Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 4(4) July 2026
DOI: https://dx.doi.org/10.4314/jobasr.v4i4.23

Genetic-Lifestyle Risk Prediction Model for Hypertension and Diabetes Using Logistic Regression

Agani, S.N.[1*], Saleh I.M.[2] & Akeyede I.[3]
[1,2,3]Department of Statistics, Faculty of Physical Sciences, Federal University of Lafia, Nigeria
[*]Corresponding Author Email: nehesimeon@gmail.com

ABSTRACT
This study focused on genetic-lifestyle risk prediction modelling of hypertension and diabetes using the logistic regression method. The sample sizes, n =1000 patients selected from a projected population, N= 567,694 in Lafia, Nasarawa State. A cross-sectional analytical design was employed, utilizing secondary data obtained from the District Health Information System (DHIS2) database. Classical logistic regression and LASSO logistic regression models were fitted to identify significant predictors of hypertension and diabetes. The results show (38.2%) had hypertension, and (29.5%) had diabetes. Many factors are indicated as diabetes risk: older age (OR = 1.053, p = .002), higher BMI (OR = 1.099, p = .001), family history (OR = 2.410, p = .001), and smoking (OR = 1.419, p = .041). Staying physically active, show protective indicator (OR = 0.657, p = .018). Hypertension indicate that age (OR = 1.046, p = .001), BMI (OR = 1.085, p = .001), family history (OR = 2.055, p < .001), and smoking (OR = 1.507, p = .021) increase risk, while physical activity gives protection (OR = 0.694, p = .030). The models show good testing, with AUC values of 0.78 for diabetes and 0.82 for hypertension which is suitable for detecting individual at higher risk. The findings show both diseases share a common genetic and lifestyle factor, which means a single predictive model could realistically screen for both at once which helps health workers in the study area a statistical tool for detecting individuals at risk early and prevention intervention.

Keywords:
Diabetes, Hypertension, Logistic Regression, Genetic-Lifestyle Factors, Risk Prediction Model, Gene-Environment Interaction

INTRODUCTION

Hypertension and diabetes mellitus are among the most persistent non-communicable diseases, contributing significantly to cardiovascular morbidity, mortality, and healthcare expenditure worldwide. Hypertension affects an estimated 1.28 billion adults globally, while diabetes affects over 500 million individuals, with projections indicating continued increases over the coming decades (WHO,2023; Lazzarini *et al*.,2023). The burden of hypertension and diabetes is particularly marked in low- and middle-income countries, including Nigeria, where rapid population growth and lifestyle transitions have contributed to a rise in risk factors such as physical inactivity, overweight, smoking, and unhealthy dietary patterns (Lazzarini *et al*.,2023). The coexistence of both conditions within the same individual is common, owing to shared risk factors and pathophysiological mechanisms, thereby elevating the risk of cardiovascular disease, renal failure, stroke, and mortality.

Given the growing epidemiological burden in Nigeria, particularly in Lafia town, recent research indicates a significant rise in the prevalence of type 2 diabetes mellitus (T2DM), with family history, physical inactivity, and high body mass index (BMI) identified as key contributing factors (Adeyemi *et al*.,2024). Also, hypertension remains highly prevalent among adults and is frequently underdiagnosed and poorly controlled, especially among individuals with coexisting comorbidity (Adegoke *et al*.,2025). These tendencies emphasize the necessity for improved approaches for early detection and prevention of this high-risk among individuals with both inherited exposure and variable lifestyle in the study area.

Hypertension and diabetes are multifactorial diseases ensuing from the interaction of genetic tendency and environmental factors, risk assessment which continue to assess these indicators in separation. Genetic susceptibility, often signified by family history or polygenic risk indicators, poses a reference point of susceptibility to development of the disease. However,

How to cite this article: Agani, S.N., Saleh I.M. & Akeyede I. (2026). Genetic-Lifestyle Risk Prediction Model for Hypertension and Diabetes Using Logistic Regression. *Journal of Basics and Applied Sciences Research*, 4(4), 196-202. https://dx.doi.org/10.4314/jobasr.v4i4.23

the indicator of this susceptibility is strongly formed by lifestyle factors such as BMI, physical activity, smoking, alcohol consumption, and dietary habits (Takase *et al*.,2024). Therefore, individuals that has a similar hereditary history may experience noticeably diverse disease consequences dependent on their behavioural and environmental factors. Their interaction between inherited and variable factors suggests that models focusing on the simple main effects which may underrate the positive disease risks and limit healthcare preventive measures in the study area. Hypertension and diabetes occur together in a patient with shared risk contributors such as age, body mass index (BMI), and genetic indicators. Hypertension and diabetes pose major health problems, affecting approximately 1.28 billion and 537 million adults worldwide, with the prevalence of diabetes expected to rise to 1.3 billion by 2050. diabetes and high blood pressure (BP) frequently co-occur in the same individual (WHO,2023; IDF,2021; Colussi *et al*.,2020). Hypertension and diabetes can lead to kidney diseases in the human body if not properly cared for. Hypertension and diabetes are classic examples of complex diseases that come from the interactions between inherited tendency and environmental conditions. Epidemiological studies consistently report trends of high prevalence of hypertension among individuals with diabetes, reinforcing the need to combine risk modeling (Adegoke *et al*.,2025). Importantly, hypertension and diabetes have some common pathophysiological systems, such as insulin resistance, vascular inflammation, endothelial dysfunction, obesity, and oxidative stress, proposing a cross-talk between diabetes and hypertension that could contribute to the development of other human diseases (Hezam *et al*.,2024).

Empirical reviews support the role of both genetic and lifestyle predictors in hypertension and diabetes development. (Liu *et al*.,2025) established a significant relationship between high BMI and high risk of hypertension among adolescents, stressing that overweight is a key variable predictor. (Tremblay et al.,2021) revealed that polygenic risk scores can improve the prediction of T2DM by capturing inherited susceptibility elsewhere which is a clinical indicator. Also, (Takase *et al*.,2024) further established that the combination of lifestyle factors and polygenic risk which improved the prediction of hypertension which indicates that genetic susceptibility and lifestyle factors affects the independent effects on the disease prevalence. In spite of these developments, gaps in methodological approaches remain as many existing studies focused on either hypertension or diabetes in isolation, rather than focusing on their co-occurrence and shared risk indicators. Furthermore, some studies have focused on relationship analyses without developing risk prediction models that join in genetic and lifestyle variables simultaneously. Some predictive models have been developed with limited consideration to assess the model calibration, discrimination, and their interaction effects, especially in the study populations where genetic, environmental factors that is different from the models that existed in those. This represents a significant knowledge gap in absence of specific model, the validated models may reduce the applicability of current prediction tools for routine clinical and public health use in Lafia town in particular. Statistically, logistic regression gives a suitable and understandable modelling binary disease outcomes such as hypertension and diabetes status. Logistic regression allows for the assessment of individual risk probabilities and quantify the independent and combined effects of several predictors through odds ratios. Additionally, the introduction of interaction terms allows the assessment on the effect of genetic tendency if it varies across levels of lifestyle factor, hence, focusing on the gene–environment interaction highlight multifaceted disease development. The methodological strength of logistic regression techniques is suitable introducing several indicators while upholding quantifiable interpretability, especially in developing risk prediction models that should be use for real-world application in healthcare centers.

MATERIALS AND METHODS

A cross-sectional analytical research design is adopted. This design allows for the assessment of associations between genetic and lifestyle variables and disease outcomes at a single point in time and is suitable for risk prediction modeling of hypertension and diabetes using logistic regression. The study was conducted in Lafia town with estimated Population, N=567,694 according world Population Review, (WPR,2026) City's population with a sample size, n=1000 drawn from the population using the Event per variable formula by (Peduzzi *et al.*,1996) given as;

$$n = \frac{EPV \times K}{p} \quad (1)$$

where; K is the number of predictor variables and P is the proportion of small outcomes in the population.

Data Collection and Analysis

A secondary data was collected from Discrete Health Information System (DHIS2) of Nasarawa state and Kaggle.com containing information on hypertension status, diabetes status, genetic risk indicators (such as family history, Body Mass Index (BMI), and lifestyle characteristics and was analyzed using IBM SPSS version 26. The analyses involve data cleaning, descriptive statistics, multicollinearity assessment, model fitting, and predictive performance evaluation.

MODEL SPECIFICATION

Classical Logistic Regression Model

Multiple logistic regressions are used to model the probability of developing hypertension or diabetes as a function of genetic-lifestyle predictors. Regression coefficients will be interpreted using odds ratios and confidence intervals. The classical multiple logistic regression models will be specified as:

$$\text{In}\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad (2)$$

Where, p is the probability of having the disease, $\left(\frac{p}{1-p}\right)$ is the Odds ratio of the probability of success to the probability of failure, β_0 is the intercept and the log-odds when all predictors are zero, β_1, β_2,… is the regression coefficients in log-odds and X_1, X_2, ...is the Predictors variables.

LASSO Logistic Regression

This performs variable selection by shrinking insignificant coefficients to zero. Least Absolute Shrinkage and selection Operators (LASSO) is used to construct a risk prediction nomogram for hypertension in T2DM patients, effectively reducing variables to a high-performing subset (Li, *et al.*, 2024). The LASSO Logistic Regression model with its probability function is shown below;

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_{i_1} + \beta_2 X_{i_2} + \cdots + \beta_p X_p \quad (3)$$

$$\text{Minimize}\left\{-\ \ell(\beta) + \lambda \sum_{j=1}^{p} |\beta_j|\right\} \quad (4)$$

Where; $\ell(\beta)$ is the log probability of the logistic model, λ is the tuning parameter that controls the degree of shrinkage and β_j is the regression coefficients.

RESULTS AND DISCUSSION

Descriptive statistics were used in the analysis of variables to summarize the whole of the study population.

Table 1: Distribution of the study of Categorical Variables

Variable	Category	Percentage (%)	Parameter Coding
Hypertension	Yes	38.2	1
	No	61.8	0
Diabetes	Yes	29.5	1
	No	70.5	0
Gender	Male	52.4	1
	Female	47.6	0
Smoking	Yes	34.1	1
	No	65.9	0
Physical Activity	Active	41.3	1
	Inactive	58.7	0
Alcohol	Yes	54.4	1
	No	45.6	0
Family History	Yes	44.6	1
	No	55.3	0

From the result above, it shows that 29.5% of patients in the population study had diabetes, while 70.5% did not have diabetes, while 38.2 % of the patients had hypertension, while 61.8% did not have hypertension. 52.4 % of the patients in the population study area were male, while 47.6 % of the gender was female. The patients who smoke in the population study are 34.1while 65.9% do not smoke. The patients in the study area that physically active had 41.3%, while 58.7% are inactive, with a mean of 0.67. The respondents in the study area who had a family history is 44.6%, while those who do not have is 55.3 %. Those that takes alcohol is 54.4 % while 45.6% do not take alcohol. The table above shows that there is an occurrence of non-communicable diseases and lifestyle risk factors like physical inactivity, alcohol, and gender in the study area. This implies that there is a high rate of health challenges in developing countries, as lifestyle habits and environment lead to the exposure of hypertension and diabetes risk factors (WHO, 2021).

Identification of significant genetic and lifestyle risk factors associated with hypertension and diabetes

Table 2: Diabetes Model

Variables	B	S.E.	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Age	0.052	0.015	12.044	0.002	1.053	1.022	1.084
BMI	0.095	0.020	22.560	0.001	1.099	1.057	1.143
Physical Activity	-0.420	0.177	5.621	0.018	0.657	0.464	0.929
Family History	0.880	0.164	29.156	0.001	2.410	1.751	3.318
Smoking	0.350	0.171	4.191	0.041	1.419	1.015	1.984

From the table above, it shows that Age has a coefficient of 0.052, a p-value of 0.002, an odds ratio of 1.053 with a 1.020 – 1.073 confidence interval at 95%, which shows that it is a statistically significant predictor. This shows that when the age of an individual increases, the individual is likely to develop diabetes. BMI has a coefficient of 0.095, p-value of 0.001, odds ratio of 1.099 with a 1.050 – 1.120 confidence interval at 95%, and BMI is also a significant variable of diabetes. This implies that age, BMI increase the risk of diabetes. This is in line with the report of the American Diabetes Association (ADA), in 2024, that excessive body weight contributes to diabetes and insulin resistance, and its mechanism.

Physical activity has a coefficient of -0.420, p = 0.018, the odds ratio 0.657, and a 0.464 – 0.929 confidence interval at 95%, which shows that physically active patients were likely to have diabetes, showing a significant negative connection with diabetes. This implies that Physical activity helps to lower the risk of diabetes by enhancing glucose metabolism through frequent exercise. Family history has a coefficient of 0.880, p- value of 0.001 with odds ratios of 2.410 and confidence interval of 1.751 – 3.318 at 95% confidence interval, which shows that family History is a significant predictor of diabetes, and the individuals in the study area were more likely to develop the disease compared to those without family history. This implies that Family History is also a strong and significant predictor, which becomes a key predictor of diabetes, showing the genetic susceptibility. Smoking has a coefficient of 0.350, p-value = 0.041, odds ratio of 1.419, and a 1.015 – 1.984 Confidence interval at 95%. This shows that the percentage of those that smokes is 41.9% and are more likely to have diabetes than those who do not smoke in the study area. The model revealed that genetic-lifestyle risk factors significantly contribute to diabetes prevalence in the study area.

Table 3: Hypertension Model

Variables	B	S.E.	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Age	0.045	0.012	14.062	0.001	1.046	1.020	1.073
BMI	0.082	0.012	20.756	0.001	1.085	1.050	1.120
Physical Activity	-0.365	0.168	4.731	0.030	0.694	0.500	0.960
Family History	0.720	0.154	22.421	0.000	2.055	1.520	2.770
Smoking	-0.365	0.168	5.302	0.030	1.509	1.060	2.140

From the table above, it shows that Age has the coefficient of 0.045, p-value of 0.001, odds ratio of 1.046 with 1.020 – 1.071 confidence interval which shows that it is statistically significant predictor. This implies that as individuals advances in age, they increase the odds of hypertension development in the study area. This implies that Age is connected to hypertension which shows the tendency of patients to develop the disease as he or she advance in years. This is in line with the finding of (Adegoke *et al*., 2022) on the epidemiological proving as age chip in to high blood pressure and vascular inflexibility.

BMI has a coefficient of 0.082, p-value of 0.001, odds ratio of 1.085 with 1.048 – 1.124 confidence interval at 95%, BMI is also a significant variable, patients with higher BMI are likely to become hypertensive as there is a strong connection between BMI and hypertension. This emphasizes the purpose of excess weight in the body involving the heart and blood vessels

Physical activity has a coefficient of -0.365, p = 0.030, odds ratio with 0.694 and 0.500 – 0.960 confidence interval at 95%. This reveals that those who participate in physical activity regularly were less prone to hypertension which suggests the important of physical activity in the management of hypertension. Family history has a p-value of 0.001, odds ratios of 2.054 with 95% confidence interval of 1.525 - 2.767, and a coefficient of 0.720, which shows that it is statistically significant and increases hypertension. The individual with this genetic tendency is more likely to have hypertension than those who do not have a family history. Smoking has a coefficient of 0.410, p = 0.021, an odds ratio of 1.507 with a 1.063 – 2.137 Confidence interval at 95%, which shows that it increases the risk moderately and individuals that smokes were more likely to have hypertension than those who do not smoke. This implies that smoking causes damage on the blood vessels and the entire cardiovascular system.

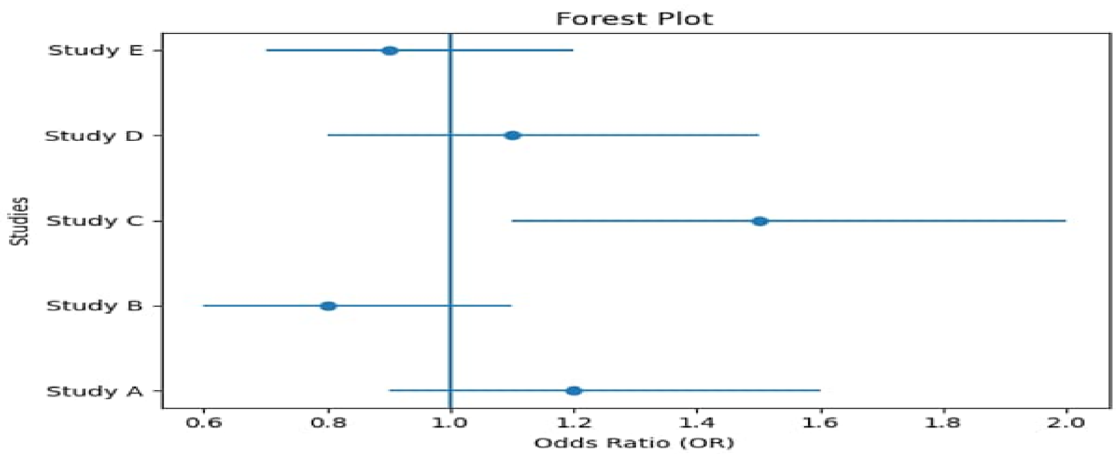

Figure 1: Forest plots of Odds Ratios

The figure above is a forest plot, which indicates a vertical line at an odds ratio of 1, which shows no effect. The variables to the right of the vertical line raise the risk of diseases, while the variables to the left of the vertical line reduce the risk of disease. If the variables are not significant, it shows that the confidence interval crosses the value 1; if it does not cross the value 1, it means that the variables are significant.

The plot reveals that family history is far to the right and a higher risk, body mass index, and smoking are clear risk factors. Physical activity reveals protective effects on the left side. Also, the figure reveals a steady increase in age.

Performance of the Models Using Discrimination and Calibration Measures

The area under the curve, specificity and sensitivity is used to assess the model's performance

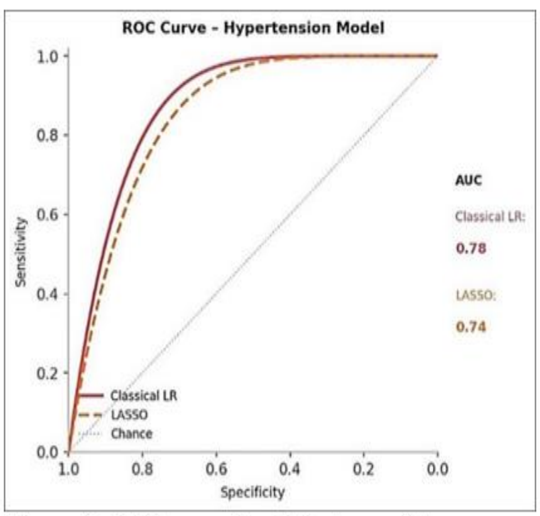

Figure 3: AUC curve for diabetes model

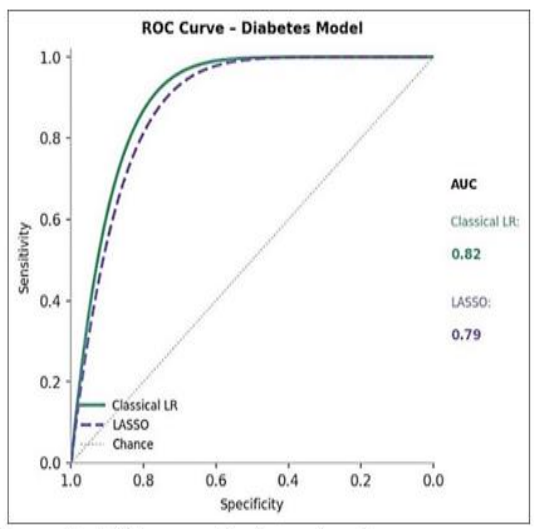

Figure 4: AUC curve for hypertension

Figure 3 and 4 show the area under curve for diabetes model reads 0.78 which shows that the model can accurately carry out discriminately assessment of diabetes while hypertension reads 0.82 which show that the hypertension model can predict the performance well and group individuals with the diseases and without the diseases. The area under the curve shows the goodness-of-fit tests to predict the performance of the diseases. The models show the level of sensitivity and specificity acceptability which is necessary for screening among the individuals in the study area.

The findings from the study show that genetic -lifestyle factors are significant in hypertension and diabetes development amongst the individuals in the study area. The findings showed that BMI, age, physical activity, family history and smoking significantly have effect on the condition of hypertension and diabetes. This implies that several inherited traits interact with environmental factors causes hypertension and diabetes highlighting the usefulness of combining genetic -lifestyle factors in the prediction and prevention of hypertension and diabetes.

CONCLUSION

The study showed that age is significantly connected with hypertension and diabetes which show that the likelihood of having hypertension and diabetes rises as the individuals grows older. The findings are line with the studies of (Adegoke *et al*., 2022) and (Adeyemi *et al*.,2024), that growing old significantly affects and increase the problem of heart and blood vessels diseases in developing countries. The suggests that the health of older adults need to be monitored and take some preventive measures to reduce the complexity of these diseases.

The findings showed that Body Mass Index (BMI) is also the strongest and significant predictors of hypertension and diabetes. The individual with normal body weight has less odds of developing hypertension and diabetes more than the Individuals with higher BMI as they have significant and higher odds of having hypertension and diabetes. This finding is in line with previous studies that overweight and obesity adds to unusual metabolic processes in the body, insulin resistance, and stress in the heart and blood vessels. The findings assert to the report of the American Diabetes Association (2024), which underscore obesity as a key risk factor for hypertension and diabetes. The suggests that good diet, having regular exercise is important in managing weight in order to reduce heart diseases condition.

The findings showed that smoking status of the individual in the study area was also significant predictor of hypertension and diabetes. Those that smoke has a higher tendency of having hypertension and diabetes diseases compared to those that does not smoke. The finding agrees with previous studies epidemiological which revealed that smoking causes harm to endothelial, rise blood pressure, and metabolic dysfunction. The findings suggest that smoking cessation programs should be organized by public health as part of their intervention programs to prevent heart diseases that may lead to hypertension and diabetes in the study area.

The findings showed that physical activity had a protective effect against hypertension and diabetes. Individuals who participate in exercise are less likely to have hypertension and diabetes than those who do not participate in physical activity, which shows the importance of physical activity in enhancing the flow of blood in the body, burning accumulated fat from the body, and maintaining the functioning of the heart and blood vessels. These findings support the report of the World Health Organization (2021), which indicates that physical inactivity is a key factor that generally affects non-communicable diseases. This suggests that active lifestyles should be encouraged through community-based health education programs, which significantly reduce occurrence of hypertension and diabetes in the study area.

Family history revealed that it is a significant predictor in the models, which showed that individuals with a family history of hypertension or diabetes were at higher risk of having the diseases. These results underscore the contribution of inherited genetic susceptibility to disease occurrence. This result asserts to the work of Carlsson (2022), who highlighted that family inherited tendency raises the exposure of individuals to chronic diseases, particularly when it joined with other bad lifestyles.

The present study addresses the identified gap by developing and validating a genetic–lifestyle risk prediction model for hypertension and diabetes using logistic regression techniques specifically, to identify the significant genetic and lifestyle risk factors associated with hypertension and diabetes, apply logistic regression to evaluates individual disease risk probabilities. In addition, the study evaluates the predictive performance of the model using discrimination and calibration measures, including receiver operating characteristic (ROC) analysis, with Area under the Curve (AUC) values of 0.78 for hypertension and 0.82 for diabetes indicating good discriminative ability of the developed models in identifying between those that have the disease and those that does not have the disease which revealed that the predictive accuracy is acceptable which shows the usefulness of penalized regression methods in handling multivariable disease prediction models, especially in studies involving correlated predictors. The result is in agreement with Pavlou *et al*. (2016) who emphasized that LASSO regression enhances model reliability and reduces overfitting.

CONFLICT OF INTEREST

There is no conflict of interest among the authors.

REFERENCE

Adegoke, T. A., Olorunfemi, O. A., & Odeniyi, I. A. (2025). Clinical validation of risk score systems for cardio-metabolic disorders in Western sub-Saharan Africa. *African Journal of Medicine and Medical Sciences*, 54(1), 45–56.

Adeyemi, O. M., Alare, K., Afolabi, S. A., et al. (2024). Prevalence and risk factors of type 2 diabetes mellitus in

Nigeria: A systematic review and meta-analysis. *Clinical Diabetes and Endocrinology,* 10(43).

American Diabetes Association. (2023). Cardiovascular disease and risk management: Standards of care. *Diabetes Care*;46(Supplement_1): S1–S4 https://doi.org/10.2337/dc23-Sint

Carlsson, S. (2022). Lifestyle or environmental influences and their interaction with genetic susceptibility on the risk of LADA. Frontiers in Endocrinology,13: 917850. DOI: 10.3389/fendo.2022.917850

Colussi, G. L., Da Porto, A. & Cavarape, (2020) A. Hypertension and type 2 diabetes: lights and shadows about causality. *Journal of Human Hypertension* vol. 34 91–93. Preprint at doi:10.1038/s41371-019-0268-x CrossRef PubMed Google Scholar

Hezam, A. A. M., Shaghdar, H. B. M., & Chen, L. (2024). The connection between hypertension and diabetes and their role in heart and kidney disease development. *Journal of Research in Medical Sciences*, 29(1), 22. https://doi.org/10.4103/jrms.jrms_470_23.

International Diabetes Federation. IDF (2021). Diabetes Atlas, 10th edition. Brussels, Belgium: Website: https://www.diabetesatlas.org

Lazzarini, P. A., Pacella, R. E., Armstrong, D. G., & van Netten, J. J (2023) Global epidemiology of diabetes-related foot ulcers and cardiovascular multi-morbidities: A systematic review and meta-analysis. *Diabetologia,* 66(4), 613–627. https://doi.org/10.1007/s00125-022-05856-w

Li, G. (2025). Research of prediction diabetes risk using logistic regression models. Highlights in Science, Engineering and Technology, 140

Liu, R., Hu, H., Cao, C., Han, Y., Bai, Y., & Feng, W. (2025). Sex differences in the relationship between body mass index in Chinese adolescents and future risk of hypertension: a decade-long cohort study. BMC pediatrics, 25(1), 187

Pavlou, M., et al. (2016). Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine*, 35(7), 1159-1177. https://doi.org/10.1002/sim.6782

Takase, M., Hirata, T., Nakaya, N., Nakamura, T., Kogure, M., Hatanaka, R.,Nakaya, K., Chiba,I., Kanno, I., Nochioka, K., Tsuchiya, N., Narita, A., Metoki, H., Satoh, M., Obara, T., Ishikuro, M., Ohseto, H., Uruno, A., Kobayashi, T., Fuse, N. (2024). Associations of combined genetic and lifestyle risks with hypertension and
home hypertension. *Hypertension Research* 47(8),2064 - 2074.https://doi.org/10.1038/s41440-024-01705-8

Tremblay, J., & Hamet, P. (2021). Genetics of hypertension: 2021 update. Current Hypertension Reports, 23(4), 22. https://doi.org/10.1007/s11906-021-01138-1

World Health Organization. (2021). Global status report on noncommunicable diseases Geneva: WHO

World Health Organization. (2023) Hypertension [cited Feb 27, 2023]. Available from: https: // www.who.int/news-room/fact-sheets/detail/hypertension.

World Population Review (2026, May 18). City's Projected Population https://worldpopulationreview.com/cities/nigeria/lafia.Retrieved from google