Proceedings of ICBioMed 2024 Workshop: Workshop on Intelligent Medical Data Analysis for Precision Medicine
DOI: 10.54254/2753-8818/50/2024AU0145

A comprehensive analysis of diabetes risk factors: Effects of gender, region, and lifestyle

Lidong Zhu

School of Arts and Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, 08901, US

lz460@scarletmail.rutgers.edu

Abstract. The prevention and nursing of diabetes is one of the research hotspots in the medical field. Diabetes is a prevalent illness that can seriously harm a patient's health if prompt, efficient prevention and treatment are not received. Studying diabetes nursing and prevention is crucial as a result. Many theories about the causative elements of diabetes are held by the medical community. For instance, obesity has a major role in the onset of diabetes. This paper will conduct a detailed statistical analysis of publicly available diabetes datasets, descriptive statistics, chi-square tests, and a series of regression analyses for factors such as total cholesterol, high-density lipoprotein, blood pressure, and gender to determine whether these are key predictors of diabetes. Diabetes was found to be substantially correlated with obesity, high cholesterol, and the high density lipoprotein (HDL) values. The study's findings offer a crucial statistical foundation for the treatment and avoidance of diabetes.

Keywords: Diabetes prevention, Obesity, Total cholesterol, High-density lipoprotein.

1. Introduction

Diabetes is a dangerous chronic illness that is becoming more commonplace globally every year. Type 2 diabetes affects an estimated 26 million people in the US, and many of them simultaneously struggle with obesity and cardiovascular disease [1]. Globally, diabetes is predicted to increase in frequency, affecting the social, medical, and economic domains. By 2045, there will likely be 629 million cases worldwide, up from 425 million in 2017 [2]. It is imperative to find quick fixes to halt or even reverse this tendency, particularly when it comes to investing in modifiable factors such as weight, physical activity, and diet [3]. As an example, bariatric surgery has proven to be an effective treatment approach. Research from randomized controlled trials shows that people with obesity and newly diagnosed type 2 diabetes who have weight-loss surgery have a 60–80% diabetes remission rate [4]. Enhancing insulin sensitivity and providing anti-inflammatory, antioxidant, and anti-cholesterol qualities that are critical for managing and preventing type II diabetes have also been associated with regular use of functional foods [5]. Remarkably, there may be a greater risk of diabetes in the years immediately following smoking cessation, underscoring the intricate relationship between risk factors for diabetes and lifestyle choices [6]. Recent studies have shown that there are gender disparities in diabetes mellitus; women with the condition often have higher body mass indices (BMIs) and have been afflicted for longer than males. This appears to be connected to the significant effects that sex hormones have on vascular function, inflammation responses, body composition, and energy expenditure [7]. The region may also

DOI: 10.54254/2753-8818/50/2024AU0145

be a potential factor in diabetes. Low socioeconomic position is frequently disregarded, although it can dramatically raise the risk for T2DM and prediabetes [8]. Findings describe an "elderly obese white female with hypertension living in a metropolitan area" as a typical person with managed diabetes [9]. Understanding the risk factors and prevalence of diabetes is essential for public health interventions.

The study conducted an in-depth analysis of a diabetes dataset containing multiple health indicators. The Chi-square test, logistic regression analysis, and descriptive statistics were among the study approaches used. In addition to identifying important diabetes predictors, the research attempts to highlight variations in health indicators by gender and geography [10]. Focus on the role of tobacco use, cholesterol levels, and other health indicators. By understanding these relationships, effective strategies can be developed to prevent and manage diabetes.

2. Data and methods

2.1. Data description

Several research scalars, including blood pressure, body weight, high-density lipoprotein (HDL), cholesterol, and glucose levels, are included in the dataset. The data are sourced from PubMed, specifically from the study titled "The Framingham Heart Study: 34-year follow-up of coronary heart disease in a male cohort" (PubMed ID: 9258308). Gender and region classifications are included, so differences can be compared for gender and different regions. First, the data set is subjected to descriptive statistical analysis, and each variable's mean, mode, and standard deviation are determined.

2.2. Descriptive statistical analysis

To support comparative analysis and describe the distribution of relevant diseases or health status, relevant characteristics, and exposure factors in the population, descriptive statistical analysis generally refers to the use of routine test records or data obtained through special investigations, depending on different regions, different times, and different population characteristics. The mode, mean, and sample standard deviation for the total sample, by gender, and by area are determined in this study using Excel functions AVERAGE, MODE, and STDEV.S, respectively. See Tables 1 and 2 for details. Using these descriptive statistics, the research was able to gain an initial understanding of the differences in key health indicators between the sexes and between the two regions.

Table 1. Descriptive statistics by gender

Female	cholesterol	glucose levels	high-density lipoprotein	ratio	glyhb	age	height	weight	blood pressure	waist	hip
Mean	208.44	102.65	52.11	4.36	5.49	45.83	63.73	174.49	135.96	38.12	44.35
Mode	179	81	46	3.90	4.40	40	63	165	130	40	43
Sample standard deviation	43.75	43.63	17.26	1.71	2.13	16.60	2.85	41.24	24.63	6.00	5.99
Male	**cholesterol**	**glucose levels**	**high-density lipoprotein**	**ratio**	**glyhb**	**age**	**height**	**weight**	**blood pressure**	**waist**	**hip**
Mean	207.02	112.24	48.13	4.75	5.72	48.26	69.12	181.92	137.21	37.59	41.23
Mode	179	85	36	3.60	4.67	40	63	170	130	37	39
Sample standard deviation	44.53	52.86	17.28	1.73	2.22	16.36	3.93	40.26	22.21	5.68	5.59

DOI: 10.54254/2753-8818/50/2024AU0145

Table 2. Descriptive statistics by region

Buckingham	cholesterol	glucose levels	high-density lipoprotein	ratio	glyhb	age	height	weight	blood pressure	waist	hip
Mean	205.08	106.41	50.11	4.55	5.68	47.08	66.50	178.71	136.04	38.14	42.71
Mode	179	92	34	4.90	4.44	31	63	170	140	33	41
Sample standard deviation	44.73	50.54	17.54	1.89	2.22	16.75	4.01	43.77	20.82	6.34	6.02
Louisa	cholesterol	glucose levels	high-density lipoprotein	ratio	glyhb	age	height	weight	blood pressure	waist	hip
Mean	210.56	106.94	50.78	4.49	5.51	46.63	65.55	176.49	136.91	37.66	43.37
Mode	215	85	46	4	4.66	40	67	179	110	37	39
Sample standard deviation	44.11	55.59	17.03	1.55	2.26	15.91	3.78	36.71	23.75	5.05	5.26

3. Chi-square test, logistic regression analysis

The chi-square test is a popular hypothesis testing technique that can be used to compare frequency distribution goodness-of-fit tests, change trend tests of multiple rates, correlation analyses between two categorical variables, and two or more sample rates or composition ratios in completely random designs. It can also be used to compare sample rates in paired designs.

In this study, excel is used to calculate the chi-square test. The steps include creating contingency tables, calculating expected frequency, calculating chi-square statistics, and calculating p-values. Using the CHISQ.DIST. RT function in Excel, you can get the chi-square value and p-value of the calculated variable. An extended version of linear regression analysis, logistic regression analysis is primarily used to examine the connection between independent and dependent variables. It works particularly well when the dependent variable is a numerical variable that represents the occurrence of an event or a binary variable. The independent variables in a logistic regression analysis might be continuous or categorical data, but the dependent variables, such as the prevalence of diabetes, can be qualitative, discrete, or categorical.

Table 3. High-Density Lipoprotein (HDL) levels by gender

Category	Number of women	Number of men	Total
HDL≥35	216	127	343
HDL<35	18	42	60
Total	234	169	403

According to the relationship between sex and HDL, Table 3 was established for chi-square test analysis, the chi-square value was 22.8 and the P-value was 1.8E-06, indicating that there were significant differences in HDL content between different genders.

Table 4. Ideal Body Weight (IBW) categories by gender

Category	Number of women	Number of men	Total
IBW<120%	79	92	171
120%≤IBW≤140%	70	50	120
IBW>140%	85	27	112
Total	234	169	403

DOI: 10.54254/2753-8818/50/2024AU0145

Chi-square test analysis was carried out by establishing a flashtable for gender and IBW, Table 4 was established for chi-square test analysis, and chi-square value was 24.51 and P-value was 4.76E-06, indicating that there were significant differences in obesity levels between different genders.

Table 5. High-Density Lipoprotein (HDL) levels by region

Category	Number of Buckingham	Number of Louisa	Total
HDL≥35	165	177	342
HDL<35	35	26	61
Total	200	203	403

Table 5 was developed for chi-square test analysis based on the association between various regions and HDL; The P-value was 0.000234 and the chi-square value was 13.5, indicating that there were notable variations in the HDL content among the different areas.

Table 6. Ideal Body Weight (IBW) categories by region

Category	Number of Buckingham	Number of Louisa	Total
IBW<120%	90	81	171
120%≤IBW≤140%	56	64	120
IBW>140%	54	58	112
Total	200	203	403

To perform a chi-square test analysis, Table 6 was created. The chi-square test study of IBW and the regions showed that there was no variation in the amount of obesity across the different areas. The P-value was 0.569 and the chi-square value was 1.128. In conclusion, the Chi-square test findings show that while obesity levels do not change significantly between areas, there are variations between genders in high-density lipoprotein levels and obesity. There were only discernible variations in high-density lipoprotein levels. The steps of using SPSS for logistic regression include importing data, specifying variables, and running the analysis. Table 7 shows the logistic regression model results. It is shown that obesity was significantly associated with total cholesterol, high-density lipoprotein, and sex. Logistic regression analysis can help us understand the contribution of each variable to diabetes risk.

Table 7. Logistic regression model results for various effects

Effect	Likelihood of Model	Chi-Square test	Degree Freedom	Significant
Intercept	257.744	0	0	.
Chol	263.725	5.980	2	.050
Hdl	269.340	11.595	2	.003
Age	264.834	7.090	6	.313
Gender	289.847	32.102	2	<.001
location	258.764	1.020	2	.600

4. Discussion

According to the study, obesity, poor HDL cholesterol, and total cholesterol are significant risk factors for diabetes. These results further support the significance of these characteristics in predicting the risk of diabetes and are in line with previous research For example, research has shown that bariatric surgery significantly impacts remission in newly diagnosed obese individuals with type 2 diabetes [4] noted in the introduction. Improved insulin sensitivity and anti-inflammatory properties have also been associated with regular functional food consumption, which may help prevent type 2 diabetes [5]. By giving up smoking, managing weight, and consuming functional foods rich in fiber, antioxidants, and anti-inflammatory qualities, people can effectively prevent and manage diabetes. The incidence of

DOI: 10.54254/2753-8818/50/2024AU0145

diabetes can be effectively decreased and patient quality of life can be enhanced by combining the aforementioned preventive methods.

5. Conclusion

Using logistic regression analysis, Chi-square testing, and descriptive statistics, this study identified many important diabetes risk variables. Basic health indicators were distributed throughout the population in various classifications by descriptive statistical analysis; chi-square tests revealed significant differences in some indicators by region and by sex; and logistic regression analysis further identified important diabetes predictors. Through a detailed statistical analysis of published diabetes data, the paper identified several significant risk factors and preliminarily identified differences between men and women in different indicators. We employed logistic regression analysis, descriptive statistics, and the Chi-square test to better understand how various factors affect diabetes risk. However, due to the limited sample size, the results obtained in this paper need further confirmation, including the differences between different indicators in different regions, and more regions should be selected for statistical analysis. Overall, however, this paper provides insights into the role of different health indicators in predicting diabetes and underscores the importance of continuous monitoring and management of high-risk populations. Future studies should further understand the specific predictive role of these indicators.

References

[1] John, J. -R., & Groves, T. (2013). BMJ Open Diabetes Research and Care. BMJ Open Diabetes Research and Care, 1(1), e000001.
[2] Forouhi, N. G. (2018). Dietary and nutritional approaches for prevention and management of type 2 diabetes. BMJ. British Medical Journal, k2234.
[3] ElSayed, N. A. (2022). 3. Prevention or delay of type 2 diabetes and associated comorbidities: Standards of care in diabetes—2023. Diabetes Care, 46(Supplement_1), S41–S48.
[4] Backholer, K. (2013). Diabetes prevention and treatment strategies. Diabetes Care, 36(9), 2714–2719.
[5] Alkhatib, A. (2017). Functional foods and lifestyle approaches for diabetes prevention and management. Nutrients, 9(12), 1310.
[6] Cunningham-Myrie, C. A. (2015). Associations between neighborhood effects and physical activity, obesity, and diabetes: The Jamaica Health and Lifestyle Survey 2008. Journal of Clinical Epidemiology, 68(9), 970–978.
[7] Ciarambino, T. (2022). Influence of gender in diabetes mellitus and its complication. International Journal of Molecular Sciences, 23(16), 8850.
[8] Kyrou, I. (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 Endocrine Disorders, 20(S1).
[9] Nguyen, T. H. (2023). Factors associated with controlled glycemic levels in type 2 diabetes patients: Study from a large medical center and its satellite clinics in southeast region in the USA. Healthcare, 12(1), 26.
[10] Willems, J. P. (1997). Prevalence of coronary heart disease risk factors among rural Blacks: A community-based study. Southern Medical Journal, 90(8), 814–820.