Journal of Clinical and Nursing Research, 2024, Volume 8, Issue 7
http://ojs.bbwpublisher.com/index.php/JCNR
Online ISSN: 2208-3693
Print ISSN: 2208-3685

Study on Key Biological Indicators of Diabetes Based on Statistical Tests

Shuaibin Yang*

School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu Province, China.

Corresponding author: Shuaibin Yang, yangshuaibin@nuist.edu.cn

Abstract: Normality testing is a fundamental hypothesis test in the statistical analysis of key biological indicators of diabetes. If this assumption is violated, it may cause the test results to deviate from the true value, leading to incorrect inferences and conclusions, and ultimately affecting the validity and accuracy of statistical inferences. Considering this, the study designs a unified analysis scheme for different data types based on parametric statistical test methods and non-parametric test methods. The data were grouped according to sample type and divided into discrete data and continuous data. To account for differences among subgroups, the conventional chi-squared test was used for discrete data. The normal distribution is the basis of many statistical methods; if the data does not follow a normal distribution, many statistical methods will fail or produce incorrect results. Therefore, before data analysis and modeling, the data were divided into normal and non-normal groups through normality testing. For normally distributed data, parametric statistical methods were used to judge the differences between groups. For non-normal data, non-parametric tests were employed to improve the accuracy of the analysis. Statistically significant indicators were retained according to the significance index *P*-value of the statistical test or corresponding statistics. These indicators were then combined with relevant medical background to further explore the etiology leading to the occurrence or transformation of diabetes status.

Keywords: Diabetes diagnosis; Statistical test; Nonparametric statistics; Normality test

Online publication: August 8, 2024

1. Introduction

In recent years, diabetes mellitus has become the most common pattern of illness worldwide and poses a significant public health challenge. The scientific problems within this domain have attracted extensive attention from biomedical and applied statistical researchers, leading to interdisciplinary studies involving artificial intelligence, data science, and medical informatics. The early diagnosis of diabetes is critical for prevention and effective treatment. Diabetes mellitus (DM) is characterized by persistent hyperglycemia and includes several types such as Type 1 (T1DM), Type 2 (T2DM), gestational (GDM), and other less common forms [1]. T1DM typically results from the autoimmune-induced destruction of islet cells [2], while T2DM is primarily characterized by insulin resistance (IR) and β-cell dysfunction [3]. Despite significant medical advancements, a complete cure for diabetes has yet to be realized. Many risk factors associated with diabetes have been identified, yet its pathogenesis remains not fully elucidated [4,5].

According to the World Health Organization (WHO), T2DM accounts for over 90% of all diabetes cases [6]. Yang *et al.* [7] conducted a large-scale study on the prevalence of T2DM among adults in China, revealing a prevalence rate of 10.9%, compared to a 0.09% rate for T1DM. Patterson *et al.* [8] studied the incidence of T1DM in European children from 1989 to 2003 and predicted cases for 2005 to 2020, noting an increasing trend among European children, yet significantly lower compared to T2DM. These studies collectively indicate a considerably higher prevalence rate for T2DM than other diabetes types.

Diabetes patients face higher morbidity and mortality risks than the general population. The global prevalence of adult diabetes has been climbing for decades, from an estimated 30 million in 1964 [9] to an alarming 171 million in less than 40 years [10], with projections suggesting a rise to 642 million by 2040 [11]. Diabetes has become a global public health issue, with its progression from asymptomatic early stages to critical onset, accompanied by complex physiological and metabolic changes. Thus, identifying high-risk individuals and implementing early interventions is critical for diabetes prevention and control.

In this study, a multidimensional analysis strategy is used to analyze the normality of the continuous data in the diabetes data set. This analysis is critical to understanding the distribution characteristics of biomarkers in patients with diabetes, as many biostatistical methods rely on the normality hypothesis. This step allows researchers to assess whether the data is suitable for further statistical analysis of the parameters. However, due to the complexity of medical data, not all data follow a normal distribution. Therefore, the study also incorporates non-parametric statistical methods that do not rely on the distribution of data assumptions, thus providing the possibility to accurately extract key biological indicators of the progression of diabetic patients from a "free" to a "diseased" state. The identification of these indicators contributes to an in-depth understanding of the pathogenesis of diabetes and provides key information for early diagnosis and intervention.

The structure of this paper is organized as follows: **Section 2** presents some research on diagnostic factors of diabetes. **Section 3** covers the details of data presentation and experimental procedure. **Section 4** summarizes the main findings and contributions of this study and looks forward to future research.

2. Related works

The diagnosis of diabetes is influenced by a variety of factors. Yokoyama *et al.* [12] pointed out in their 2003 study that the pathogenesis of diabetic nephropathy may be related to atherosclerosis, and further studies are needed to clarify its practicability and predictive value. In 2015, Lorenz *et al.* [13] used carotid intima-media thickness (CIMT), which predicts cardiovascular disease (CVD) events in the general population, to determine the relationship between CIMT changes and CVD events in diabetic patients. The results do not support the use of CIMT progression as an alternative endpoint for clinical trials in patients with diabetes. Therefore, more studies are still needed to analyze the key biological indicators of diabetes from multiple dimensions. In 2020, Katsiki *et al.* [14] reviewed the research on the correlation between diabetes and carotid artery disease, concluding that further research is needed in this field to optimize the treatment of such patients. Wong *et al.* [15] evaluated the health preference scores of Chinese people with DM according to different socio-demographic characteristics of the Hong Kong region population. They found that complications caused by diabetes can have a serious impact on people's quality of life and that modifying individual disease guidelines may help improve their health-related quality of life.

Diabetes is a complex disease influenced by genetic and environmental risks, and numerous studies have shown the role of microbial dysbiosis in its pathogenesis. Studies comparing microbiome characteristics between type 2 diabetes and healthy controls have reported inconsistent results. For further research, Que *et al.* [16] conducted a study on the characteristics of intestinal bacteria in type 2 diabetes patients in 2021 and observed significant differences in intestinal bacteria data between the experimental group and the control group, providing new ideas for the prevention and improvement of diabetes. Li *et al.* [17] evaluated the role of ncRNAs in the development of DM and CVD, as well as the evidence of ncRNAs as potential therapeutic targets, and used bioinformatics to analyze the differential ncRNAs with potential functions in DM, providing a new direction for diabetes research. Guo *et al.* [18] believed that dyslipidemia and abnormal glucose metabolism in T2DM patients were closely related to carotid IMT. Clinically, T2DM patients can be monitored for dyslipidemia, atherosclerosis, and other related complications by assessing CIMT.

While it is important to use biological indicators to improve the management of T2DM patients, Diriba *et al.* [19] pointed out that encouraging self-management practices, as well as ongoing family or friend support, are necessary to improve the quality of life of patients with type 2 diabetes. In 2023, Liao *et al.* [20] revealed the relationship between carotid artery ultrasound parameters and microvascular and macrovascular complications of diabetes. They suggested that carotid artery ultrasound parameters may be a non-invasive method for the early detection of long-term complications of diabetes.

3. Research method

3.1. Dataset introduction

The dataset was obtained retrospectively from a study conducted by the Affiliated Hospital of Nanjing University of Chinese Medicine between May 2017 and April 2021. The dataset comprises 283 subjects, including 104 patients and 179 individuals from the normal population. This data comes from the same database as the atherosclerosis study conducted by Yang *et al.* [21].

Table 1 provides a detailed description of the dataset.

Table 1. Dataset description

	Number
Number of patients	283
Risk group	104
Control group	179

3.2. Analytical process

At this stage, this study analyzed the pathogenesis of DM by statistical analysis methods. There were 23 independent variables in total, including 4 discrete variables: gender (0 male, 1 female), hypertension (HTN; 0 negative, 1 positive), atherosclerosis (AS; 0 negative, 1 positive), and smoking (0 non-smoking, 1 smoking). There were 19 continuous variables, which included biological indicators such as body mass index (BMI), age, and low-density lipoprotein (LDL), as well as ultrasound imaging indicators such as CIMT and carotid artery ultrafast pulse wave velocity (UFPWV). The dependent variable was diabetes (0 normal, 1 diseased).

3.2.1. Discrete variable analysis

To explore the important characteristics of the critical state of diabetes, a chi-squared test was performed on discrete variables to study the impact of diabetic complications on diabetic patients. The chi-squared test is a commonly used statistical method to compare whether the difference between the observed frequency and the expected frequency is significant. It can be used to assess the impact of different discrete data on diabetes and determine whether the impact is statistically significant. The results are shown in **Table 2**.

Table 2. Chi-squared test of discrete variables and diabetes mellitus

Chi-squared test	Gender - DM	HTN - DM	AS - DM	Smoking - DM
P-value	0.548	0.000	0.009	0.055

In the results, $P < 0.05$ was obtained for HTN and AS, indicating that there is a certain deviation between the observed data and the expected data. This means that the occurrence of complications may have a different impact on the development of diabetes, and not just be the result of random factors. It can be concluded that HTN and AS have a significant impact on the occurrence of diabetes. This is of great significance for the clinical treatment of diabetes, as it can help doctors plan treatment strategies in advance and provide reliable data support for decision-making.

3.2.2. Continuous variable analysis

To categorize continuous variables into normal and non-normal groups, the Kolmogorov-Smirnov (KS) test was used to test whether the data conform to a normal distribution. In the KS normality test, the difference between the cumulative distribution function of the sample data and the cumulative distribution function of the theoretical normal distribution is compared. This method is particularly useful for normality tests that are not limited by sample size. The statistic and the corresponding *P*-value are calculated, and if the *P*-value is less than the significance level (usually 0.05), the hypothesis that the data conforms to a normal distribution is rejected. Through the KS test, the data can be more objectively evaluated for normal distribution, providing an important reference for subsequent data analysis. The results are shown in **Table 3**.

Table 3. Kolmogorov-Smirnov (KS) test for continuous variables

KS test	BMI	Age	LDL	TG	Cr	Urea	UA
P	0.48	0.033	0.228	0.000	0.000	0.000	0.006
KS test	**Hb**	**WBC**	**RBC**	**PLT**	**ALT**	**AST**	**LCCA-IMT**
P	0.001	0.007	0.002	0.002	0.000	0.000	0.000
KS test	**RCCA-IMT**	**LCCA-RI**	**RCCA-RI**	**LCCA-BS**	**LCCA-ES**		
P	0.000	0.031	0.019	0.6092	0.7841		

Abbreviation: KS test, Kolmogorov-Smirnov test; BMI, body mass index; LDL, low-density lipoprotein; TG, triglycerides; Cr, creatinine; UA, uric acid; Hb, hemoglobin; WBC, white blood cell; RBC, red blood cell; PLT, platelet; ALT, alanine aminotransferase; AST, aspartate aminotransferase; LCCA-IMT, left common carotid artery-intima media thickness; RCCA-IMT, right common carotid artery-intima media thickness; LCCA-RI, left common carotid artery-resistive index; RCCA-RI, right common carotid artery-resistive index; LCCA-BS, left common carotid artery-beginning of systole; LCCA-ES, left common carotid artery-end of systole.

According to the results of the normality test, if the data conforms to a normal distribution, statistical methods based on normal distribution can be used in subsequent analysis. If the data do not conform to a normal distribution, non-parametric statistical methods should be considered.

From the above analysis, it was found that there were 15 instances of $P < 0.05$, indicating sufficient reasons to reject the hypothesis that the data conform to a normal distribution. The remaining 4 variables are BMI, LDL, LCCA-BS, and LCCA-ES. This result signifies the need to consider the non-normal distribution of the data in subsequent analysis and modeling processes. In medical data analysis, the normality of data is typically a precondition for statistical inference and hypothesis testing. Non-parametric statistical methods may be required to ensure the reliability and accuracy of the analysis results.

For the four normal variables, variance analysis was used for statistical testing to compare the mean differences between the two groups. The results are shown in **Table 4**.

Table 4. Analysis of variance (ANOVA) of normal data

ANOVA	BMI - DM	LDL - DM	LCCA-BS - DM	LCCA-ES - DM
P	0.037	0.015	0.000	0.000

From these results, BMI, LDL, LCCA-BS, and LCCA-ES indicators were significantly different among different groups of diabetic patients. For variables that do not follow a normal distribution, the rank sum test was used to examine the difference between each variable and the diabetic population, with the results shown in **Table 5**.

Table 5. Nonparametric rank test

Rank test	Age - DM	TG - DM	Cr - DM	Urea - DM
P	0.049	0.000	0.689	0.687
Rank test	**UA - DM**	**Hb - DM**	**WBC - DM**	**RBC - DM**
P	0.000	0.150	0.612	0.002
Rank test	**PLT - DM**	**ALT - DM**	**AST - DM**	**LCCA-IMT - DM**
P	0.081	0.847	0.409	0.079
Rank test	**RCCA-IMT - DM**	**LCCA-RI - DM**	**RCCA-RI - DM**	
P	0.338	0.18	0.412	

According to the rank sum test results, the *P*-values of age, triglyceride (TG), urea, uric acid (UA), and red blood cell count (RBC) were all less than 0.05 for comparisons among diabetic patients (DM), indicating significant differences in these indicators among different diabetic patients. These findings are significant for understanding the differences in these physiological indicators among groups of diabetic patients and provide useful references for personalized treatment and management.

4. Conclusion

When assessing the risk and disease status of diabetic patients, traditional methods often rely on parametric statistical tests, such as the t-test and analysis of variance. Although these methods can provide some degree of information, they often ignore the nonlinear relationships and interactions between variables. To more fully assess the risk of patients with diabetes, non-parametric statistical tests, such as the rank sum test, can be introduced. These methods do not rely on distribution assumptions of the data and can better handle nonlinear relationships and outliers, thus improving the accuracy of the evaluation.

A comprehensive approach that combines parametric and non-parametric statistical tests can provide a more in-depth and holistic analysis of diabetes research. For instance, combining the results of both parametric and non-parametric tests allows consideration of both linear and nonlinear relationships, enabling a more accurate assessment of the risk and condition status of diabetic patients. In clinical practice, these evaluation methods can provide doctors with more scientific support, helping them better understand the condition, develop personalized treatment plans, and predict the development trend of the patient's disease.

Ethics approval

The study was approved by the decision of the Ethical Committee of the Affiliated Hospital of Nanjing University of Chinese Medicine where it was conducted. The study is an open controlled clinical trial.

Consent to participate

A written contract was signed with each of the patients, in which they gave their consent to information processing; the contract guaranteed non-disclosure, complete anonymity of information, and observance of moral and ethical norms.

Funding

(1) National Natural Science Foundation of China (No. 12271261)

(2) Postgraduate Research and Practice Innovation Program of Jiangsu Province, China (Grant No. SJCX23 0368)

Disclosure statement

The author declares no conflict of interest.

References

[1] Redondo MJ, Hagopian WA, Oram R, et al., 2020, The Clinical Consequences of Heterogeneity Within and Between Different Diabetes Types. Diabetologia, 63(10): 2040–2048. https://doi.org/10.1007/s00125-020-05211-7

[2] Eizirik DL, Pasquali L, Cnop M, 2020, Pancreatic β-Cells in Type 1 and Type 2 Diabetes Mellitus: Different Pathways to Failure. Nat Rev Endocrinol, 16(7): 349–362. https://doi.org/10.1038/s41574-020-0355-7

[3] Galicia-Garcia U, Benito-Vicente A, Jebari S, et al., 2020, Pathophysiology of Type 2 Diabetes Mellitus. Int J Mol Sci, 21(17): 6275. https://doi.org/10.3390/ijms21176275

[4] Siqueira ISL, Alves Guimarães R, Mamed SN, et al., 2020, Prevalence and Risk Factors for Self-Report Diabetes Mellitus: A Population-Based Study. Int J Environ Res Public Health, 17(18): 6497. https://doi.org/10.3390/ijerph17186497

[5] Zeru MA, Tesfa E, Mitiku AA, et al., 2021, Prevalence and Risk Factors of Type-2 Diabetes Mellitus in Ethiopia: Systematic Review and Meta-Analysis. Sci Rep, 11(1): 21733. https://doi.org/10.1038/s41598-021-01256-9

[6] World Health Organization (WHO), Diabetes, viewed June 20, 2023, https://www.who.int/news-room/fact-sheets/detail/diabetes

[7] Yang W, Lu J, Weng J, et al., 2010, Prevalence of Diabetes among Men and Women in China. N Engl J Med, 362(12): 1090–1101. https://doi.org/10.1056/NEJMoa0908292

[8] Patterson CC, Dahlquist GG, Gyürüs E, et al., 2009, Incidence Trends for Childhood Type 1 Diabetes in Europe during 1989–2003 and Predicted New Cases 2005–20: A Multicentre Prospective Registration Study. Lancet, 373(9680): 2027–2033. https://doi.org/10.1016/S0140-6736(09)60568-7

[9] Entmacher PS, Marks HH, 1965, Diabetes in 1964; A World Survey. Diabetes, 14: 212–223. https://doi.org/10.2337/diab.14.4.212

[10] Wild S, Roglic G, Green A, et al., 2004, Global Prevalence of Diabetes: Estimates for the Year 2000 and Projections for 2030. Diabetes Care, 27(5): 1047–1053. https://doi.org/10.2337/diacare.27.5.1047

[11] Ogurtsova K, da Rocha Fernandes JD, Huang Y, et al., 2017, IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract, 128: 40–50. https://doi.org/10.1016/j.diabres.2017.03.024

[12] Yokoyama H, Hirasawa K, Aoki T, et al., 2003, Brachial-Ankle Pulse Wave Velocity Measured Automatically by Oscillometric Method is Elevated in Diabetic Patients with Incipient Nephropathy. Diabet Med, 20(11): 942–945. https://doi.org/10.1046/j.1464-5491.2003.01043.x

[13] Lorenz MW, Price JF, Robertson C, et al., 2015, Carotid Intima-Media Thickness Progression and Risk of Vascular Events in People with Diabetes: Results from the PROG-IMT Collaboration. Diabetes Care, 38(10): 1921–1929. https://doi.org/10.2337/dc14-2732

[14] Katsiki N, Mikhailidis DP, 2020, Diabetes and Carotid Artery Disease: A Narrative Review. Ann Transl Med, 8(19): 1280. https://doi.org/10.21037/atm.2019.12.153

[15] Wong ELY, Xu RH, Cheung AWL, 2020, Measurement of Health-Related Quality of Life in Patients with Diabetes Mellitus Using EQ-5D-5L in Hong Kong, China. Qual Life Res, 29(7): 1913–1921. https://doi.org/10.1007/s11136-020-02462-0

[16] Que Y, Cao M, He J, et al., 2021, Gut Bacterial Characteristics of Patients With Type 2 Diabetes Mellitus and the Application Potential. Front Immunol, 12: 722206. https://doi.org/10.3389/fimmu.2021.722206

[17] Li C, Wang D, Jiang Z, et al., 2022, Non-Coding RNAs in Diabetes Mellitus and Diabetic Cardiovascular Disease. Front Endocrinol (Lausanne), 13: 961802. https://doi.org/10.3389/fendo.2022.961802

[18] Guo HJ, Li CC, Bian XY, et al., 2023, Correlation Study on the Relationship between Dyslipidemia and Carotid Intima-Media Thickness in Patients with Diabetes Mellitus. Pak J Med Sci, 39(3): 875–879. https://doi.org/10.12669/pjms.39.3.6866

[19] Diriba DC, Leung DYP, Suen LKP, 2023, Factors Predicted Quality of Life of People with Type 2 Diabetes in Western Ethiopia. PLoS One, 18(2): e0281716. https://doi.org/10.1371/journal.pone.0281716

[20] Liao M, Chen S, Guo R, 2023, Association between Carotid Ultrasonographic Parameters and Microvascular and Macrovascular Complications in Diabetes: A Systematic Review and Meta-Analysis. J Diabetes Complications, 37(8): 108554. https://doi.org/10.1016/j.jdiacomp.2023.108554

[21] Yang M, He L, Liu W, et al., 2024, Performance Improvement of Atherosclerosis Risk Assessment Based on Feature Interaction. Comput Methods Programs Biomed, 249: 108139. https://doi.org/10.1016/j.cmpb.2024.108139

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.