



Simultaneous Modeling of Cardiovascular Disease and Diabetes Incidence with Age, BMI, Smoking, and Physical Activity as Predictors

Are, Stephen O.^a, Aako, Olubisi L.^{a*} and Ojo, Gabriel O.^a

^a *Department of Mathematics & Statistics, Federal Polytechnic Ilaro, Ogun State, Nigeria.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: <https://doi.org/10.9734/ajrcd/2026/v8i1164>

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://pr.sdiarticle5.com/review-history/161608>

Original Research Article

Received: 15/05/2026

Accepted: 11/07/2026

Published: 16/07/2026

Abstract

This study investigated the simultaneous occurrence of cardiovascular disease (CVD) and diabetes and examined the linear and non-linear effects of age, body mass index (BMI), smoking status, and physical activity on both outcomes while accounting for their possible dependence. Data from 315 participants were analysed using descriptive statistics, correlation analysis, logistic regression, and Vector Generalized Additive Models (VGAMs). The prevalence of CVD and diabetes was 23.81% and 40.95%, respectively, and the co-occurrence of both conditions was 15.24%. Logistic regression showed that age and BMI were significant predictors of diabetes, whereas none of the selected predictors had a statistically significant linear association with CVD. The VGAM identified a significant non-linear relationship between age and diabetes incidence ($p = 0.0134$), while the non-linear effects of age and BMI on CVD were not statistically significant. The estimated residual correlation parameter was 0.552, indicating moderate positive dependence between CVD and diabetes after adjustment for covariates. The diabetes prediction model showed reasonable

*Corresponding author: E-mail: olubisi.aako@federalpolyilaro.edu.ng;

discriminatory ability (AUC = 0.76), whereas the CVD model showed poor discriminatory ability (AUC = 0.51). The VGAM joint model also produced lower AIC and BIC values than the restricted independence model, and the likelihood ratio test confirmed significant dependence between the two outcomes (LRT = 20.60, $p < 0.001$). These findings support the use of joint modelling for related cardiometabolic outcomes.

Keywords: *Vector Generalized Additive Model (VGAM); Cardiovascular Disease (CVD); diabetes mellitus; prevalence; comorbidity analysis.*

1. Introduction

Cardiovascular disease (CVD) and diabetes mellitus (DM) are important public health problems that are major causes of morbidity, mortality, and healthcare burden worldwide. According to Regassa et al. (2021), CVD comprises a group of disorders affecting the heart and blood vessels, including coronary heart disease (CHD), cerebrovascular disease, peripheral arterial disease, deep vein thrombosis, and pulmonary embolism. Diabetes mellitus, by contrast, is a chronic metabolic disorder associated with insulin resistance and a gradual decline in pancreatic β -cell function, resulting in poor glucose regulation. It is well established that diabetes and CVD often co-exist and that biological mechanisms and risk factors link the two conditions. Obesity has been recognised as a key factor connecting metabolic dysfunction, systemic inflammation, insulin resistance, and cardiovascular complications (Lopez-Jimenez et al., 2022). In addition, Piché et al. (2020) noted that differences in obesity phenotypes also have strong links to cardiometabolic risk, underlining the close relationship between obesity and cardiometabolic diseases. There are various risk factors for diabetes and CVD, some of which are modifiable, such as BMI, smoking, and physical inactivity, while others, such as age, are not. These factors affect disease progression in a complex and non-linear manner. Despite these relationships, current statistical methods have conventionally analysed diabetes and CVD as separate outcomes. Such separation restricts the scope for identifying common risk profiles, interdependence, and disease dynamics. With the global burden of cardiometabolic disease increasing, more integrative analytical models are needed to reflect real-world disease interactions. In this application, Vector Generalized Additive Models (VGAMs) offer a flexible and powerful approach for modelling several related response variables while supporting non-linear relationships between predictors and responses. Hence, this study is intended to model the incidence of cardiovascular disease and diabetes simultaneously using VGAMs, taking account of key factors that affect the risk of both diseases, including age, BMI, smoking status, and physical activity. This study aims to provide deeper insight into the joint dynamics, interaction effects, and non-linear relationships that underlie the development of these chronic conditions. Lopez-Jimenez et al. (2022) highlighted the growing literature showing that cardiovascular disease and diabetes share many similar risk factors. Obesity is a well-established risk factor for both diseases and is known to contribute to chronic inflammation, dyslipidaemia, and insulin resistance. In support of this, Piché et al. (2020) showed that disease heterogeneity in obesity phenotypes is reflected in different trends in cardiometabolic risk. Age is another important factor, as higher rates of both CVD and diabetes are significantly associated with older age. Cardiovascular complications are reported to be highly prevalent in older people with type 2 diabetes mellitus (T2DM) compared with their counterparts without T2DM (Regassa et al., 2021). Exercise is an important protective factor against cardiovascular disease and diabetes. Hu et al. (2005) adopted Cox proportional hazards regression to determine the effect of moderate and high levels of physical activity on cardiovascular and all-cause mortality in people with diabetes. According to Nagata et al. (2022), moderate-to-vigorous physical activity (MVPA) from early adulthood through middle age is highly protective against metabolic diseases. Furthermore, Martinez-Gomez et al. (2019) showed that exercise was an effective intervention for lowering cardiovascular disease risk factors, even when participants did not lose weight. Qian et al. (2023) reported that the timing of moderate-to-vigorous physical activity was associated with changes in glycaemic control among adults with type 2 diabetes over a four-year follow-up period.

Recent global evidence indicates that cardiovascular disease and diabetes remain major and increasing burdens, while high BMI, smoking, and related metabolic risks continue to contribute substantially to non-communicable disease outcomes (Mensah et al., 2023; GBD 2021 Diabetes Collaborators, 2023; GBD 2021 Risk Factors Collaborators, 2024).

In type 2 diabetes, adiposity-related and atherogenic indices have also been linked to insulin resistance and cardiovascular risk, which supports the inclusion of BMI in the present joint modelling framework (Briggs et al., 2025).

In addition, smoking is a significant modifiable risk factor for cardiovascular disease and diabetes. Smoking, obesity, and physical inactivity significantly decrease a person's healthy and disease-free life expectancy (Stenholm et al., 2016). For instance, Regassa et al. (2021) used multivariate logistic regression to determine factors associated with cardiovascular disease in diabetic patients. Similarly, Adeboye et al. (2020) used parametric and semi-parametric survival models to evaluate survival outcomes among patients with asthma, demonstrating the importance of advanced regression-based survival techniques in medical research. They found that parametric models, especially the log-normal model, predicted patients' survival better than the Cox proportional hazards model when modelling survival experiences. Similarly, Aako et al. (2022) used survival modelling techniques to identify prognostic factors among breast cancer patients in a tertiary hospital in Nigeria. The study applied Kaplan-Meier estimation, Cox regression analysis, Weibull and log-logistic models, and found that tumour stage and treatment methods were important factors affecting patient survival. These studies emphasise the increasing use of regression and survival analysis in the study of disease progression and health outcomes.

Recent preventive-care literature further supports integrated screening and lifestyle-based intervention across obesity, diabetes, and CVD, reinforcing the relevance of analysing these conditions within a common cardiometabolic framework (Josiah et al., 2026).

Additionally, Hossain et al. (2023) introduced a multi-disease comorbidity prediction framework that considers diabetes, hypertension, and cardiovascular risk trajectories, highlighting the need to model interdependencies among diseases. Currently, however, most analyses either consider only one outcome or are unable to fully account for non-linear relationships and interactions between predictor variables. This limitation reduces their capacity to address the complexity of cardiometabolic diseases. Analytical methods that can model multiple correlated outcomes simultaneously are needed to handle the complexity and interrelatedness of cardiovascular disease and diabetes. Generalized Additive Models (GAMs) extend generalized linear models by including smooth, non-linear functions of predictors and modelling multiple response variables simultaneously. VGAMs extend GAMs to multiple response variables that include smooth, non-linear functions of predictors. Generalized additive models have been shown to model flexible non-linear effects in epidemiological and biomedical data sets; however, few studies have applied these models to multivariate disease models. A great deal of research has been conducted on factors affecting cardiovascular disease and diabetes, but gaps remain in the literature. Existing research largely examines these conditions individually rather than jointly and often does not account for the interdependencies and common mechanisms underlying them. In addition, few studies have focused on the non-linear relationships between key predictors and disease outcomes, and interaction effects among risk factors are generally ignored. Moreover, only a small number of studies have used sophisticated multivariate modelling techniques such as VGAMs that can model correlated responses. Thus, the specific research gap addressed in this study is the limited use of a simultaneous modelling strategy that can evaluate correlated CVD and diabetes outcomes while allowing possible non-linear effects of age and BMI and interaction effects among behavioural risk factors. This study is designed to overcome these limitations and uses a simultaneous modelling approach that captures both the complexity and interrelatedness of cardiometabolic diseases.

2. Methodology

2.1 Models Specification and Estimation

The effects of selected risk factors on the prevalence of cardiovascular disease (CVD) and diabetes were investigated using cross-sectional data obtained from patients attending one Federal Medical Centre and two State Government Hospitals with the help of binary logistic regression and Vector Generalized Additive Models (VGAMs). A purposive sampling method was used to ensure that participants were drawn from the eligible study population, and participants with incomplete information on the variables of interest were excluded from the analysis. The explanatory variables included age (years), BMI (kg/m^2), smoking status (smoker/non-smoker), and physical activity (physically active/inactive), while the disease outcomes were coded as binary variables (1 = disease present, 0 = disease absent). The probability of disease occurrence for each disease outcome was estimated using binary logistic regression, and the VGAM framework was used to model the correlated binary disease outcomes and account for possible non-linear relationships between the explanatory variables and response variables. All statistical analyses were conducted with R statistical software using binary logistic regression and Vector Generalized Additive Models.

2.2 Multiple Binary Logistic Regression Model

The multiple binary logistic regression model was employed to estimate the probability of cardiovascular disease (CVD) or diabetes based on selected risk factors. Since the response variable is binary, the logit link function was used to relate the probability of disease occurrence to the explanatory variables. The model is expressed as:

$$\log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 AGE_i + \beta_2 BMI_i + \beta_3 SMOKING_i + \beta_4 PA_i \quad (1)$$

$$\eta_i = \log\left(\frac{P_i}{1-P_i}\right) \quad (2)$$

The corresponding probability function is given by:

$$P_i = \frac{e^{\beta_0 + \beta_1 AGE_i + \beta_2 BMI_i + \beta_3 SMOKING_i + \beta_4 PA_i}}{1 + e^{\beta_0 + \beta_1 AGE_i + \beta_2 BMI_i + \beta_3 SMOKING_i + \beta_4 PA_i}} \quad (3)$$

where, P_i is the probability of disease occurrence for participant i , β_0 is the intercept, and $\beta_1, \beta_2, \beta_3, \beta_4$ are regression coefficients associated with Age, BMI, Smoking and PA, respectively, where η_i represents the linear predictor (log-odds of disease occurrence). The regression coefficients measure the effect of each predictor on the log-odds of disease while holding the other variables constant.

2.3 Vector Generalized Additive Models (VGAM)

The correlated binary outcomes of cardiovascular disease and diabetes were modelled jointly using the Vector Generalized Additive Model (VGAM) to capture potential non-linearity between the continuous predictor variables and the disease variables. The general form of the VGAM is given by

$$g_i(\mu_j) = \eta_j = \beta_{0j} + \sum_{k=1}^p f_k(x_k) \quad (4)$$

Furthermore, this study re-expressed the VGAM as:

$$\eta_j = \beta_{0j} + f_1(AGE_i) + f_2(BMI_i) + \beta_{3j} SMOKING_i + \beta_{4j} PA_i \quad (5)$$

where,

$g_i(\cdot)$ represent the link function, μ_j denotes the expected response for disease outcome j , $f_k(\cdot)$ represents smooth functions of the explanatory variables, η_j is the linear predictor for disease outcome j (CVD or Diabetes), β_{0j} represents the intercept parameter, $f_1(AGE_i)$ and $f_2(BMI_i)$ denote the smooth spline functions for age and BMI, $SMOKING_i$ represents smoking status, PA_i represents physical activity status and $j = 1, 2, \dots$ represent CVD and diabetes.

2.4 Smooth Non-linear Function Models

The smoothing functions in the VGAM framework from equations (4) and (5) were approximated by spline-based smoothing functions.

$$S(x) = \sum_{m=1}^M \alpha_m B_m(x) \quad (6)$$

where α_m represents the corresponding spline coefficients, $B_m(x)$ denotes the spline basis functions, M is the number of spline basis functions.

Consequently, the smooth functions for age and BMI are written as

$$S(AGE_i) = \sum_{m=1}^M \alpha_m B_m(AGE_i) \quad (7)$$

$$S(BMI_i) = \sum_{m=1}^M \alpha_m B_m(BMI_i) \quad (8)$$

These spline functions allow the relationship between the continuous predictors and the disease outcomes to be estimated without imposing a strict linear assumption.

2.5 Model Evaluation and Selection

To assess the predictive performance and goodness-of-fit of the fitted models, several model evaluation criteria were employed, including the receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the likelihood ratio test (LRT).

2.5.1 Receiver Operating Characteristic (ROC) model – AUC

The ROC curve is used to evaluate classification performance at different thresholds. The area under the curve (AUC) is given as:

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx \quad (9)$$

TPR denotes the true positive rate (sensitivity), and FPR denotes the false positive rate (1 - specificity). An AUC value close to one indicates excellent predictive performance, whereas a value close to 0.5 suggests no better classification than random chance.

2.5.2 Akaike Information Criterion (AIC) and Bayesian Information (BIC)

Model selection and comparison were performed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC is computed as

$$AIC = -2 \log(L) + 2k \quad (10)$$

The Bayesian Information Criterion (BIC) is expressed as

$$BIC = -2 \log(L) + k \log(n) \quad (11)$$

L denotes the likelihood function, k represents the number of estimated parameters, and n denotes the sample size. Models with smaller AIC and BIC values are considered to provide a better balance between model fit and complexity for model selection.

2.5.3 Likelihood Ratio Test (LRT)

The likelihood ratio test (LRT) is used to compare nested competing models. The test statistic is given as

$$LRT = -2[\log(L_0) - \log(L_1)] \quad (12)$$

L_0 denotes the likelihood of the null model, and L_1 denotes the likelihood of the alternative model. Under the null hypothesis, the LRT statistic follows a chi-square distribution with degrees of freedom equal to the difference in the number of estimated parameters between the competing models. A statistically significant result indicates that the alternative model provides a significantly better fit to the data than the null model.

3. Results

The findings of the study are presented in this section. To assess the prevalence, distribution, and co-occurrence of CVD and diabetes in the study sample, descriptive analyses were performed first. Subsequently, logistic regression and Vector Generalized Additive Models (VGAMs) were used to examine the relationships between age, BMI, smoking status, physical activity, and disease occurrence. The fitted models were also compared with respect to their predictive performance and adequacy using ROC analysis, AUC, AIC, BIC, and likelihood ratio tests.

Table 1. Prevalence of CVD, Diabetes, and Their Co-occurrence

Disease Condition	Prevalence (%)
Cardiovascular Disease (CVD)	23.81
Diabetes	40.95
Joint Prevalence of CVD and Diabetes	15.24

Source: Rstudio Computation

Table 1 shows that the prevalence of cardiovascular disease (CVD) was 23.81% and that of diabetes was 40.95% among the study population. In addition, 15.24% of participants had both diabetes and CVD, showing a considerable degree of comorbidity among participants. Based on these results, the population faced a high burden of chronic cardiometabolic diseases, and integrated risk assessment and prevention strategies are warranted.

Table 2. Cross-tabulation of CVD and Diabetes Status Among Participants

CVD Status	Diabetes = 0	Diabetes = 1	Total
No CVD (0)	159	81	240
CVD Present (1)	27	48	75
Total	186	129	315

Source: Rstudio Computation

The distribution of cardiovascular disease (CVD) and diabetes among the 315 participants in the study is shown in Table 2. A higher proportion of participants (50.48%) did not have either condition, 25.71% had diabetes alone, and 8.57% had CVD alone. Comorbidity between CVD and diabetes was found in 48 (15.24%) participants. Moreover, 64.0% of those diagnosed with CVD had diabetes, compared with 37.2% of those diagnosed with diabetes who had CVD, indicating that diabetes occurred more frequently in individuals with CVD.

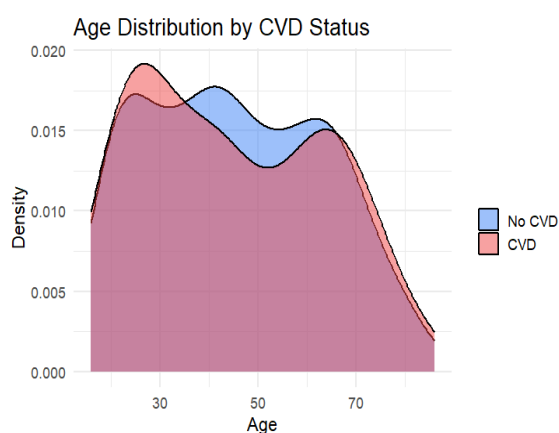


Fig. 1. Age Distribution by CVD Status

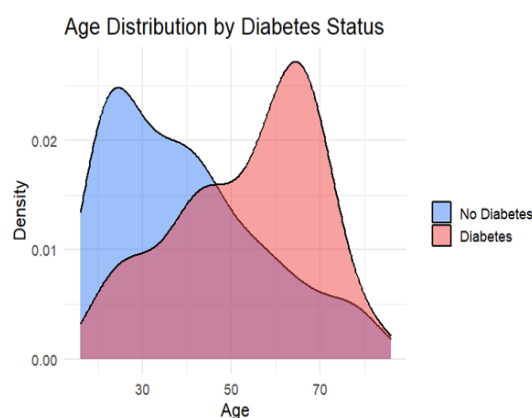


Fig. 2. Age Distribution by Diabetes Status

In Figs. 1 and 2, the distributions are positively associated with cardiovascular disease (CVD) and diabetes, respectively. The age distribution of diabetic patients is mostly skewed, with the majority of patients in the older age groups (around 60-70 years), compared with non-diabetic patients, who are in the younger age groups (30-40 years), indicating a higher prevalence of diabetes with age. Likewise, the age distributions of individuals with CVD are older than those of individuals without CVD, but the overlap between the distributions is larger, suggesting that the separation between individuals with and without CVD is weaker. In general, the results indicate that age is a risk factor for both diabetes and cardiovascular disease, and that the association appears to be stronger for diabetes than for cardiovascular disease.

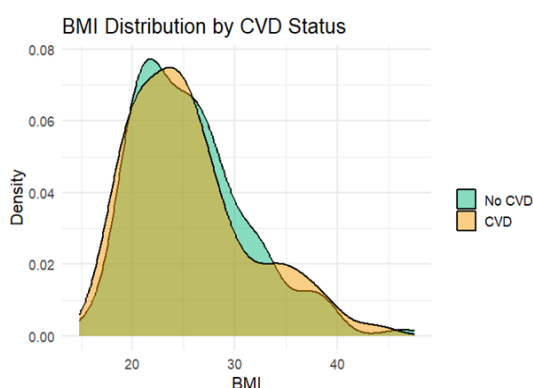


Fig. 3. BMI Distribution by CVD Status

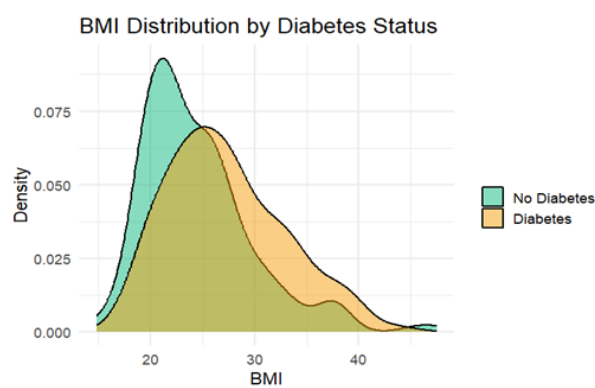


Fig. 4. BMI Distribution by Diabetes Status

For BMI, the distribution plots shown in Figs. 3 and 4 indicate a positive relationship with both cardiovascular disease (CVD) and diabetes. The BMI distribution is shifted to the right in individuals with diabetes, as evidenced by higher densities in the overweight and obese BMI categories, reflecting a strong association between high BMI and diabetes prevalence. Likewise, there is a small difference in BMI values between those with and without CVD, but there is considerable overlap in BMI distribution between the groups. The greater separation for diabetes implies that, in the study population, BMI might be a better predictor of diabetes than of CVD. In conclusion, excess body weight is an important cardiometabolic risk factor linked to both diabetes and cardiovascular disease.

Table 3. Prevalence of Cardiovascular Disease (CVD) and Diabetes by Smoking Status and Physical Activity Level

Smoking Status	Physical Activity	Sample Size	CVD Prevalence (%)	Diabetes Prevalence (%)
Non-smoker	Inactive	124	25	38.7
Non-smoker	Active	141	22.7	42.6
Smoker	Inactive	19	31.6	36.8
Smoker	Active	31	19.4	45.2

Source: Rstudio Computation

The prevalence of cardiovascular disease (CVD) and diabetes according to smoking and physical activity groups is presented in Table 3. The highest prevalence of CVD was observed among inactive smokers (31.6%), and the lowest prevalence was observed among active smokers (19.4%). Among non-smokers, there was little difference in CVD prevalence between inactive and active participants, with inactive participants having a slightly higher prevalence of CVD (25.0%) than active participants (22.7%), suggesting that physical inactivity may be linked to a higher burden of CVD. Diabetes prevalence differed less across groups, ranging from 36.8% among inactive smokers to 45.2% among active smokers, with the highest prevalence among active smokers. The prevalence of both CVD and diabetes varied across smoking and physical activity categories, but the differences observed here were not statistically significant. A more detailed model and statistical analysis were therefore needed to determine the independent and/or combined effects of smoking status and physical activity on the prevalence of CVD and diabetes.

The correlation analysis in Fig. 5 and Table 4 shows important relationships between the variables studied. Age showed a moderate positive correlation with BMI ($r = 0.378$) and diabetes ($r = 0.377$), signifying a moderate correlation between age, BMI, and diabetes among older people. BMI also had a slight positive correlation with diabetes ($r = 0.242$), indicating that greater BMI may be associated with impaired metabolism and a higher risk of diabetes. In addition, there was a positive correlation between the two chronic conditions, namely CVD and diabetes ($r = 0.262$), suggesting a weak-to-moderate comorbidity association between the two. On the other hand, the correlations between CVD and age ($r = -0.008$) and BMI ($r = -0.010$) were non-significant, suggesting that the linear relationships between these variables and CVD may be limited in the study sample. Therefore, the current dataset indicates that diabetes has a stronger association with age and BMI than CVD does. The positive

correlation between the two diseases, CVD and diabetes, observed here further justifies the use of a joint modelling framework, such as the Vector Generalized Additive Model (VGAM), which considers dependence between correlated disease outcomes.

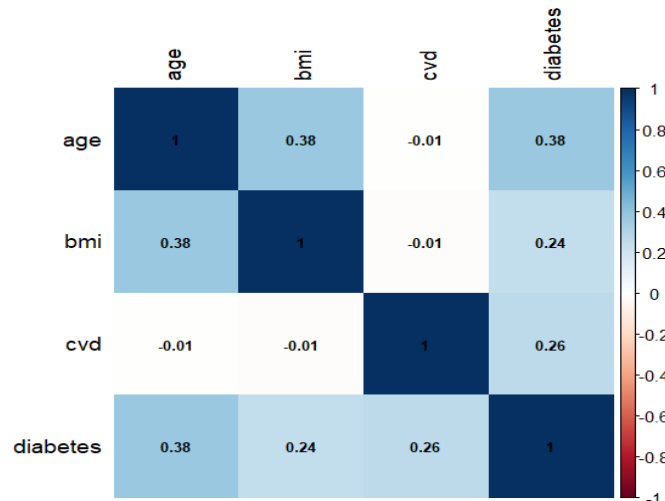


Fig. 5. Correlation Heat-map Plot of Age, BMI, CVD, and Diabetes

Table 4. Correlation Matrix for Age, BMI, CVD, and Diabetes

Variables	Age	BMI	CVD	Diabetes
Age	1	0.378	-0.008	0.377
BMI	0.378	1	-0.01	0.242
CVD	-0.008	-0.01	1	0.262
Diabetes	0.377	0.242	0.262	1

Source: Rstudio Computation

Table 5. Univariate Logistic Regression Model for Cardiovascular Disease (CVD)

Predictor	Estimate (β)	Standard Error	z-value	p-value
Intercept	-0.8977	0.6578	-1.365	0.172
Age	-0.0013	0.008	-0.157	0.875
BMI	-0.0038	0.0257	-0.149	0.881
Smoking (Smoker)	0.0292	0.3628	0.081	0.936
Physical Activity (Active)	-0.2164	0.2671	-0.81	0.418

Source: Rstudio Computation

The univariate logistic regression model shown in Table 5 was used to evaluate the effects of age, BMI, smoking, and physical activity on the prevalence of cardiovascular disease (CVD). The results showed that none of the explanatory variables was statistically significant at the 5% level. Age ($p = 0.875$) and BMI ($p = 0.881$) showed very slight negative associations with the likelihood of CVD, suggesting that differences in age and body mass index did not significantly explain the occurrence of cardiovascular disease in the study population. Smoking status was not significantly associated with the development of CVD and had a very small positive coefficient ($\beta = 0.0292$, $p = 0.936$), indicating that smokers were slightly more likely than non-smokers to develop CVD. Physical activity had a negative association with CVD ($\beta = -0.2164$, $p = 0.418$), suggesting that physically active participants had a lower likelihood of developing cardiovascular disease than inactive participants, although this association was not statistically significant. In addition, the results indicate that the influence of the selected predictors on the occurrence of CVD in the study sample was not sufficiently strong in the linear model; therefore, the use of more flexible models, such as VGAMs that can account for possible non-linear relationships and interactions between predictors, is recommended.

Table 6. Univariate Logistic Regression Model for Diabetes Incidence

Predictor	Estimate (β)	Standard Error	z-value	p-value
Intercept	-3.8732	0.6709	-5.773	<0.001 ***
Age	0.0427	0.0077	5.532	<0.001 ***
BMI	0.0493	0.0234	2.102	0.0355 *
Smoking (Smoker)	0.0582	0.3399	0.171	0.864
Physical Activity (Active)	0.4227	0.2557	1.653	0.0983 .

Source: Rstudio Computation

The univariate logistic regression model in Table 6 was used to investigate the impact of age, BMI, smoking, and physical activity on the occurrence of diabetes. The findings indicated that age and BMI were statistically significant at the 5% level of significance for diabetes. In particular, diabetes incidence was positively associated with age ($p < 0.001$, $\beta = 0.0427$), suggesting that diabetes incidence was higher as age increased. Similarly, BMI had a positive and statistically significant relationship with diabetes ($\beta = 0.0493$, $p = 0.0355$), meaning that people with higher BMI values were at greater risk of developing diabetes. Smoking status had a small positive coefficient ($\beta = 0.0582$, $p = 0.864$), but this effect was not statistically significant; therefore, there was insufficient evidence of an independent effect of smoking on the occurrence of diabetes in the study population. Physical activity showed a positive association with diabetes ($\beta = 0.4227$, $p = 0.0983$), which was marginally significant at the 10% level and may indicate a relationship between physical activity status and diabetes incidence. Therefore, the results show that age and BMI are significant factors for diabetes incidence, while smoking and physical activity showed less significant associations in the fitted linear logistic model. This supports the justification for using VGAMs as more flexible approaches to test potential non-linear effects and interactions among the predictors.

Table 7. VGAM Bivariate Model Results for Simultaneous Modeling of CVD and Diabetes

Predictor	CVD Equation ((μ_1)) Coefficient	Diabetes Equation ((μ_2)) Coefficient
Intercept	-0.6689	-2.2602
Smooth Age Effect (s(age))	0.0001	0.026
Smooth BMI Effect (s(bmi))	0.0006	0.0259
Smoking (Smoker)	0.2088	0.0141
Physical Activity (Active)	-0.0974	0.2525
Smoking $\times$ Physical Activity	-0.3582	0.0764
Residual Correlation Parameter ((ρ))		{1.2426 (rhobit scale)}

Source: Rstudio Computation

Table 8. Approximate Chi-Square Tests for Nonlinear Effects in the VGAM Model

Smooth Term	Disease Outcome	Chi-Square Statistic	p-value
(s(age))	CVD	2.133	0.5398
(s(age))	Diabetes	10.701	0.0134
(s(bmi))	CVD	1.8262	0.6052
(s(bmi))	Diabetes	3.7396	0.2789

Source: Rstudio Computation

Tables 7 and 8 present the VGAM fitted to jointly model the incidence of cardiovascular disease (CVD) and diabetes with possible dependence between the two outcomes. Smooth non-linear effects of age and BMI, as well as smoking status, physical activity, and smoking status $\times$ physical activity interaction effects, were incorporated into the model. The results showed that the non-linear effect of age on diabetes was statistically significant ($\chi^2 = 10.701$, $p = 0.0134$), which means that the risk of diabetes changes non-linearly across age levels. The non-linear effect of age on CVD, however, was not significant ($p = 0.5398$). The same applied to the smooth BMI effects for both chronic conditions, which were not statistically significant, although BMI had relatively stronger positive effects on diabetes than on CVD. There was a stronger positive relationship between

smoking and CVD, and a slightly protective association between physical activity and CVD, compared with diabetes. Smoking and physical activity were found to have a negative association with CVD, indicating that physical activity may partially buffer the harmful effect of smoking on cardiovascular health. Moreover, the positive residual correlation parameter supported the use of the joint modelling framework, as the explanatory variables were not the only influences in the model. The positive residual correlation also suggested that a positive association between CVD and diabetes remained after controlling for the explanatory variables. Thus, the role of age in diabetes incidence is complex and non-linear, and the concurrent incidence of both diseases indicates shared underlying risk mechanisms and the potential value of integrated prevention and healthcare intervention strategies. Using the estimated coefficients of the VGAM bivariate model, the simultaneous models for CVD and diabetes can be expressed as follows (recall equations (4) and (5)):

Estimated VGAM Bivariate Model for CVD and Diabetes:

$$\eta_{CVD} = -0.6689 + 0.0001s(Age) + 0.0006s(BMI) + 0.2088(Smoking) - 0.0974(Physical\ Activity) - 0.3582(Smoking \times Physical\ Activity) \quad (13)$$

$$\eta_{Diabetes} = -2.2602 + 0.0260s(Age) + 0.0259s(BMI) + 0.0141(Smoking) + 0.2525(Physical\ Activity) + 0.0764(Smoking \times Physical\ Activity) \quad (14)$$

where, η_{CVD} and $\eta_{Diabetes}$ are the linear predictors for CVD and diabetes incidence.

Table 9. Estimated Residual Correlation Parameter from the VGAM Bivariate Mode

Parameter	Estimate
Linear Predictor for Correlation ((η_{ρ}))	1.2426
Residual Correlation ((ρ))	0.552

Source: Rstudio Computation

After adjusting for age, BMI, smoking status, physical activity, and their interactions, the estimated residual correlation parameter in Table 9 was $\rho = 0.5520$, suggesting a moderately strong positive correlation between cardiovascular disease (CVD) and diabetes. This suggests that the two diseases are not statistically independent and that significant unexplained variation remains in both diseases after adjustment for the observed predictors. The residual dependence indicates that unobserved risk factors or common biological pathways may underlie the co-occurrence of both diseases, such as metabolic dysfunction, inflammation, genetics, or lifestyle-related factors that were not explicitly represented in the model. This further indicates that people with one condition are more likely to have the other condition, with the likelihood increasing as the strength of the correlation increases. The outcome confirms that using a multivariate joint modelling approach, such as the VGAM bivariate model, is preferable to separate independent logistic regression models.

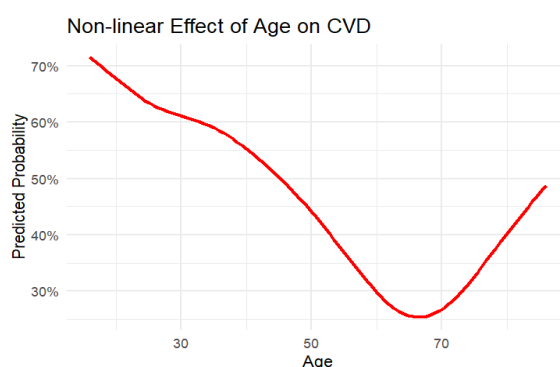


Fig. 6. Non-linear Effects of Age on the Predicted Predicted Probability of CVD

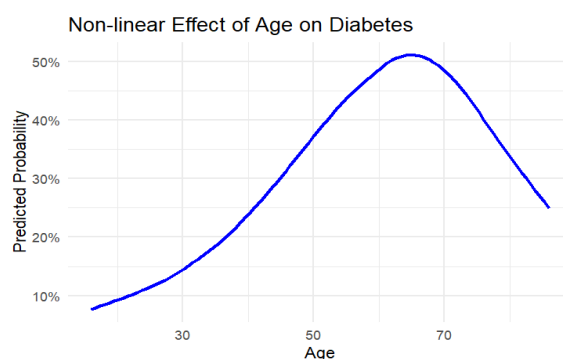


Fig. 7. Non-linear Effects of Age on the Probability of Diabetes

The estimated smooth effects of age on the probability of cardiovascular disease (CVD) and diabetes, respectively, are shown in Figs. 6 and 7. The relationship between age and CVD is not linear, as shown in Fig. 6: the predicted risk of CVD decreases from younger ages, reaches a minimum around 65-68 years, and then gradually increases at older ages. The curvature of age in relation to CVD prevalence, however, was not statistically significant ($p = 0.5398$), and thus the observed age effect was not sufficient to conclude that there was a meaningful non-linear age effect in the study population. In contrast, the statistically significant non-linear relationship between age and diabetes in Fig. 7 shows that the predicted probability of diabetes rises steadily with age, peaks at about 65-70 years, and then decreases slightly at older ages. This finding further suggests that age is an important parameter to consider in assessing diabetes prevalence, implying that the risk of diabetes generally tends to rise with age, possibly reflecting age-related changes in metabolism and body physiology that affect glucose regulation.

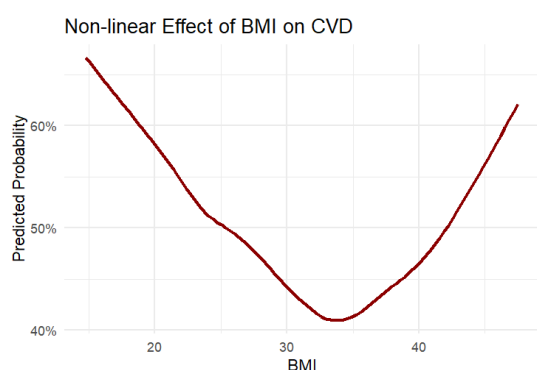


Fig. 8. U-shaped non-linear relationship between cardiovascular disease (CVD) risk

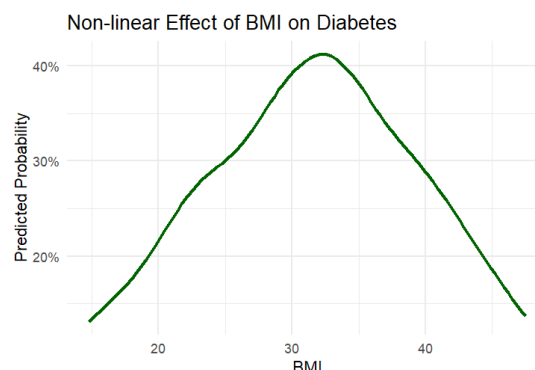


Fig. 9. Inverted U-shaped non-linear BMI and relationship between BMI and diabetes risk

As seen in Fig. 8, the association between BMI and the risk of CVD is U-shaped. BMI levels of around 66% have a predicted probability of just below 41% for CVD, while higher BMI levels have a predicted probability of over 60%. The pattern indicates that lower and higher BMI values might correspond to increased predicted probabilities of CVD, while moderate BMI values might be associated with decreased predicted probabilities of CVD. To assess the possible non-linear association between BMI and cardiovascular outcomes, non-linear terms should be explored to reflect the shape of the curve. However, the results shown in Table 8 indicate that the non-linear effect of BMI was not significant. Thus, the observed U-shaped curve should not be taken as definitive proof of a meaningful non-linear association between BMI and the risk of CVD. Fig. 9 illustrates a non-linear relationship between diabetes risk and BMI, which appears to be inverted U-shaped. The risk of developing diabetes continues to increase with each successive BMI value until about BMI 32-33, at which point the risk approaches 41%, and then tapers off as BMI increases further. This pattern suggests that those with moderate to high BMI levels have the greatest risk of being diagnosed with diabetes, while those with low BMI and very high BMI have relatively lower risks. The bend in the observed relationship indicates that the relationship between BMI and diabetes is not strictly linear but instead has a statistical non-linearity.

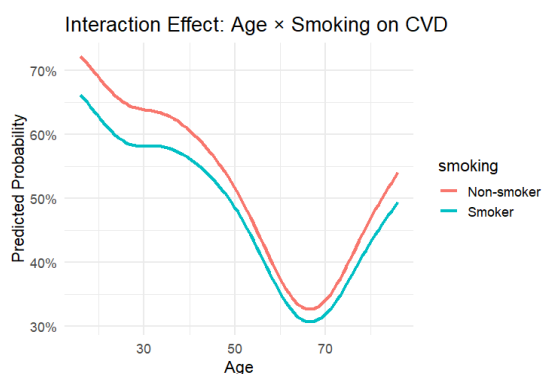


Fig. 10. Interaction Effect of Age and Smoking Status on Predicted Probability of CVD

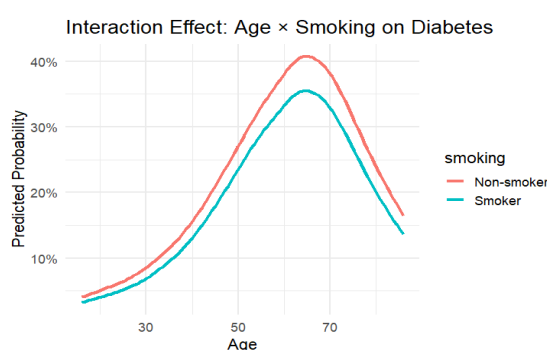


Fig. 11. Interaction Effect of Age and Smoking Status on Predicted Probability of Diabetes

The interaction effects between age and smoking status on the predicted probabilities of CVD and diabetes are shown in Figs. 10 and 11, respectively, using the VGAM. The relationships between age and disease incidence are clearly non-linear, suggesting that the impact of age on both CVD and diabetes is not linear throughout the lifespan. The predicted probability of CVD is seen to decline gradually with age from younger ages to around 65-68 years and then rise again at older ages, creating a U-shaped curve for both smokers and non-smokers in Fig. 10. The fitted curves for smokers and non-smokers are similar in shape, with smokers having slightly lower predicted probabilities than non-smokers; however, the similarity of the curves indicates that age is a more important determinant of cardiovascular risk than smoking status alone in this dataset. The predicted probability of diabetes rises gradually with age, reaches a maximum at about 65-70 years, and then decreases slightly with age for both smoking groups, suggesting that the risk of diabetes is higher with increasing age because of cumulative metabolic and physiological changes. In both figures, the curves for smokers and non-smokers are very close, indicating that the interaction effects between age and smoking status are not large, although smoking still plays a role in variation in disease risk. These patterns are relevant to the research objectives of the study, which are to explore non-linear and interaction effects among the predictors within the VGAM framework.

Fig. 12 shows the joint predicted probability of having both CVD and diabetes for physically inactive smokers, based on the fitted VGAM. It shows clear non-linear trends in the risk of joint disease across combinations of age and BMI. Lighter colours correspond to higher joint probabilities and are mostly seen at younger and older age ranges and at very low or very high BMI values, suggesting higher combined disease susceptibility in these ranges. By contrast, darker areas clustered around middle-to-late adulthood and moderate BMI levels suggest relatively low joint risk. The smooth transitions between heat-map cells indicate that the interaction between age and BMI in the co-occurrence of CVD and diabetes is complex and non-linear rather than constant across the predictor space. From a statistical perspective, these results emphasise the interplay between demographic and lifestyle-related risk factors in determining chronic disease comorbidity and reinforce the positive residual dependence observed in the VGAM bivariate model.

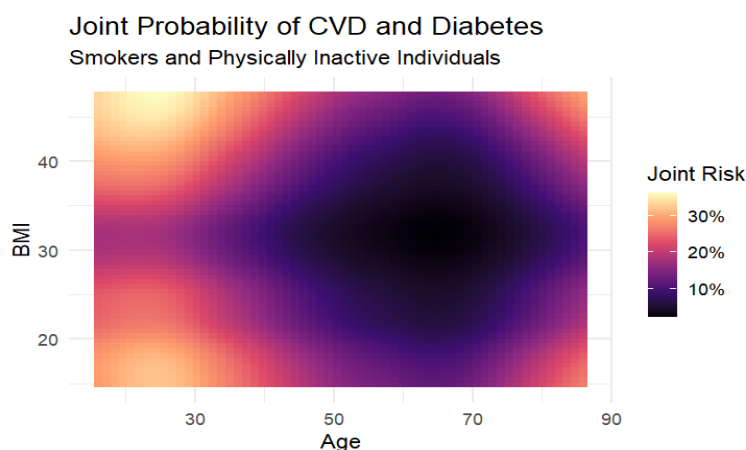


Fig. 12. Joint Predicted Probability of CVD and Diabetes among Smokers and Physically Inactive Individuals

Table 10. Model Performance Evaluation

Outcome	AUC
CVD	0.5099
Diabetes	0.7598

Source: Rstudio Computation

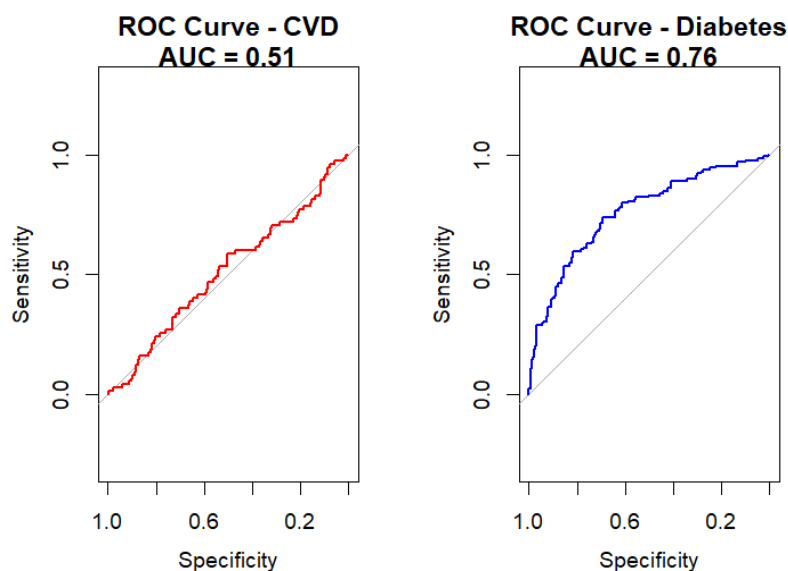


Fig. 13. ROC Curves for Cardiovascular Disease (CVD) and Diabetes Prediction Model

Table 10 and Fig. 13 show the model performance evaluation, ROC curves, and AUC values for the predictive models of CVD and diabetes. The AUC for the CVD model was 0.51, indicating very poor discriminative ability and performance only slightly better than random classification. The ROC curve is close to the diagonal, indicating limited ability to correctly distinguish between those with and without CVD. The diabetes model, however, had an AUC of 0.76, indicating acceptable-to-good predictive accuracy. The ROC curve for diabetes is well above the diagonal line, indicating good sensitivity-specificity trade-offs and good overall classification accuracy. Therefore, these results show that the discriminatory power and reliability of the diabetes prediction model are higher than those of the CVD model.

Table 11. Model Comparison

Model	AIC	BIC
VGAM Joint Model	710.97	752.25
Restricted Independence Model	718.1	811.4

Source: Rstudio Computation

The model comparison in Table 11, based on AIC and BIC, indicates that the VGAM joint model outperformed the restricted independence model. The VGAM joint model had a smaller AIC (710.97) and BIC (752.25) than the restricted independence model, indicating that the model fitted better and was more statistically efficient. Overall, these findings suggest that joint modelling is a more appropriate approach for capturing the dependence and co-occurrence between cardiovascular disease and diabetes in the study population.

Table 12. Likelihood Ratio Test

Statistic	Value
LRT Statistic	20.6037
Degrees of Freedom	1
P-value	0.000006

Source: Rstudio Computation

The likelihood ratio test presented in Table 12 supported the VGAM joint model and revealed a statistically significant dependence between cardiovascular disease and diabetes (LRT = 20.60, $p < 0.001$), indicating the presence of shared latent risk mechanisms. Increasing age, elevated BMI, smoking, and physical inactivity were associated with increased disease risk.

4. Summary of Findings

In this study, a VGAM framework was adopted to examine the co-occurrence of CVD and diabetes, and it found a significant burden of cardiometabolic disease among the study population. The prevalence analysis revealed that 23.81% of the participants had CVD, 40.95% had diabetes, and 15.24% had both conditions, suggesting a high level of comorbidity. Cross-tabulation also showed high co-occurrence of the diseases, with about 64% of those with CVD also having diabetes. The correlation analysis showed a positive correlation between age, BMI, diabetes, and CVD, with the highest correlation between age and diabetes. The univariate logistic regression model showed that age and BMI were significant predictors of diabetes incidence, while none of the predictors had statistically significant linear effects on the occurrence of CVD, indicating that the relationship between the predictors and disease outcomes might be more complex than conventional linear models can explain. The VGAM bivariate model therefore provided more insight, as it showed a significant non-linear effect of age on diabetes incidence, but not on CVD. Smoking was more strongly associated with CVD than with diabetes, and physical activity was protective against cardiovascular disease. Importantly, the estimated residual correlation coefficient ($\rho = 0.552$) showed a moderate positive correlation between CVD and diabetes after adjusting for age, BMI, smoking status, and physical activity, confirming the presence of shared underlying risk mechanisms. The VGAM joint model also performed better than the restricted independence model in terms of lower AIC and BIC values, and the likelihood ratio test demonstrated statistically significant dependence between the two diseases. The diabetes model had good discriminatory ability (AUC = 0.76), while the CVD model had poor predictive accuracy (AUC = 0.51). Overall, these results show that the effects of age, BMI, smoking, and physical activity on the risk of CVD and diabetes differ and highlight the need for joint modelling approaches in cardiometabolic disease analysis. The findings of this study are broadly in line with recent literature on cardiovascular disease, diabetes, and cardiometabolic multimorbidity. The positive residual dependence found in this study is consistent with Han et al. (2024), who reported that lifestyle factors and metabolic health are associated with cardiometabolic multimorbidity, and Gupta et al. (2025), who described obesity, metabolic syndrome, diabetes, and cardiovascular diseases as interconnected conditions. The strong non-linear association of age with diabetes is also consistent with Jiang et al. (2024) and Zhang et al. (2025), who linked biological ageing to cardiometabolic multimorbidity and incident type 2 diabetes. Likewise, the positive correlation between BMI and diabetes is consistent with Han et al. (2024) and Gupta et al. (2025), who identified excess adiposity and obesity-related mechanisms as relevant to cardiometabolic disease. In addition, the positive residual dependence between CVD and diabetes is consistent with Cheng et al. (2024) and Wang et al. (2025), who reported that chronic low-grade inflammation, insulin resistance, and dyslipidaemia are associated with

cardiometabolic multimorbidity. The superior performance of the VGAM joint model also lends support to recent methodological research that has promoted joint modelling approaches for correlated chronic diseases, as these approaches can better account for shared disease pathways and outcome dependence than separate regression models.

5. Conclusion

In this study, a Vector Generalized Additive Model (VGAM) was used to explore the interaction and dependence between cardiovascular disease (CVD) and diabetes. The results indicated that diabetes was more common than CVD, and a significant proportion of participants had both conditions. The selected predictors did not have statistically significant linear associations with CVD, while logistic regression analysis revealed that age and body mass index (BMI) were significant predictors of diabetes. The VGAM also revealed a significant non-linear effect of age on diabetes, showing the importance of flexible modelling methods for identifying complex disease patterns. The estimated residual correlation parameter, adjusted for age, BMI, smoking, and physical activity, indicated moderate positive dependence between CVD and diabetes. The diabetes model had good discriminatory ability (AUC = 0.76), while the CVD model had poor predictive ability (AUC = 0.51), suggesting the need for additional risk factors to better predict CVD. These results indicate that chronic conditions such as CVD and diabetes may be related and may share common risk patterns. The study also shows that the VGAM framework is a useful analytical tool for modelling multiple correlated disease outcomes, with the possibility of uncovering non-linear relationships that may not be well represented by traditional regression models.

6. Limitation

This study was limited by its cross-sectional design, which restricts causal interpretation and temporal inference. The sample size was 315, and some smoking and physical activity subgroups were small, which may have affected the stability of estimates. The hospital-based recruitment strategy may limit generalisability. Important clinical and behavioural variables, including blood pressure, lipid profile, dietary patterns, medication use, family history, disease duration, and socioeconomic status, were not included. In addition, the CVD prediction model had poor discriminatory performance, indicating that the selected predictors were insufficient for accurate CVD classification.

Consent

It is not applicable.

Ethical Approval

It is not applicable.

Declaration of AI Use

This manuscript was prepared through the combined contributions of all author(s), including contributions to the study design, data, content development, results, interpretation, and related scholarly work. The author(s) acknowledge the use of Grammarly and ChatGPT to assist with grammar checking, language refinement, reference formatting. These AI-assisted tools were not used as authors and did not replace the intellectual contributions or scholarly judgment of the author(s). All AI-assisted outputs, including content, references, and interpretations, were carefully reviewed, revised, verified, and approved by the author(s). The author(s) accept full responsibility for the accuracy, integrity, and final content of the manuscript.

Competing Interests

Authors have declared that no competing interests exist.

References

- Aako, O. L., Adewara, J. A., & Are, S. O. (2022). Risk factor analysis of breast cancer patients in a Nigerian tertiary hospital. *FUDMA Journal of Sciences*, 6(3), 95–99. <https://doi.org/10.33003/fjs-2022-0603-975>
- Adeboye, N. O., Ajibode, I. A., & Aako, O. L. (2020). On the survival assessment of asthmatic patients using parametric and semi-parametric survival models. *Occupational Diseases and Environmental Medicine*, 8(2), 50–63. <https://doi.org/10.4236/odem.2020.82004>
- Briggs, O. N., Sheudeen, A. A., & Elechi-Amadi, K. N. (2025). Evaluation of lipid accumulation product, visceral adiposity and atherogenic risk indices in type 2 diabetics in Port Harcourt, Nigeria. *Asian Journal of Research in Biochemistry*, 15(6), 92–104. <https://doi.org/10.9734/ajrb/2025/v15i6450>
- Cheng, W., Du, Z., & Lu, B. (2024). Chronic low-grade inflammation associated with higher risk and earlier onset of cardiometabolic multimorbidity in middle-aged and older adults: A population-based cohort study. *Scientific Reports*, 14, Article 22635. <https://doi.org/10.1038/s41598-024-72988-7>
- GBD 2021 Diabetes Collaborators. (2023). Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 402(10397), 203–234. [https://doi.org/10.1016/S0140-6736\(23\)01301-6](https://doi.org/10.1016/S0140-6736(23)01301-6)
- GBD 2021 Risk Factors Collaborators. (2024). Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 403(10440), 2162–2203. [https://doi.org/10.1016/S0140-6736\(24\)00933-4](https://doi.org/10.1016/S0140-6736(24)00933-4)
- Gupta, A., Shah, K., & Gupta, V. (2025). Interconnected epidemics: Obesity, metabolic syndrome, diabetes and cardiovascular diseases—insights from research and prevention strategies. *Discover Public Health*, 22, Article 106. <https://doi.org/10.1186/s12982-025-00496-8>
- Han, W., Chen, S., Kong, L., Li, Q., Zhang, J., Shan, G., & He, H. (2024). Lifestyle and clinical factors as predictive indicators of cardiometabolic multimorbidity in Chinese adults: Baseline findings of the Beijing Health Management Cohort (BHMC) study. *Computers in Biology and Medicine*, 168, Article 107792. <https://doi.org/10.1016/j.compbiomed.2023.107792>
- Hossain, M. Z., Khan, M. M., Islam, R., Nahar, K., & Kabir, M. F. (2023). Formulation of a multi-disease comorbidity prediction framework: A data-driven case analysis on of diabetes, hypertension, and cardiovascular risk trajectories. *Journal of Computer Science and Technology Studies*, 5(3), 161–182. <https://doi.org/10.32996/jcsts.2023.5.3.12>
- Hu, G., Jousilahti, P., Barengo, N. C., Qiao, Q., Lakka, T. A., & Tuomilehto, J. (2005). Physical activity, cardiovascular risk factors, and mortality among Finnish adults with diabetes. *Diabetes Care*, 28(4), 799–805. <https://doi.org/10.2337/diacare.28.4.799>
- Jiang, M., Tian, S., Liu, S., Wang, Y., Guo, X., Huang, T., Lin, X., Belsky, D. W., Baccarelli, A. A., & Gao, X. (2024). Accelerated biological aging elevates the risk of cardiometabolic multimorbidity and mortality. *Nature Cardiovascular Research*, 3, 332–342. <https://doi.org/10.1038/s44161-024-00438-8>
- Josiah, P. A., Tony-Okpalaobi, O. H., Raji, G. T., Nwoye, E. O., Ihebie, C. E., & Asorho, B. M. (2026). Preventive care and early intervention strategies for obesity, diabetes and cardiovascular disease. *Journal of Disease and Global Health*, 19(1), 294–315. <https://doi.org/10.56557/jodagh/2026/v19i110481>
- Lopez-Jimenez, F., Almahmeed, W., Bays, H., Cuevas, A., Di Angelantonio, E., le Roux, C. W., Sattar, N., Sun, M. C., Wittert, G., Pinto, F. J., & Wilding, J. P. H. (2022). Obesity and cardiovascular disease: Mechanistic insights and management strategies. A joint position paper by the World Heart Federation and World Obesity Federation. *European Journal of Preventive Cardiology*, 29(17), 2218–2237. <https://doi.org/10.1093/eurjpc/zwac187>
- Martinez-Gomez, D., Lavie, C. J., Hamer, M., Cabanas-Sanchez, V., Garcia-Esquinas, E., Pareja-Galeano, H., Struijk, E. A., Sadarangani, K. P., Ortega, F. B., & Rodríguez-Artalejo, F. (2019). Physical activity without weight loss reduces the development of cardiovascular disease risk factors—A prospective cohort study of more than one hundred thousand adults. *Progress in Cardiovascular Diseases*, 62(6), 522–530. <https://doi.org/10.1016/j.pcad.2019.11.010>
- Mensah, G. A., Fuster, V., Murray, C. J. L., Roth, G. A., & Global Burden of Cardiovascular Diseases and Risks Collaborators. (2023). Global burden of cardiovascular diseases and risks, 1990–2022. *Journal of the American College of Cardiology*, 82(25), 2350–2473. <https://doi.org/10.1016/j.jacc.2023.11.007>
- Nagata, J. M., Vittinghoff, E., Pettee Gabriel, K., Garber, A. K., Moran, A. E., Rana, J. S., Reis, J. P., Sidney, S., & Bibbins-Domingo, K. (2022). Moderate-to-vigorous intensity physical activity from young adulthood to middle age and metabolic disease: A 30-year population-based cohort study. *British Journal of Sports Medicine*, 56(15), 847–853. <https://doi.org/10.1136/bjsports-2021-104231>

- Piché, M.-E., Tchernof, A., & Després, J.-P. (2020). Obesity phenotypes, diabetes, and cardiovascular diseases. *Circulation Research*, 126(11), 1477–1500. <https://doi.org/10.1161/CIRCRESAHA.120.316101>
- Qian, J., Xiao, Q., Walkup, M. P., Coday, M., Erickson, M. L., Unick, J., Jakicic, J. M., Hu, K., Scheer, F. A. J. L., Middelbeek, R. J. W., & Look AHEAD Research Group. (2023). Association of timing of moderate-to-vigorous physical activity with changes in glycemic control over 4 years in adults with type 2 diabetes from the Look AHEAD trial. *Diabetes Care*, 46(7), 1417–1424. <https://doi.org/10.2337/dc22-2413>
- Regassa, L. D., Tola, A., & Ayele, Y. (2021). Prevalence of cardiovascular disease and associated factors among type 2 diabetes patients in selected hospitals of Harari Region, Eastern Ethiopia. *Frontiers in Public Health*, 8, Article 532719. <https://doi.org/10.3389/fpubh.2020.532719>
- Stenholm, S., Head, J., Kivimäki, M., Kawachi, I., Aalto, V., Zins, M., Goldberg, M., Zaninotto, P., Hanson, L. M., Westerlund, H., & Vahtera, J. (2016). Smoking, physical inactivity and obesity as predictors of healthy and disease-free life expectancy between ages 50 and 75: A multicohort study. *International Journal of Epidemiology*, 45(4), 1260–1270. <https://doi.org/10.1093/ije/dyw126>
- Wang, H., Zhang, H., Zhang, L., Zheng, X., & Yang, M. (2025). Combined effects of insulin resistance and dyslipidaemia on cardiometabolic multimorbidity in Chinese middle-age and older adults: A longitudinal cohort study. *Scientific Reports*, 15, Article 38060. <https://doi.org/10.1038/s41598-025-21850-5>
- Zhang, Z., Wang, J., Yu, B., Sun, Y., Chen, Y., Lu, Y., Wang, N., & Xia, F. (2025). Accelerated biological aging, mediating amino acids, and risk of incident type 2 diabetes: A prospective cohort study. *Journal of Endocrinological Investigation*, 48, 435–443. <https://doi.org/10.1007/s40618-024-02436-w>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2026): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<https://pr.sdiarticle5.com/review-history/161608>