Original Article

# Risk Prediction of high blood glucose among women (15–49 years) and men (15–54 years) in India: An analysis from National Family Health Survey-5 (2019–21)

**Anjan Kumar Karri, Visweswara Rao Guthi, Sri Sai Githa P**

*Department of Community Medicine, SVIMS-Sri Padmavathi Medical College for Women, Tirupati, Andhra Pradesh, India*

## Abstract

**Context:** Approximately 500 million individuals worldwide are known to have diabetes, representing roughly 1 out of every 11 adults in the world. Approximately 45.8% of adult diabetes cases are believed to be undiagnosed. **Aim:** This study aimed to identify the predictors for high blood glucose and to develop a risk score which helps in early detection of high blood glucose among Indian men (15–54 years) and women (15–49 years). **Methods and Material:** This study utilised data from the National Family Health Survey-5, which were gathered between 2019 and 2021. The study population comprises women aged 15–49 years and men aged 15–54 years in India. **Statistical Analysis Used:** A logistic regression analysis was conducted to determine the predictors of high blood glucose. The results were expressed as odds ratios with 95% confidence intervals. The risk score for high blood glucose was derived through variable shrinking and by employing regression coefficients obtained from the standard logistic regression model. Data were analysed using IBM SPSS version 26. **Results:** The prevalence of high blood glucose in India was 9.3%. The study findings indicated an association between age and the occurrence of high blood glucose levels. The prevalence of high blood glucose was higher among males (11.1% vs 7.5%), individuals living in urban areas (10.7% vs 8.9%), those with a waist circumference exceeding the specified limit (11.7% vs 5.9%), and individuals who were overweight or obese (11.3%). The prevalence of high blood glucose was higher among alcoholics (13.2% vs 8.8%) and various forms of tobacco users (12.1% vs 8.4%). **Conclusions:** Age, sex, place of residence (urban), consumption of alcohol, hypertension, and waist circumference were found to be the significant predictor variables and were used to develop the risk prediction score using the logistic regression model.

**Keywords:** High blood glucose India, NFHS, predicting the risk, risk factors

## Introduction

Approximately 500 million individuals worldwide are known to have diabetes, representing roughly 1 out of every 11 adults in the world.[1] Approximately 45.8% of adult diabetes cases are believed to be undiagnosed.[2] Diabetes is a significant global health crisis in the 21st century, being one of the top ten causes of death along with cardiovascular disease (CVD), respiratory disease, and cancer.[3] It is projected that over 592 million individuals will succumb to diabetes by the year 2035.[4] The World Health Organisation (WHO) reports that the occurrence of diabetes is increasing at a fast pace in nations with lower and moderate incomes.[5]

The rapid socioeconomic transformation, along with the process of urbanisation and industrialisation, are the primary catalysts for

**Address for correspondence:** Dr. Visweswara Rao Guthi, Department of Community Medicine, SVIMS-Sri Padmavathi Medical College for Women, Tirupati, Andhra Pradesh, India.
E-mail: vissumbbs@gmail.com

**Received:** 29-05-2024 **Revised:** 28-07-2024
**Accepted:** 12-08-2024 **Published:** 18-11-2024

Access this article online
Quick Response Code:
Website: http://journals.lww.com/JFMPC
DOI: 10.4103/jfmpc.jfmpc_929_24



**How to cite this article:** Karri AK, Guthi VR, Githa PS. Risk prediction of high blood glucose among women (15–49 years) and men (15–54 years) in India: An analysis from National Family Health Survey-5 (2019–21). J Family Med Prim Care 2024;13:5312-9.

the worldwide surge in the diabetes epidemic. Other contributing factors, such as population growth, unhealthy dietary patterns, and a sedentary way of life, also exert significant influence.[6] Unhealthy food, high blood pressure, high fasting plasma glucose, high cholesterol, and overweight have all experienced a surge in every state of India. These risk factors now collectively account for 25% of the overall disease burden in the country.[7] Diabetes has experienced the most significant rise in disease burden, as measured by DALYs, among the major non-communicable diseases since 1990. The prevalence of diabetes rose from 26 million individuals in 1990 to 65 million individuals in 2016. The incidence of DALY rate due to diabetes has significantly risen in all states of India from 1990 to 2016, with a four-fold difference in DALY rate between the states in 2016.

The less developed states have experienced the largest increase in diabetes prevalence and DALY rate, after adjusting for age differences. The prevalence of obesity, a significant contributing factor to diabetes, has nearly risen in every state of India since 1990. In 2016, the number of individuals with diabetes in India was twice as high as the global average for every 100 overweight individuals, highlighting the elevated risk of diabetes in India.[8] Diabetes is a chronic condition that gradually worsens and results in severe complications, which are linked to higher expenses for the individual's family, the community, and the healthcare system. Unregulated diabetes results in a higher likelihood of developing vascular disease.

The majority of the impact of type 2 diabetes is due to damage in the larger blood vessels (cardiovascular, cerebrovascular, and peripheral artery disease) and the smaller blood vessels (diabetic retinopathy, nephropathy, and neuropathy).[6,9] Risk prediction models, which rely on patient features as predictors, are fundamental tools in contemporary clinical care for forecasting health outcomes.[10] Risk models are constructed by including many risk factors, usually derived from patient characteristics, which are believed to be associated with the specific health event under consideration. The selection of these predictions is typically based on clinical expertise and a thorough examination of the literature. The risk model can utilise patient characteristics to accurately determine the likelihood of a patient experiencing the event.[11]

The NFHS-5 survey was carried out by the Indian Ministry of Health and Family Welfare from 2019 to 2021. The survey provides crucial data on health and family well-being and measurements of biomarkers such as height, weight, waist circumference, blood pressure, and blood glucose levels. In this study, we aimed to identify the predictors for high blood glucose and to develop a risk score which helps in early detection of high blood glucose among Indian men (15–54 years) and women (15–49 years).

## Material and Methods

### Data source

The present study utilised data obtained from individuals of reproductive age group, both male and female, as part of the National Family Health survey-5. The NFHS-5 sample has been designed to generate accurate measurements of important indicators at both the national and state levels. Additionally, it also provides estimates for the majority of critical indicators at the district level, including all 707 districts in India as of 31 March 2017. The overall sample size comprises over 610,000 households. The NFHS-5 has implemented a standardised sample design that utilises a two-stage cluster sampling methodology. During the initial phase, primary sampling units (PSUs) were chosen using the proportionate to size technique (PPS). In rural regions, the PSU was the village, while in urban areas, it was the census enumeration block (CEB). In the second stage, a set number of 22 households (PSUs) were picked using systematic random sampling from newly constructed lists of households residing in the selected PSUs. The list of households is generated by the process of mapping and conducting a household listing operation in each selected PSU prior to the household selection in the second stage.

Ethical policy and Institutional Review board statement: The Demographic and Health Surveys (DHS) Program's website hosts the NFHS-5, an anonymous dataset that is made available to the public and cannot be used to identify the survey respondents. The respondents provided signed consent after being fully informed about the survey's purpose and procedures. Only interviews were done after obtaining proper consent from each participant. We have obtained approval from the DHS IRB before doing analysis. Ethical approval was not needed as the analysis used secondary data available in the public domain. However Institutional Review board of Demographic and Health surveys programme approved the study protocol (AuthLetter_177766 dated 30, Dec 2022). The guidelines for data use as required by the DHS programme were strictly followed.

### Operational definitions

**Dependent Variable**: The study considered high blood glucose as the dependent variable. The participants of the NFHS-5 had their random blood glucose level checked using a finger-stick blood samples using the Accu-Chek Performa glucometer with glucose test strips for blood glucose testing. The study participants were classified as having high blood glucose if they responded affirmatively to the questions, they were diagnosed with high blood glucose by a doctor or other healthcare professionals on two or more occasions, currently taking medication to lower blood glucose, or if their random blood glucose level was ≥ 140 mg/dl.

**Independent variables**: Age, gender, place of residence (rural/urban), educational status, wealth index, BMI, waist circumference, WHR, use of tobacco, use of alcohol, hypertension status, religion, and caste. Age in completed years: According to NFHS-5, the whole country is divided into northern, northeast, south, east, and western regions and the place of residency into urban or rural areas. For educational status, individuals are categorised based on no schooling, those who completed pre-school or primary, secondary, or higher school. Wealth index was used instead of socio-economic scale, and its sub-scale ranges from poorest, poorer, middle, and richer to richest. Using BMI, individuals are categorised into thin, normal, overweight, and obese. This was the first time that NFHS-5 included waist and hip circumference measurements provided by using Gulick tapes for both eligible women and men for measurements of abdominal obesity. Current use of tobacco and alcohol is in any form.

## Statistical analysis

An initial descriptive analysis was conducted to ascertain the basic characteristics of the data. The results were presented in frequencies and percentages, along with 95% confidence intervals. Continuous values were expressed as mean ± standard deviation (SD). The bivariate analysis was performed using Pearson's Chi-square test. A logistic regression model was later adopted using the enter and forward likelihood ratio approach. The significance of the model was tested using the omnibus Chi-square test (18833616), the 2-log likelihood test (291761.86), and the pseudo-R square value of 0.077. The odds ratio, together with its related 95% confidence intervals, was computed to identify the relevant predictors for high blood glucose. Predictors that had a $P$ value less than 0.05 were considered to be statistically significant. The risk score for developing high blood glucose was calculated by applying the aforementioned techniques to reduce the variables and utilising regression coefficients derived from the standard logistic regression model.[11]

Individual risk for developing high blood glucose = exp (Individual risk score) ÷ [1 + exp (Individual risk score)]

The statistical package for data analysis was Statistical package for Social Sciences for Windows, version 26.0 (IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp).

## Results

Out of the total sample of 14,38,132, 7,21,688 were females in the age group of 15–49 years and 7,16,444 were males in the age group of 15–54 (household member code file IAPR7CFL). A total of 14,38,132 individual-level data were available. Among them, 1,83,585 individuals were excluded as they did not have at least one blood sugar measurement. Hence finally, 12,54,547 individuals' data were included in the final analysis [Figure 1]. Among them, there was an increased prevalence of high blood glucose as the age increased (13.6% in > 30 years). The prevalence of high blood glucose was higher in males than in females (11.1% vs 7.5%). Persons living in urban areas had a higher prevalence of high blood glucose (10.7% vs 8.9%). Individuals with primary education had a high blood glucose of 10.7%, and individuals with no school/pre-school have a high blood glucose of 10.2%. The richest wealth index group had high blood glucose of 10.8%, followed by the richer wealth index group with 10.3%. Individuals falling into overweight and obese categories based on their BMI had a high prevalence of high blood glucose. Males with abdominal circumference >90 cm and females with abdominal circumference of >80 cm have high prevalence of high blood glucose (11.7% vs 5.9%). Males with a waist hip ratio of >0.95 and females with a waist hip ratio of >0.85 had a high prevalence of high blood glucose (9.0% vs 6.8%). Persons using tobacco and alcohol had high prevalence of high blood glucose compared to those who did not (12.1% and 13.2%, respectively). High blood glucose was associated with hypertension in 18.7% of the individuals. High prevalence of high blood glucose was seen in Hindu and Christian religions (9.4% each), followed by Muslims and others (9.0% and 8.8%, respectively). The prevalence of high blood glucose was 8.3% in scheduled tribe, 9.0% in scheduled caste, 9.6% in other backward class, and 9.8% in individuals who fall into none of the above groups [Table 1].

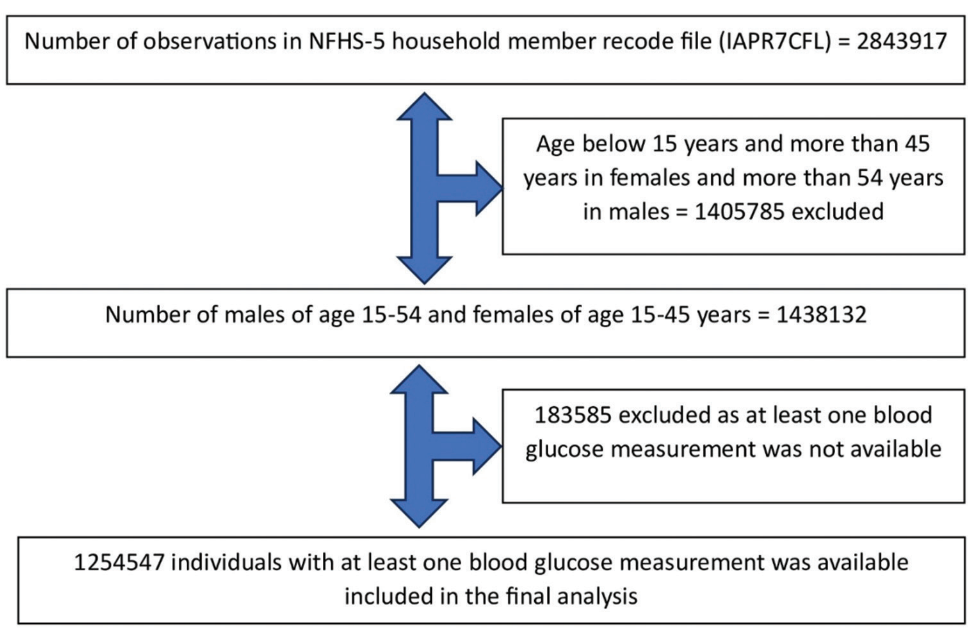


**Figure 1:** Sampling strategy

**Table 1: Distribution of high blood glucose according to socio-demographic factors, anthropometry, and habits**

| Variable | | Total | High blood glucose % (95% CI) | Chi-Square | *P* |
|---|---|---|---|---|---|
| Age | 15-18 years | 139256 | 3.1 (2.9-3.2) | 30398.18 | <0.001 |
| | 19-29 years | 456402 | 5.1 (5.0-5.1) | | |
| | >30 years | 658889 | 13.6 (13.5-13.7) | | |
| Gender | Male | 631455 | 11.1 (11.0.-11.2) | 5031.95 | <0.001 |
| | Female | 623092 | 7.5 (7.4-7.6) | | |
| Place of residence | Urban | 860922 | 3.9 (3.8-4.0) | 918.34 | <0.001 |
| | Rural | 944697 | 8.9 (8.8-9.0) | | |
| Schooling | No education, preschool | 200952 | 10.2 (10.0-10.4) | 753.92 | <0.001 |
| | Primary | 153369 | 10.7 (10.5-10.8) | | |
| | Secondary | 692864 | 9.0 (8.8-9.1) | | |
| | Higher | 206899 | 8.6 (8.5-8.8) | | |
| Wealth index | Poorest | 256720 | 7.9 (7.7-8.0) | 1648.56 | <0.001 |
| | Poorer | 279518 | 8.6 (8.4-8.7) | | |
| | Middle | 266552 | 9.4 (9.3-9.5) | | |
| | Richer | 244040 | 10.3 (10.1-10.4) | | |
| | richest | 207717 | 10.8 (10.6-10.9) | | |
| Religion | Hindu | 945769 | 9.4 (9.3-9.5) | 39.26 | <0.001 |
| | Muslim | 153580 | 9.0 (8.9-9.2) | | |
| | Christian | 93250 | 9.4 (9.2-9.6) | | |
| | Others | 118430 | 4.6 (4.5-4.7) | | |
| Caste | Scheduled caste | 247421 | 9.0 (8.8-9.1) | 454.32 | <0.001 |
| | Scheduled tribe | 245061 | 8.9 (8.8-9.1) | | |
| | Other backward class | 463125 | 9.6 (9.5-9.7) | | |
| | None of them | 242918 | 9.8 (9.6-9.9) | | |
| BMI | Underweight | 110265 | 4.7 (4.5-4.8) | 7622.21 | <0.001 |
| | Normal | 288460 | 5.5 (5.4-5.6) | | |
| | Overweight and obese | 224324 | 11.3 (11.1-11.5) | | |
| Waist circumference | <90 or<80 | 462318 | 5.9 (5.8-6.1) | 7474.94 | <0.001 |
| | >90 or>80 | 252537 | 11.7 (11.6-11.9) | | |
| Waist hip ratio | <0.950 or 0.85 | 336487 | 6.8 (6.6-6.9) | 1243.67 | <0.001 |
| | >0.950 or 0.85 | 377969 | 9.0 (8.9-9.1) | | |
| Use of tobacco | No | 950364 | 8.4 (8.3-8.5) | 3668.01 | <0.001 |
| | Yes | 302799 | 12.1 (11.9-12.2) | | |
| Use of alcohol | No | 1094598 | 8.8 (8.7-8.9) | 3168.17 | <0.001 |
| | Yes | 158348 | 13.2 (13.0-13.3) | | |
| Hypertension | Hypertensive | 220688 | 18.7 (18.5-18.58) | 28144.16 | <0.001 |
| | Normotensive | 1031251 | 7.2 (7.1-7.3) | | |

The independent variables such as age, place of living, education status, wealth index, BMI, waist hip ratio, waist circumference, use of tobacco, use of alcohol, caste, and hypertension were considered as significant predictors of having high blood glucose, and the prediction score was calculated after using forward logistic regression analysis [Table 2].

The fitted risk model was utilised to predict individual risk. Six predictors were chosen to create the risk model using standard and stepwise logistic regression based on their likelihood ratios [Table 3].

Individual risk score = intercept + ($b_{age}$ × age) + ($b_{gender}$ × gender) + ($b_{residence}$ × residence) + ($b_{\text{hypertension}}$ × hypertension) + ($b_{alcohol}$ × alcohol) + ($b_{\text{waistcircumference}}$ × waistcircumference).

b = Regression coefficients

For example, the risk score of a 45-year-old male with a waist circumference of 95 cm, living in an urban area with known hypertension, and consuming alcohol would be calculated as -5.767+ (0.051 × 45) + (-0.173 × 0) + (0.077 × 1) + (0.701 × 1) + (-0.078 × 1) + (0.021 × 95) = -0.777. His predicted risk of having high blood glucose was Exp (0.777) ÷ {1 + exp (0.777)} = 31.4% when compared to others without these risk factors.

## Discussion

This study was conducted to identify the predictors for high blood glucose and to develop a risk score which helps in early detection of high blood glucose among Indian men (15–54 years) and women (15–49 years).

The prevalence of high blood glucose in the current study increased with the increasing age of the study participants. Similar

**Table 2: Association of high blood glucose with socio-demographic factors, anthropometry, and habits**

| Variable | | P-value | Odds ratio | 95% confidence interval | |
|---|---|---|---|---|---|
| | | | | lower | upper |
| Age groups | 15-18 years | | | | |
| | 19-29 years | <0.001 | 1.249 | 1.181 | 1.321 |
| | >30 years | <0.001 | 2.338 | 2.203 | 2.480 |
| Place of living | Rural | | | | |
| | Urban | <0.001 | 0.945 | 0.921 | 0.969 |
| Education status | No education | | | | |
| | Primary | 0.031 | 1.038 | 1.003 | 1.073 |
| | Secondary | 0.005 | 1.040 | 1.012 | 1.070 |
| | Higher | 0.026 | 0.955 | 0.917 | 0.995 |
| Wealth index | Poorest | | | | |
| | Poorer | 0.072 | 1.031 | 0.997 | 1.065 |
| | Middle | 0.001 | 1.060 | 1.025 | 1.097 |
| | Richer | <0.001 | 1.085 | 1.046 | 1.125 |
| | Richest | 0.245 | 1.025 | 0.983 | 1.069 |
| BMI | Underweight | | | | |
| | Normal | 0.007 | 0.954 | 0.922 | 0.987 |
| | Overweight and obese | <0.001 | 1.346 | 1.297 | 1.398 |
| Waist hip ratio | <0.950 or 0.85 | | | | |
| | >0.950 or 0.85 | <0.001 | 1.117 | 1.091 | 1.143 |
| Waist circumference | <90 or<80 | | | | |
| | >90 or>80 | <0.001 | 1.332 | 1.298 | 1.367 |
| Use of tobacco | No | | | | |
| | Yes | 0.006 | 1.053 | 1.015 | 1.092 |
| Drinks alcohol | No | | | | |
| | Yes | 0.030 | 0.927 | 0.865 | 0.993 |
| Caste | Scheduled caste | | | | |
| | Scheduled tribe | 0.067 | 0.956 | 0.925 | 0.989 |
| | Other backward class | 0.070 | 1.017 | 0.973 | 1.063 |
| | None of them | 0.000 | 0.836 | 0.797 | 0.878 |
| Hypertension | Normotensive | | | | |
| | Hypertension | 0.000 | 2.188 | 2.138 | 2.238 |

**Table 3: Logistic Regression Model**

| Variables | Parameter coding | B-coefficient |
|---|---|---|
| Age | Actual age | 0.051 |
| Sex | Male 0 | -0.173 |
| | Female 1 | |
| Place of residence | Rural 0 | 0.077 |
| | Urban 1 | |
| Drinks alcohol | No 0 | -0.078 |
| | Yes 1 | |
| Hypertension | Normotensive 0 | 0.701 |
| | Hypertensive 1 | |
| Waist circumference | Actual waist circumference | 0.021 |
| Constant | | -5.767 |

results are observed in the studies conducted by Mathur *P* and Tripathy *et al*.[12,13]

In a study conducted by ICMR-NCDIR 2017-18, the prevalence of high blood glucose was higher in males when compared to females, which is corresponding to our present study.[14] We noted an increased prevalence of high blood glucose in urban areas and was comparable with the study conducted by Saeedi *P et al*.[1]

The results of our study indicate a significant association between hypertension and elevated blood glucose levels, which aligns with the findings of the study conducted by Tripathy *et al*.[13]

Lindstrom and Schulze MB conducted studies on prediction models using non-invasive measures. These measures were based on information such as age, sex, height, waist circumference, BMI, and history of hypertension.[15,16] In the study conducted by Tripathy *et al*.,[13] it was shown that age group, hypertension, and obesity were significant risk variables related with DM in a multivariate regression model. The above study findings correlated well with our study.

Multiple epidemiological studies have shown that cigarette smoking is a distinct risk factor for type 2 diabetes.[17,18] Elderly individuals who consumed alcohol showed a 1.56-fold increased likelihood to have elevated fasting blood glucose levels.[19] Long-term alcohol consumption is recognised as a potential contributing factor to the development of type 2 diabetes mellitus (T2DM).[20] Similar with the above studies, we also have found a significant relation between use of alcohol and tobacco with high blood glucose levels.

The increasing incidence of diabetes (high blood glucose) is attributed to rapid urbanisation, changing lifestyles, and uneven dietary patterns.[21] By leveraging machine learning techniques, medical professionals can evaluate the probability of diabetes. Numerous predictive models for diabetes have been created and utilised, with logistic regression and a machine learning algorithm-based classification tree being prominent choices. Analysing risk factors through data exploration can reveal correlations between various factors and diabetes.[22,23]

Precise forecasting of the risk of type 2 diabetes enables doctors to pinpoint individuals with a high susceptibility to the condition. This identification offers an opportunity to implement preventive actions like recommending dietary adjustments, regular physical activity, or initiating interventions like metformin to delay the onset of type 2 diabetes.[24] Addressing the significant morbidity and mortality linked to diabetes and its complications is critical. Despite awareness, some individuals persist with behavioural risk factors, heightening their complication risks. Regular prevalence assessments, awareness campaigns, and effective treatment and control measures are essential for diabetes prevention and improving care quality.[12]

In conclusion, the study emphasises the urgent necessity of tackling the widespread threat posed by diabetes mellitus risk factors in India. By implementing interventions and policies grounded in evidence, stakeholders can bring out substantial transformations. These can range from measures aimed at controlling tobacco usage to initiatives that encourage healthier lifestyles and responsible environmental practices.[25] Existence of significant shortcomings in terms of comprehensive risk factor data, insufficient coverage across different regions and demographic groups, and the absence of a standardised approach, are the major deficiencies. Overcoming these flaws is crucial for achieving superior and more efficient control of diabetes in India.[26]

The significant predictor variables utilised to generate the risk prediction score using a logistic regression model were age, sex, place of residence (urban), use of alcohol, hypertension, and waist circumference.

A study conducted by Liangjun Jiang *et al*.[27] in Guangzhou identified several significant predictors of diabetes, including age, BMI, smoking, staple food consumption, exercise frequency and duration, alcohol use, and blood pressure. In another study conducted by Sia K Nicolaisen in Denmark, the comprehensive diabetes prediction model incorporated several factors, namely, HbA1c, age, sex, body mass index (BMI), usage of antihypertensive drugs, presence of pancreatic disease or cancer, self-reported diet, medical advice to lose weight or modify dietary habits, availability of a confidant, and self-assessed health status.[28] Kedir N Turi *et al*.[29] created a spline regression model known as multivariate adaptive regression splines (MARS). This study found a non-linear relationship between systolic blood pressure, diastolic blood pressure, sleep duration, waist circumference, and the occurrence of diabetes.

In this study, we have used the logistic regression model to predict the high blood glucose. There are other methods like Support Vector Machines,[30] Random Forest Classifier,[31] Gradient Boosted Trees,[32] and Weighted Ensemble Model,[33] which can be used to predict the disease risk.

The Government of India (GoI) established the National Programme for Prevention and reduction of non-communicable diseases (NP-NCD) to reduce NCDs, including diabetes. While the program has been successful thus far, it is not without flaws.[34] Some of the programme's goals have been achieved, including the early screening and diagnosis of NCDs like diabetes, which is essential to addressing problems early on.[35] The recent ICMR-INDIAB study has highlighted that, in India, objectives pertaining to the management of diabetes are not being met optimally.[36]

The primary limitation of the study is that the diagnosis of diabetes was based on a single-finger stick random blood glucose measurement, which may not correspond to the preferred HbA1c testing method considered as the most accurate. In addition, despite a small proportion of individuals who did not respond or from whom blood glucose samples could not be collected, this potential bias does not invalidate our estimates. The study was conducted exclusively on women aged 15–49 and males aged 15–54. Therefore, it is not applicable to persons in different age groups.

Despite these limitations, this study provides a risk prediction model that aids in early detection of high blood glucose levels among the adult Indian population. These findings have the potential to assist in the monitoring and development of national protocols for the control and treatment of diabetes in the country.

## List of abbreviations

| Abbreviations | Definition |
|---|---|
| NFHS | National Family Health survey |
| DALY | Disability Adjusted Life Years |

## Acknowledgements

| Name | Role |
|---|---|
| Demographic Health Survey (DHS) | Approving and providing data set for analysis |

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## References


1. Saeedi P, Petersohn I, Salpea P. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract 2019;157:107843.
2. Beagley J, Guariguata L, Weil C. Global estimates of undiagnosed diabetes in adults. Diabetes Res Clin Pract 2014;103:150–60.
3. International Diabetes Federation. IDF Diabetes Atlas. 9th ed. Brussels, Belgium: International Diabetes Federation; 2019. 2. World Health Organization. The top 10 causes of death. Available from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [Last accessed on 2023 Jun 04].
4. Tao Z, ShiA, Zhao J. Epidemiological perspectives of diabetes. Cell Biochem Biophys 2015;73:181-5.
5. World Health Organization. Diabetes. Available from: https://www.who.int/health-topics/diabetes#tab=tab_1. [Last accessed on 2023 Jun 04].
6. Pradeepa R, Mohan V. Prevalence of type 2 diabetes and its complications in India and economic costs to the nation. Eur J Clin Nutr 2017;71:816-24.
7. Institute for Health Metrics and Evaluation (IHME). Disease burden initiative in India. Available from: https://www.healthdata.org/disease-burden-india. [Last accessed on 2023 Jun 04].
8. Public Health Foundation of India (PHFI). NCDs and Suicides State-Level Disease Burden Estimates. Available from: https://phfi.org/the-work/research/ncds-and-suicides-state-level-disease-burden-estimates. [Last accessed on 2023 Jun 04]
9. Van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. The global burden of diabetes and its complications: An emerging pandemic. Eur J Cardiovasc Prev Rehabil 2010;17(Suppl 1):S3-8.
10. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? BMJ 2009;338:b375.
11. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, *et al.* How to develop a more accurate risk prediction model when there are few events. BMJ 2015;351:h3868.
12. Mathur P, Leburu S, Kulothungan V. Prevalence, awareness, treatment and control of diabetes in india from the countrywide national NCD monitoring survey. Front Public Health 2022;10:748157.
13. Tripathy JP, Thakur JS, Jeet G, Chawla S, Jain S, Pal A, *et al.* Prevalence and risk factors of diabetes in a large community-based study in North India: Results from a STEPS survey in Punjab, India. Diabetol Metab Syndr 2017;9:8.
14. ICMR-NCDIR. National Noncommunicable Disease Monitoring Survey (NNMS) 2017–18. 2020. Available from: https://www.ncdirindia.org/nnms/resources/NNMS%20 2017-18%20-%20Report.pdf. [Last accessed on 2023 Jun 04].
15. Lindström J, Tuomilehto J. The diabetes risk score: A practical tool to predict type 2 diabetes risk. Diabetes Care 2003;26:725–31.
16. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M, *et al.* An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007;30:510–15.
17. Foy CG, Bell RA, Farmer DF, Goff DC Jr, Wagenknecht LE. Smoking and incidence of diabetes among U.S. adults: Findings from the insulin resistance atherosclerosis study. Diabetes Care 2005; 28:2501–7.
18. Beziaud F, Halimi JM, Lecomte P, Vol S, Tichet J. Cigarette smoking and diabetes mellitus. Diabetes Metab 2004;30:161-6.
19. Gupta A, Belwal R, Ramakrishnan L, Khenduja P, Kapil U. Association of tobacco and alcohol consumption with cardiovascular risk factors among elderly population in India. J Family Med Prim Care 2020;9:5242-8.
20. Kim SJ, Kim DJ. Alcoholism and diabetes mellitus. Diabetes Metab J 2012;36:108-15.
21. Ismail L, Materwala H, Al Kaabi J. Association of risk factors with type 2 diabetes: A systematic review Comput Struct Biotechnol J 2021;19:1759-85.
22. Dritsas E, Trigka M. Data driven machine-learning methods for diabetes risk prediction. Sensors (Basel) 2022;22:5304.
23. Joshi RD, Dhakal CK. Predicting type 2 diabetes using logistic regression and machine learning approaches. Int J Environ Res Public Health 2021;18:7346.
24. Hahn SJ, Kim S, Choi YS, Lee J, Kang J. Prediction of type 2 diabetes using genome-wide polygenic risk score and metabolic profiles: A machine learning analysis of population-based 10-year prospective cohort study. EBioMedicine 2022;86:104383.
25. Sharma M, Gaidhane A, Choudhari S G. A comprehensive review on trends and patterns of non-communicable disease risk factors in India. Cureus 2024;16:e57027.
26. Nethan S, Sinha D, Mehrotra R. Non communicable disease risk factor and their trends in India. Asian Pac J Cancer Prev2017;18:2005-10.
27. Jiang L, Xia Z, Zhu R, Gong H, Wang J, Li J, *et al.* Diabetes risk prediction model based on community follow-up data using machine learning. Prev Med Rep 2023;35:102358.
28. Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L. Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. BMJ Open Diabetes Res Care 2022;10:e002946.
29. Turi KN, Buchner DM, Grigsby-Toussaint DS. Predicting risk of type 2 diabetes by using data on easy-to-measure risk factors. Prev Chronic Dis 2017;14:160244.
30. Cortes C, Vapnik VN. Support-vector networks. Mach Learn 1995;20:273–97.
31. Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on document analysis and recognition. Vol 1. IEEE; 1995. p. 278–82.
32. Quinlan JR. Induction of decision trees. Mach Learn 1986;1:81–106.
33. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat 2001;29:1189–232.
34. National Programme for Prevention and Control of Non-Communicable Diseases (NP-NCD) Operational Guidelines (Revised: 2023–30) Directorate General of Health Services Ministry of Health and Family welfare Government of India 2023. Available from: https://mohfw.gov.in/sites/default/files/NP-NCD%20Operational%20Guidelines.pdf. [Last accessed on 2024 May 28].
35. Indian Council of Medical Research, Public Health

Foundation of India and Institute for Health Metrics and Evaluation. India: Health of the Nation's States-The India State -level Disease Burden Initiative. 2017. Available from: https://www.healthdata.org/sites/default/files/files/policy_report/2017/India_Health_of_the_Nation%27s_States_Report_2017.pdf. [Last accessed on 2024 May 28].

36. Anjana RM, Unnikrishnan R, Deepa M, Venkatesan U, Pradeepa R, Joshi S, *et al*. Achievement of guideline recommended diabetes treatment targets and health habits in people with self-reported diabetes in India (ICMR-INDIAB-13): A national cross-sectional study. Lancet Diabetes Endocrinol 2022;10:430-41.