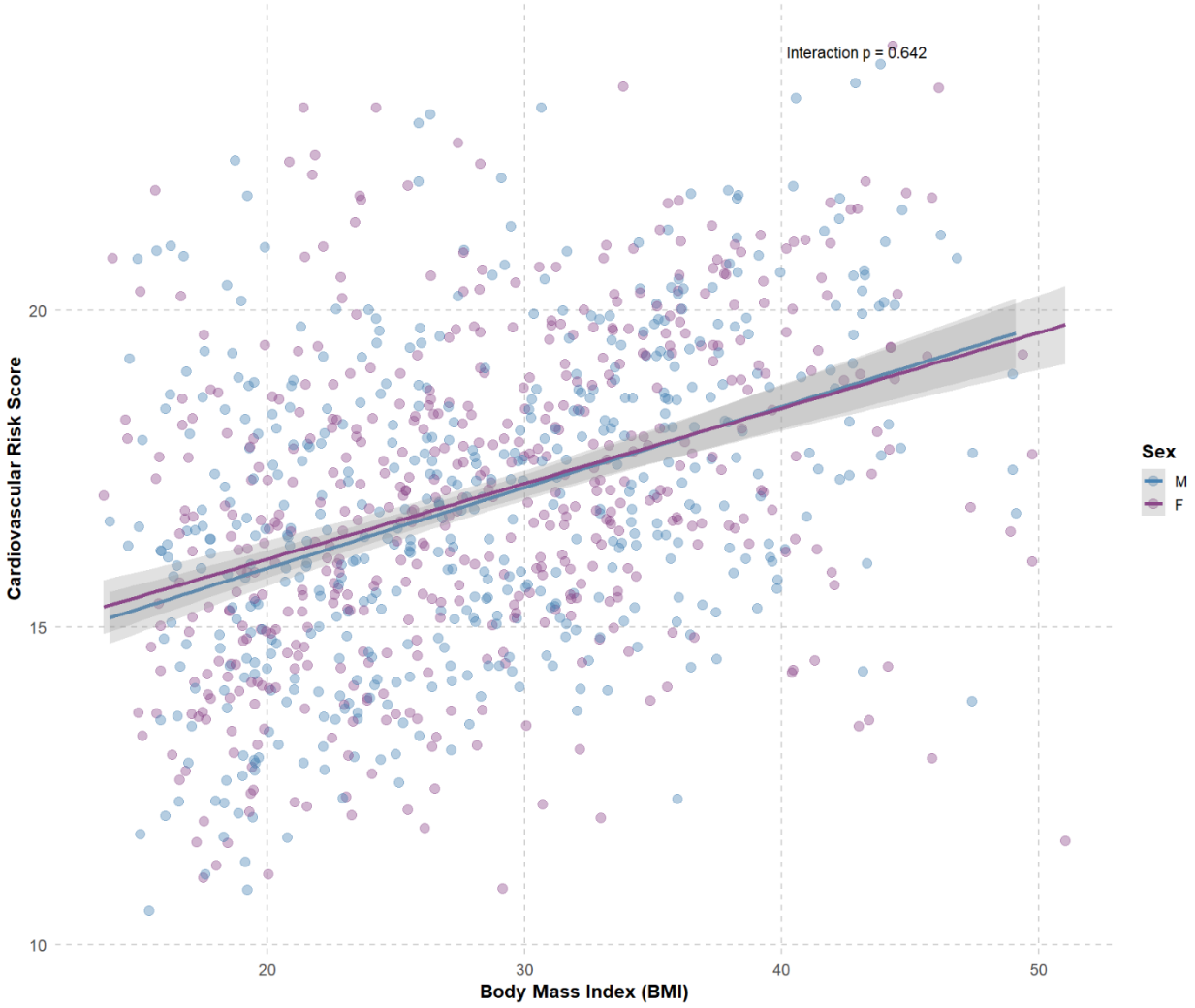
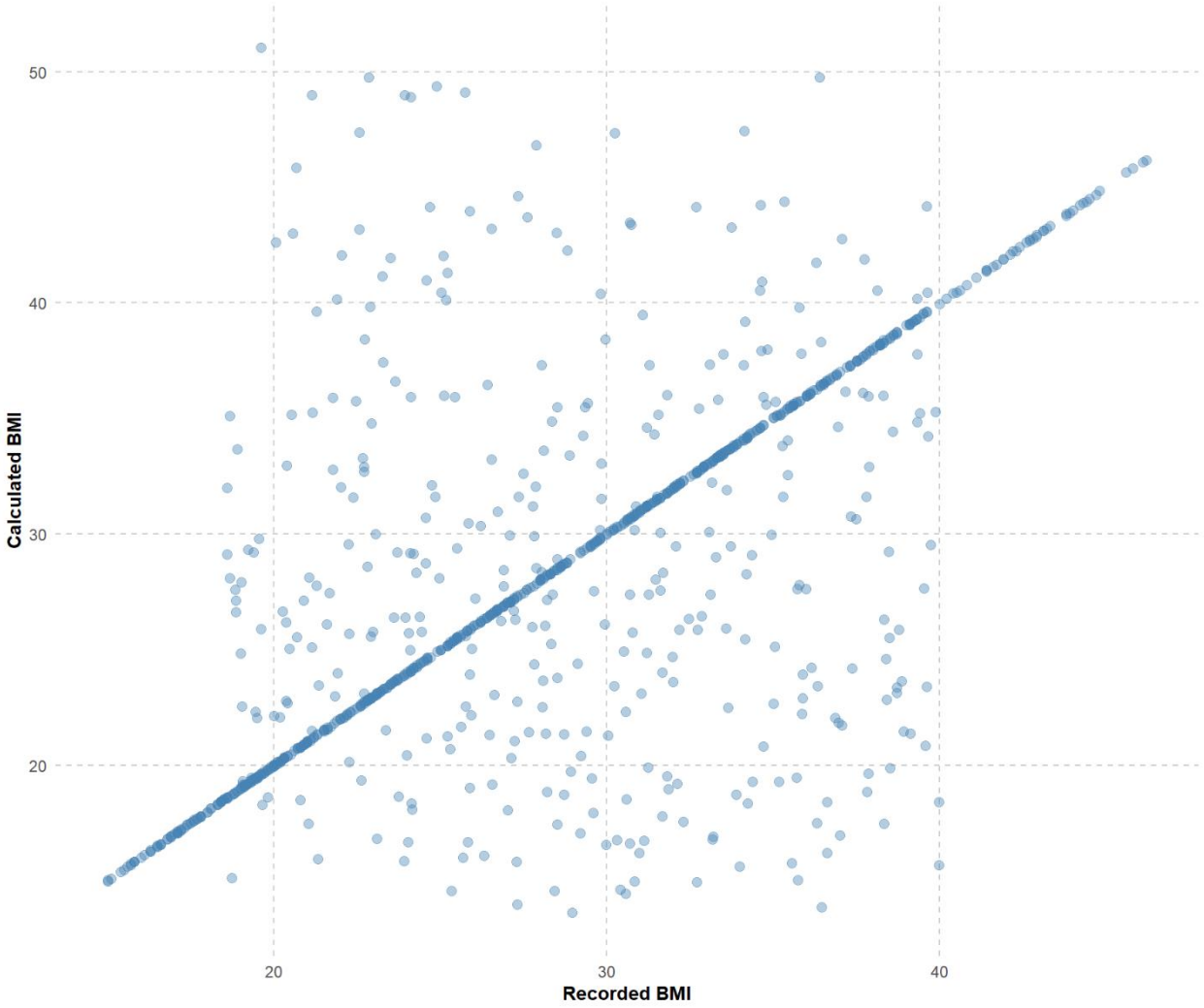


Figure 11. The association between cardiovascular disease risk score and body mass index, stratified by sex. Each point represent patient, males for blue dots and female for purple dots. Linear regression lines were fit to male and female groups with 95% confidence interval. The P-value for the interaction term between BMI and sex is displayed.

Supplemental Figure 1.



Supplementary Figure 1. Comparison of recorded BMI values with BMI calculated with height and weight measurements. Each point represents an individual observation. The substantial inconsistency of BMI values was visualized, suggesting measurement errors or differences in calculation methods.