Analysis of factors influencing the prevalence of type 2 diabetes mellitus

Kunyan Gao

School of Information Management and Systems, Capital Medical University, China

yilan@ldy.edu.rs

Abstract. With the growth of the global economy and lifestyle changes, Type 2 Diabetes Mellitus (T2DM) has become a major global health challenge. It reduces patients' quality of life and leads to severe complications, increasing medical costs and psychological stress for patients and their families. Understanding the factors driving T2DM prevalence is essential for effective prevention and control. This study investigated the factors affecting the prevalence of T2DM using a comprehensive dataset from Kaggle, which contains health data from 1879 patients. The study used t-test, chi-square test, and logistic regression to analyze variables such as lifestyle, medical history, clinical measures, and symptoms. Significant factors associated with T2DM included family history, hypertension (HTN), smoking, urinary frequency, excessive thirst, unexplained weight loss, fasting blood sugar (FBS), and HbA1c levels. These factors have an impact on the prevalence of diabetes and are important in reducing the incidence of T2DM and associated complications if early intervention and personalized prevention strategies are adopted.

Keywords: diabetes risk factors; Logistic Regression; Type 2 Diabetes Mellitus (T2DM)

1. Introduction

With the development of the global economy and lifestyle changes, the prevalence of Type 2 Diabetes Mellitus (T2DM) become a major public health challenge worldwide. This problem affects all countries, but its prevalence and influencing factors vary by region and population. Developed regions such as North America and Europe have a higher prevalence of T2DM due to changing lifestyles and aging. Developing regions such as Asia and Africa, where the prevalence of T2DM has traditionally been low, are on the rise with rapid economic development and urbanization. According to the World Health Organization (WHO), the number of T2DM patients worldwide has continued to increase over the past decades and is expected to grow further in the future. The WHO estimates that diabetes will be the 7th primary cause of fatality by 2030.

The metabolic illness known as type 2 diabetes is typified by polydipsia, polyphagia, and hyperglycemia. One of the most prevalent metabolic illnesses, T2DM, is on the rise globally and is primarily brought on by insulin resistance and insufficient insulin production [1–3].

In addition to lowering a patient's quality of life, T2DM can cause several dangerous side effects, including neuropathy, retinopathy, renal disease, and cardiovascular disease. These complications not only increase the medical burden and economic costs but also bring great psychological and social

pressure to patients and their families. Therefore, to prevent and control the condition, a thorough understanding of the factors determining the occurrence of T2DM is crucial.

First, identifying risk factors for T2DM is crucial for early prevention. The risk factors for T2DM include genetic background, unhealthy lifestyle, environmental factors, and socioeconomic factors. By identifying these factors, early intervention can be implemented in high-risk groups to reduce their risk of developing the disease. For example, targeting lifestyle factors such as obesity and physical inactivity can help people control their weight and improve their metabolic health by promoting healthy diets and increasing physical activity. Early prevention can not only reduce the incidence of T2DM but also minimize the occurrence of related complications, thus reducing the burden of disease on individuals and society [4].

Second, studying the factors influencing the prevalence of T2DM can help identify and promote protective factors. Protective factors such as healthy eating habits, regular physical activity, and good mental health can significantly reduce the risk of T2DM. Public health policies can accordingly formulate corresponding health promotion measures, such as increasing community exercise facilities and conducting health education activities, to comprehensively enhance public health awareness and quality of life [5, 6].

Third, studies on the variables influencing the onset of T2DM can serve as a foundation for the creation of individualized preventative and therapeutic initiatives. Through in-depth analysis of these differences, personalized prevention and treatment plans can be developed for high-risk individuals to improve the effectiveness of intervention. For example, for individuals with a family record of the disease, genetic screening and early monitoring can be conducted to take preventive measures promptly; for individuals with unhealthy lifestyles, intervention can be carried out through customized diet and exercise programs. Personalized prevention and treatment programs can not only effectively reduce the incidence of T2DM, but also improve patients' quality of life and treatment outcomes [7, 8].

In addition, studying the factors influencing the prevalence of T2DM has important socioeconomic implications. T2DM and its complications place a huge burden on the healthcare system and socioeconomics. Effective prevention and control measures can significantly reduce the incidence of T2DM and complications, thereby reducing healthcare costs and increasing the productivity of the working population. A healthy working population is the basis for socioeconomic development, and effective control of T2DM not only reduces the pressure on the healthcare system, but also promotes sustainable socioeconomic development.

The study aims to gain insight into the pathogenesis of T2DM by systematically analyzing the impact of multiple factors, including lifestyle, medical history, substance use, environmental exposures, and health behaviors, on the development of T2DM. In addition, the study also aims to raise public awareness of T2DM and its risk factors and promote the adoption of a healthy lifestyle by publicizing the findings. Ultimately, these efforts will help promote further development in the fields of medicine and public health and enhance society's overall health.

2. Methods

The dataset used in this work is taken from the Kaggle website, which makes it freely accessible. The 1,879 patients in this dataset have extensive health data, including demographics, lifestyle characteristics, clinical measures, medical histories, prescription drug consumption, symptoms, quality of life ratings, exposure to the environment, and health-related behaviors. To ensure privacy and secrecy, each patient is assigned to a private physician.

This research employs SPSS26.0 software for data organization and analysis. The t-test is employed for measurement data that follows a normal distribution, the chi-square test is utilized for count data comparison, the logistic regression model is employed to examine the factors that influence the prevalence of type 2 diabetes, and a test level of $\alpha = 0.05$ was employed.

3. Results

Table 1 presents the study participants' baseline characteristics. Comprehensive health data from 1879

eligible participants are included in the dataset used in this paper to explore factors associated with diabetes.

Table 1. Baseline characteristics of included participants.

Variable	Case	Diabetes		Chi-Square	P-Value
		Yes(1)	No(0)		
Age				1.647	0.8
20-35	402(21.4)	167(41.5)	235(58.5)		
35-50	391(20.8)	161(41.2)	230(58.8)		
50-65	399(21.2)	151(37.8)	248(62.2)		
65-80	400(21.3)	162(40.5)	238(59.5)		
80-90	287(15.3)	111(38.7)	176(61.3)		
Gender				0.259	0.611
Male	963(51.3)	380(39.5)	583(60.5)		
Female	916(48.7)	372(40.6)	544(59.4)		
Ethnicity				2.332	0.506
Caucasian	1099(58.5)	444(40.4)	655(59.6)		
African American	357(19.0)	151(42.3)	206(57.7)		
Asian	206(11.0)	76(36.9)	130(63.1)		
Other	217(11.5)	81(37.3)	136(62.7)		
Socioeconomic Status				1.775	0.412
Low	557(29.6)	210(37.7)	347(62.3)		
Middle	780(41.5)	320(41.0)	460(59.0)		
High	542(28.8)	222(41.0)	320(59.0)		
Education Level				1.595	0.66
None	163(8.7)	60(36.8)	103(63.2)		
High School	615(32.7)	252(41.0)	363(59.0)		
Bachelor's	725(38.6)	296(40.8)	429(59.2)		
Higher	376(20.0)	144(38.3)	232(61.7)		
BMI				0.962	0.327
<25	730(38.9)	282(38.6)	448(61.4)		
≥25	1149(61.1)	470(40.9)	679(59.1)		
Smoking				5.445	0.02
No	1350(71.8)	518(38.4)	832(61.6)		
Yes	529(28.2)	234(44.2)	295(55.8)		
Alcohol Consumption				2.521	0.471
0-5	486(25.9)	203(41.8)	283(58.2)		
5-10	437(23.3)	172(39.4)	265(60.6)		
10-15	456(24.3)	170(37.3)	286(62.7)		
15-20	500(26.6)	207(41.4)	293(58.6)		
Physical Activity				0.629	0.428
0-5	891(47.4)	365(41.0)	526(59.0)		
5-10	988(52.6)	387(39.2)	601(60.8)		

A total of 1879 people are investigated, with an overall prevalence of 40.02%, 39.46% male and 40.61% female. The age is 20-90 years old, with an average age of (55.04±20.516) years. There are 963 males with a mean age of (55.06±20.247) years and 916 females with a mean age of (55.03±20.806) years. The results of the statistical analysis indicated that there is a significant difference ($p<0.05$) between the smoking status and the diagnosis of diabetes. The results are shown in Tables 1 and 2.

Table 2. Clinical measurements and lifestyle factors results.

Variable	Male		Female		t-Value	P-value
	Mean	Std.dev.	Mean	Std.dev.		
Age	55.06	20.247	55.03	20.806	0.037	0.970
BMI	27.54	7.049	27.84	7.338	-0.886	0.376
Systolic BP	134.95	25.635	133.11	25.572	1.556	0.12
Diastolic BP	90.13	17.542	89.58	17.105	0.693	0.488
Fasting Blood Sugar	135.84	37.438	134.54	37.606	0.747	0.455
HbA1c	6.91	1.746	7.05	1.730	-1.830	0.067
Serum Creatinine	2.81	1.307	2.76	1.309	0.899	0.369
BUN Levels	27.74	13.173	27.86	12.404	-0.194	0.846
Cholesterol Total	225.21	43.421	224.79	43.333	0.207	0.836
Cholesterol LDL	123.04	42.606	126.36	43.187	-1.676	0.094
Cholesterol HDL	59.71	23.132	60.44	23.516	-0.681	0.496
Cholesterol Triglycerides	226.61	98.793	228.21	103.460	-0.342	0.732
Alcohol Consumption	10.32	5.816	9.86	6.009	1.668	0.096
Physical Activity	5.18	2.811	5.22	2.811	-0.29	0.772
Diet Quality	4.89	2.925	4.91	2.807	-0.128	0.898
Sleep Quality	7.03	1.741	7.01	1.718	0.25	0.802

Clinical Measurements included systolic blood pressure (SystolicBP), diastolic blood pressure (DiastolicBP), fasting blood sugar levels (FBS), HbA1c levels, serum creatinine levels (SC), blood urea nitrogen levels (BUN), total cholesterol levels (Total-C), LDL cholesterol levels (LDL-C), HDL cholesterol levels (HDL-C), and triglycerides levels(T-C). Statistical results showed that the differences in clinical measurements are not statistically significant between males and females (all P>0.05). The results are shown in Table 2.

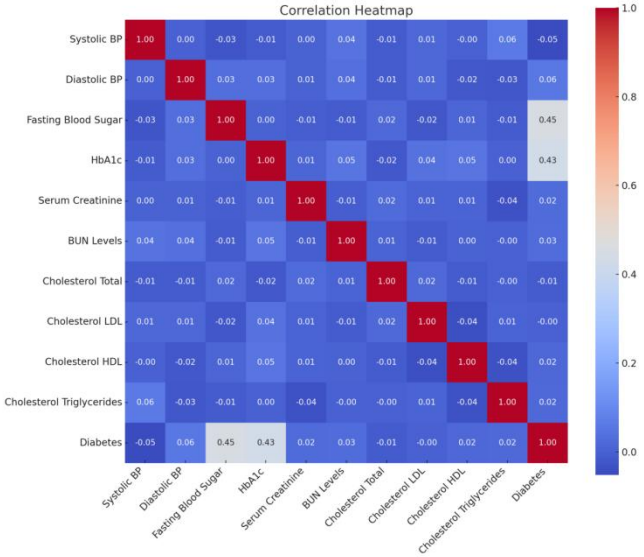

Figure 1. Heat map of clinically measured data and the prevalence of T2DM (Original).

Systolic and diastolic blood pressure, as well as FBS and HbA1c values, have significance levels of less than 0.05 and less than 0.01, respectively. At the 0.05 level, there is a significant positive correlation (r=0.055, p<0.05) between DiastolicBP and the prevalence of T2DM, and a significant negative correlation (r=-0.052, p<0.05) between SystolicBP and the prevalence of T2DM. At the 0.01

level, there is a significant positive correlation (r=0.452, p<0.01) between the FBS level and the prevalence of T2DM, and at the 0.01 level, there is a significant positive correlation (r=0.431, p<0.01) between HbA1c and the prevalence of T2DM.

It indicates that SystolicBP, DiastolicBP, FBS, and HbA1c level affect the prevalence of T2DM. Meanwhile, other clinical measures are not significantly correlated with the prevalence of T2DM (p>0.05). The results are shown in Figure 1.

Table 3. One-way analysis of variance.

Variable	Classification	Diabetes		Chi-Square	P-Value
		Yes(1)	No(0)		
Antihypertensive Medications	No	546	796	0.863	0.353
	Yes	206	331		
Statins	No	458	662	0.877	0.349
	Yes	294	465		
Antidiabetic Medications	No	538	818	0.243	0.622
	Yes	214	309		
Frequent Urination	No	548	960	43.13	0.000
	Yes	204	167		
Excessive Thirst	No	580	936	10.158	0.001
	Yes	172	191		
Unexplained WeightLoss	No	652	1021	7.001	0.008
	Yes	100	106		
Blurred Vision	No	670	1030	2.762	0.097
	Yes	82	97		
Slow Healing Sores	No	673	1013	0.074	0.785
	Yes	79	114		
Tingling Hands Feet	No	665	1005	0.253	0.615
	Yes	87	122		
Heavy Metals Exposure	No	710	1071	0.346	0.556
	Yes	42	56		
Occupational Exposure Chemicals	No	676	1009	0.065	0.800
	Yes	76	118		
Water Quality	Good	589	913	2.031	0.154
	Bad	163	214		
Family History Diabetes	No	554	877	4.272	0.039
	Yes	198	250		
Gestational Diabetes	No	669	1023	1.647	0.199
	Yes	83	104		
Polycystic Ovary Syndrome	No	711	1084	2.829	0.093
	Yes	41	43		
Previous Pre Diabetes	No	640	950	0.228	0.633
	Yes	112	177		
Hypertension	No	593	998	32.682	0.000
	Yes	159	129		

The results of the one-way analysis demonstrated the significant relationships between the prevalence of T2DM and the following factors: smoking, family history of T2DM, HTN, DiastolicBP, SystolicBP, FBS, HbA1c, Frequent Urination, Excessive Thirst, and Unexplained Weight Loss (Table 3). These relationships allowed regression analyses to be conducted. The VIF values suggest that there is no severe multicollinearity among the predictors.

Table 4. Logistic regression analysis.

Variable	B	Standard Error	Wald	OR(95%CI)	P-Value
Family History Diabetes					
No				1	
Yes	0.216	0.111	3.828	1.241(1.000-1.542)	0.050
Hypertension					
No				1	
Yes	0.726	0.129	31.398	2.066(1.603-2.663)	0.000
Smoking					
No				1	
Yes	0.242	0.104	5.434	1.274(1.039-1.562)	0.020
Frequent Urination					
No				1	
Yes	0.756	0.118	41.106	2.129(1.690-2.682)	0.000
Excessive Thirst					
No				1	
Yes	0.377	0.119	10.020	1.459(1.155-1.843)	0.002
Unexplained Weight Loss					
No				1	
Yes	0.370	0.150	6.064	1.448(1.079-1.945)	0.014
SystolicBP	-0.004	0.002	3.004	0.996(0.991-1.001)	0.083
DiastolicBP	0.005	0.004	1.877	1.005(0.998-1.012)	0.171
FastingBloodSugar	0.039	0.002	356.168	1.040(1.036-1.044)	0.000
HbA1c	0.807	0.044	341.033	2.242(2.058-2.443)	0.000

Based on logistic statistics, the risk of developing diabetes is 1.241 times higher in people with a family history of T2DM, with a marginally significant difference ($P = 0.050$). HTN individuals are 2.066 times more likely to develop diabetes ($P < 0.001$), and smokers are 1.274 times more likely ($P = 0.020$). Among symptom indicators, those experiencing frequent urination are 2.129 times more likely to develop diabetes ($P < 0.001$), while excessive thirst increases the likelihood by 1.459 times ($P = 0.002$), and unexplained weight loss by 1.448 times ($P = 0.014$). Each unit increase in FBS raises the diabetes risk by 1.040 times ($P < 0.001$), highlighting its importance in diabetes diagnosis. Additionally, a one-unit increase in HbA1c raises the risk by 2.242 times ($P < 0.001$). There is no significant relationship between systolic and diastolic blood pressure and diabetes, indicating these factors may influence diabetes risk without reaching statistical significance. The results are shown in Table 4.

4. Discussion

Among the variables discussed in this article, family history, HTN, smoking, frequent urination, excessive thirst symptoms, unexplained weight loss symptoms, FBS, and HbA1c are the factors most strongly associated with T2DM. Understanding the impact of these factors can help in early recognition and prevention of diabetes.

Family history is the result of a combination of genetics and lifestyle, which suggests that genetic factors play an important role in diabetes risk [9]. HTN is strongly associated with insulin resistance, which may be due to endothelial dysfunction and inflammatory response [10]. Smoking may lead to insulin resistance and impairment of pancreatic beta cell function, which increases the risk of diabetes [11, 12]. Frequent urination is a direct symptom of hyperglycemia because high blood sugar increases the burden on the kidneys, leading to increased urination. Excessive thirst is also a direct symptom of hyperglycemia and is the body's attempt to dilute the sugar in the blood by increasing the amount of water consumed. Unexplained weight loss may be due to increased breakdown of fat and muscle due

to diabetes, as the body is unable to utilize blood sugar efficiently [13, 14]. HbA1c is an indicator of long-term glycemic control and significantly elevated HbA1c indicates persistent high blood glucose levels and increased risk of diabetes [15].

5. Conclusion

This study identifies key factors that are strongly associated with T2DM, including family history, HTN, smoking, and symptoms such as frequent urination, excessive thirst, and unexplained weight loss. Additionally, FBS and HbA1c levels are significant indicators of diabetes risk. Understanding the impact of these factors is crucial for the early detection and prevention of diabetes. By highlighting these associations, the study provides valuable insights into the factors influencing the prevalence of diabetes, serving as a foundation for future research and practical applications. It is hoped that these findings will contribute to advancements in the field and lead to more meaningful discussions about diabetes prevention and management.

The strength of this study, is the comprehensive dataset from Kaggle, which includes detailed health data on 1,879 patients, covering demographics, lifestyle factors, medical history, clinical measurements, and more. This allowed it possible to conduct a comprehensive study using a variety of statistical techniques, including logistic regression models, chi-square tests, and t-tests. This helped to identify the factors that influence the occurrence of T2DM and provided a solid scientific foundation for individualized prevention and treatment plans.

This paper has some strengths but also limitations. The potential geographical or population biases in the data, self-reporting biases for some variables, and the constraints of statistical methods may have overlooked confounders or errors. Additionally, as a cross-sectional study, it cannot establish causal relationships, and the sample size may not represent diverse cultural backgrounds, limiting the generalizability of the findings.

References

[1] World Health Organization (WHO) 2017 Global Report on Diabetes (Geneva: WHO) Available online: http://www.who.int/diabetes/global-report/en/ (accessed on 22 September 2018).

[2] Mathers C D and Loncar D 2006 Projections of global mortality and burden of disease from 2002 to 2030 PLoS Med. 3 e442

[3] World Health Organization and International Diabetes Federation 2006 Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia: Report of a WHO/IDF Consultation (Geneva: WHO)

[4] Bellou V, Belbasis L, Tzoulaki I and Evangelou E 2018 Risk factors for type 2 diabetes mellitus: an exposure-wide umbrella review of meta-analyses PloS One 13 e0194127

[5] Zeru M A, Tesfa E, Mitiku A A, Seyoum A and Bokoro T A 2021 Prevalence and risk factors of type-2 diabetes mellitus in Ethiopia: systematic review and meta-analysis Sci. Rep. 11 21733

[6] Ismail L, Materwala H and Al Kaabi J 2021 Association of risk factors with type 2 diabetes: A systematic review Comput. Struct. Biotechnol. J. 19 1759

[7] Kyrou I, Tsigos C, Mavrogianni C, Cardon G, Van Stappen V, Latomme J et al. 2020 Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe BMC Endocr. Disord. 20 1

[8] Diab M, Barhoosh H A, Daoudi B, AlMukdad S I, Zaghloul N H, Ashour M et al. 2019 Prevention and screening recommendations in type 2 diabetes: review and critical appraisal of clinical practice guidelines Prim. Care Diabetes 13 197

[9] Lascar N, Brown J, Pattison H, Barnett A H, Bailey C J and Bellary S 2018 Type 2 diabetes in adolescents and young adults. Lancet Diabetes Endocrinol. 6 69

[10] Yildiz M, Esenboğa K and Oktay A A 2020 Hypertension and diabetes mellitus: highlights of a complex relationship Curr. Opin. Cardiol. 35 397

[11] Durlach V, Vergès B, Al-Salameh A, Bahougne T, Benzerouk F, Berlin I et al. 2022 Smoking and diabetes interplay: a comprehensive review and joint statement Diabetes Metab. 48 101370
[12] Reed Z E, Sallis H M, Richmond R C, Attwood A S, Lawlor D A and Munafo M R 2024 Do smoking and alcohol behaviours influence risk of type 2 diabetes? A Mendelian randomisation study medRxiv 2024-07
[13] Dwivedi M and Pandey A R 2020 Diabetes mellitus and its treatment: an overview J. Adv. Pharmacol. 1 48
[14] Gregory N S 2023 Excessive thirst, hunger, and urination in diabetes Introduction to Clinical Pharmacology: From Symptoms to Treatment (New York: Elsevier) p 304
[15] Selvi N M K, Nandhini S, Sakthivadivel V, Lokesh S, Srinivasan A R and Sumathi S 2021 Association of triglyceride – glucose index (TyG index) with hbA1c and insulin resistance in type 2 diabetes mellitus Maedica 16 375