Influencing Factors of Type 2 Diabetes Mellitus Complicated with Non-Alcoholic Fatty Liver Disease based on The Logistic Regression

Gongyi Xue

College of Sciences, Qingdao University, Qingdao, China

13176393208@qdu.edu.cn

Abstract. Based on the logistic regression equation, this paper analyzes the influencing factors of Type 2 Diabetes Mellitus (T2DM) & non‑alcoholic fatty liver disease (NAFLD) NAFLD. The electronic medical record data of 602 patients with T2DM from 2018 to 2021 in a hospital in Qingdao were divided into the T2DM group (208 cases) and the T2DM combined with NAFLD group (394 cases). The results show that there are no significant differences in SBP, DBP, total cholesterol and fasting blood glucose between the two groups, while the differences in gender, age, BMI, and triglyceride (TG) were statistically significant. Multivariate logistic regression analysis finally determined BMI, TG, and gender as independent influencing factors for T2DM combined with NAFLD, and the regression equation was: Logit(P)=-12.893+0.415×BMI + 0.893×log-TG - 0.405×Gender, (male=0). This study suggests that obesity and lipid metabolism disorders are important risk factors for T2DM combined with NAFLD, and male patients are more likely to have NAFLD, which provides a reference for clinical prevention and treatment.

Keywords: T2DM, NAFLD, logistic regression analysis.

1. Introduction

At present, the prevalence of type 2 diabetes mellitus Type 2 Diabetes Mellitus (T2DM) among adults is 10.5% worldwide, and the prevalence is expected to reach 12.2% by 2045, accompanied by the widespread prevalence of metabolic-related liver diseases. At the same time, the prevalence of non**-**alcoholic fatty liver disease (NAFLD) has reached 32.4% worldwide and continues to grow [1]. Studies show that the global prevalence of NAFLD in patients with T2DM is about 65.04%, while the prevalence of T2DM with NAFLD in China is 52.56% [2]. The risk of cardiovascular disease with NAFLD is twice that of patients without NAFLD compared with T2DM patients without NAFLD, and T2DM with NAFLD and C Reactive protein, uric acid, body mass index, BMI and blood lipids are correlated [2, 3].

T2DM is a lifelong metabolic disease, and its pathological characteristics are insulin secretion insufficiency and insulin resistance caused by damage to pancreatic islet β cells. The clinical manifestations are a significant increase in blood glucose levels, and with the gradual progression of the disease, the disease may be combined with multi-tissue and organ complications. NAFLD refers to a clinicopathological syndrome characterized by hepatocellular steatosis that occurs in the absence of excessive alcohol consumption and other clear factors of liver damage, and is a liver manifestation of metabolic syndrome. Type 2 diabetes combined with non-alcoholic fatty liver disease leads to higher insulin resistance, which seriously threatens the patient's health. Studies have reported that in some Western countries, NAFLD is the key cause of chronic liver disease, and in China, NAFLD has also become the second leading cause of chronic liver disease, second only to chronic viral hepatitis. T2DM complicated by NAFLD is a common type of liver disease in clinical practice, and its pathogenesis has been discussed in various studies. T2DM and NAFLD may be risk factors for each other, and patients with T2DM combined with NAFLD have more obvious glucose and lipid metabolism disorders and insulin resistance. Among them, insulin resistance IR is a key factor, and fatty liver can promote insulin secretion from the pancreas, promote glucose metabolism disorders, and aggravate the severity of T2DM [4]. T2DM patients are more prone to obesity, BMI index is too large, and the content of adipose tissue in internal organs is excessive, leading to liver lipid

metabolism disorders and IR, which further develops NAFLD, thus forming a vicious circle of fatty liver and insulin, and then seriously affects the quality of life and life expectancy of patients [5, 6].

Based on these, this study aims to analyze the influencing factors of type 2 diabetes mellitus combined with non-alcoholic fatty liver disease based on the logistic regression equation, and calculate the impact weights of these influencing factors on T2DM with NAFLD, so as to obtain some factors affecting T2DM combined with NAFLD, and provide a scientific basis for clinical risk assessment, early screening and the formulation of intervention strategies.

2. Data and Methodology

2.1. Data Sources

Select the electronic medical records of patients visiting the Department of Endocrinology in Affiliated Hospital of Qingdao University from 2018 to 2021, provided by the tutor. In strict compliance with the ethical principles of the Declaration of Helsinki and its revised version, this study has been desensitized to retain only some of the necessary physiological and biochemical indicators of the patient, and to exclude patients with self-reported type 1 diabetes mellitus and adult occult autoimmune diabetes, and other special types of diabetes. After screening, regression analysis was conducted based on 602 T2DM patients as research samples. At the same time, according to the official website of NHANES in the United States, alcohol intake (30 g/d for men ≥≥ 20 g/d for women). These patients were selected and divided into the T2DM alone group and the T2DM combined with NAFLD group based on the "Guidelines for the Prevention and Treatment of Nonalcoholic Fatty Liver Disease (2018 Updated Edition)" published by the Fatty Liver and Alcoholic Liver Disease Group of the Hepatology Branch of the Chinese Medical Association in 2018 [7]. Among them, the number of patients with T2DM alone was 208, while the number of patients in the T2DM combined with NAFLD group was 394.

2.2. Treatment Method

For the medical data of 602 patients, the general data, such as gender, age, weight, BMI, systolic blood pressure (SBP), and diastolic blood pressure (DBP) of the two groups were first selected for comparison, where BMI=weight(kg)/height (m2). SPSS 27.0 was used to analyze the data, and some important factors of T2DM combined with NAFLD were explored by multivariate logistic regression analysis, with $P<0.05$ as the difference.

2.3. Data Processing

In the variable view of SPSS, group, numeric type, so that 1=T2DM merges NAFLD group, 2=T2DM group, Gender, numerical type, 0=male, 1=female, Age, BMI, SBP, DBP are set as scales (continuous). First, according to the calculation formula of t-value:

$$t = \frac{\bar{x}1-\bar{x}2}{\sqrt{\mathrm{Sp2}\cdot\left(\frac{1}{n1}+\frac{1}{n2}\right)}} \tag{1}$$

The combined variance Sp2 is:

$$\mathrm{Sp2} = \frac{(n1-1)S1^2+(n2-1)S2^2}{n1+n2-2} \tag{2}$$

Where $\bar{x}$ is the mean, S is the standard deviation, and n is the sample size. By inputting the general data of the two groups into the "data view", the chi-square test was performed for the gender classification, and the independent sample t-test was performed for the normally distributed variables of Age, BMI, SBP, and DBP, and the results are shown in Table 1.

Table 1. Comparison of general data of patients in the two groups

Group	Number of people (persons)	Gender (male/Femalen)	Age ($\bar{x} \pm s$)	BMI ($\bar{x} \pm s$ kg/m^2)	SBP ($\bar{x} \pm s$ mmHg)	DBP ($\bar{x} \pm s$ mmHg)
1	394	275/119	56.8±9.94	31.5±3.70	133.84±19.17	77.82±10.98
2	208	124/84	52.2±10.1	27.0±3.36	133.36±17.85	78.62±12.23
X^2/t		6.314	-5.367	-14.658	-0.299	0.817
P		<0.05	<0.01	<0.01	>0.05	>0.05

According to Table 1, there was no significant difference in SBP and DBP between the two groups, P>0.05. While the T2DM and T2DM&NAFLD's gender difference was statistically significant,P value <0.05. In addition, the age of the T2DM group alone was smaller than that of the T2DM & NAFLD group, P<0.01, which means difference was statistically significant. The BMI of the T2DM & NAFLD group was significantly greater than that of the T2DM group alone, and the difference of P<0.01 was statistically significant.

At the same time, because the three physiological and biochemical indicators of triglycerides, total cholesterol, and fasting blood glucose are generally considered to be strongly correlated with T2DM combined with NAFLD, these three physiological and biochemical indicators are also calculated separately according to the above method (n is the sample size), in order of triglycerides, total cholesterol, and fasting blood glucose.

First, the TG (triglycerides) are calculated, and the same as above, $x1 = 1.98, S1 = 1.57, n1 = 208; x2 = 3.24, S2 = 3.86, n2 = 394$,combined variance Sp2:

$$Sp2 = \frac{(208-1)\times 1.57^2+(394-1)\times 3.86^2}{600} = 10.609 \tag{3}$$

The t-value is calculated

$$t = \frac{1.98-3.24}{\sqrt{10.609\times 0.00735}} = -\frac{1.26}{0.0239} \sim -52.719, \tag{4}$$

S2 =3.86 is almost 1.2 times the mean (3.24), which indicates that the data is extremely skewed, so TG is a non-normally distributed variable, so log-TG is used to process it, and the nonparametric test Mann-Whitney U test is used instead of the t-test.

The TC (total cholesterol) was then calculated as above, $x1 = 5.16, S1 = 1.32, n1 = 208; x2 = 5.13, S2 = 1.22, n2 = 394$,calculate the combined variance

$$S_p^2: Sp2 = \frac{(208-1)\times 1.32^2+(394-1)\times 1.22^2}{600} = 1.576 \tag{5}$$

Calculate The T-Value:

$$t = \frac{5.16-5.13}{\sqrt{1.576\times\left(\frac{1}{208}+\frac{1}{394}\right)}} = \frac{0.03}{0.1076} \sim 0.279 \tag{6}$$

The same is true for the TG step, when t = 0.279, P>0.05 $df = 600$. Next, FBG (fasting blood glucose) is calculated, similar to the previous two,$x1 = 9.23, S1 = 4.53, n1 = 208; x2 = 9.72, S2 = 4.61, n2 = 394$,calculate the combined varianceSp2:

$$Sp2 = \frac{(208-1)\times 4.53^2+(394-1)\times 4.61^2}{600} = 20.998\sim 21.00 \tag{7}$$

Calculate the t-value:

$$t = \frac{9.23-9.72}{\sqrt{21.00\times 0.00735}} = \frac{-0.49}{0.3929} \sim -1.247 \tag{8}$$

The same principle yields a P value of >0.05. The calculation results of the three physiological and biochemical indicators are summarized and displayed, as shown in Table 2.

Table 2. Comparison of physiological and biochemical index levels between the two groups

Group	Number of people (persons)	log-TG (mmol/L)	T C (mmol/L)	FBG (mmol/L)
1	394	2.12 (1.45-3.24)	5.13±1.22	9.72±4.61
2	208	1.45 (1.02-2.10)	5.16±1.32	9.23±4.53
Z/t		8.345	0.279	-1.247
P		<0.01	>0.05	>0.05

According to Table 2, the TG levels of the T2DM with NAFLD group were significantly higher than those in the T2DM group, and P value <0.01 which means the difference was statistically significant, but there was no difference in TC and FBG levels between the T2DM alone and the NAFLD group.

3. Logistic Regression Model Analysis is Introduced

SPSS 27.0 statistical software was used to analyze the data, and the presence or absence of NAFLD was regarded as the dependent variable. It is assumed that there is 1 with NAFLD and 0 without NAFLD; Gender (0 for males and 1 for females), BMI, TG, and age were put separately as covariates, BMI, TG, and age were all continuous variables, and multivariate logistic regression analysis was performed using AIC as the criterion by forward step method. The basic form of the logistic regression model can be expressed by the following formula

$$logit(p) = ln[\,p\,/(1-p)] = \beta 0 + \beta 1 \times X1 + \beta 2 \times X2 + \cdots + \beta k \times Xk. \quad (9)$$

In which P represents the probability of occurrence of events, and in this paper, the probability of occurrence of T2DM combined with NAFLD is present; logit(p) is log-odds, which is the logarithmic ratio of probabilities. β_0 is the intercept of the model, which is a constant term. β_1, β_2, …, β_k is the regression coefficient of the respective variables, corresponding to the characteristic variables X_1, X_2,..., X_2, respectively.

3.1. Univariate Logistic Regression Analysis Separately

The presence or absence of NAFLD was regarded as the dependent variable. The sample size was n=602, and the P value <0.05 was considered to be statistically significant. BMI, log-TG, Gender, and Age were separately inserted, and the output results were shown in Table 3.

Table 3. Univariate logistic regression analysis of independent variables

variable	β	Wald χ^2	P-value	OR
BMI	0.423	154.76	<0.001	1.526
log-TG	0.142	40.11	<0.001	1.330
Gender	0.152	6.65	0.01	0.676
Age	0.029	2.34	0.126	1.045
TC	-0.023	0.14	0.708	0.977
FBG	0.026	3.14	0.122	1.026

It can be seen that BMI, log-TG, and gender are significant in univariate analysis and must be included, and age is not statistically significant, but it is considered important clinically, so it is also temporarily included. Next, the basic model is set up, and the forward step method is used to include variables with P<0.05 as the standard, and the gender is set as the first category (male=0) as the reference. The variable BMI with the smallest P value was selected to enter the model first, and other variables were added based on BMI, in order of log-TG, Gender, and Age, and Age did not significantly improve the model fit (P>0.05), so it was not included, and then multicollinearity was diagnosed, and the final results were sorted out as shown in Table 4.

Table 4. Multivariate logistic regression with independent variables

variable	β	S.E.	Wald χ^2	P	OR	95%CI	VIF
BMI	0.415	0.035	140.89	<0.001	1.515	1.414-1.623	1.12
Log-TG	0.893	0.142	39.56	<0.001	2.442	1.848-3.228	1.08
Gender	-0.405	0.154	6.92	<0.01	0.667	0.493-0.902	1.04
β_0	-12.893	1.145	126.76	<0.001	0.000		

In summary, according to the constant terms in the output results and the regression coefficients of each variable, the logistic regression equation can be obtained, i.e., $Logit(P) = -12.893 + 0.415 \times BMI + 0.893 \times log - TG - 0.405 \times Gender$, Nagelkerke $R2 = 0.428$. Among them, P is the probability of patients suffering from T2DM combined with NAFLD, and body mass index which also called BMI, is a continuous variable, kg/m2. log-TG is the logarithmic value (continuous variable) of triglycerides, and Gender is the categorical variable, 0=male, 1=female. This study is based on a cross-sectional survey and cannot predict long-term trends, which may lead to prediction bias. In the future, multi-center and large-sample clinical data support is still needed to further improve the credibility and generalizability of the model.

3.2. Analyze the Results

The final model included three independent variables: BMI, TG, and Gender, model χ^2=201.52, P<0.001, the model was statistically significant, Nagelkerke R^2~0.428, the model explained about 42.8% of the variation. The significance of each variable was as follows: For patients with T2DM and NAFLD, BMI was the strongest predictor of NAFLD. High TG levels also significantly increase the risk of NAFLD. As for Gender, women are 0.676 times more likely to develop NAFLD than men, which means male T2DM patients are more likely to have NAFLD.

4. Discussion

A total of 602 cases of T2DM were included in this study, of which 394 were complicated by NAFLD, accounting for 65.44%. In this paper, a multivariate logistic regression model is used to screen predictors, which can more effectively deal with the problems of overfitting and multicollinearity between factors, with higher accuracy and efficiency. The results of this study showed that the levels of SBP, DBP, TC and FBG in the T2DM group alone and the T2DM combined with NAFLD group, including age comparison, were not statistically significant in the differences in these factors, P>0.05, and the difference between the two groups was statistically significant, P<0.05, and male T2DM patients were more susceptible to NAFLD. The TG level of the T2DM with NAFLD group was significantly higher than that of the T2DM alone group, and the difference was statistically significant P<0.05, revealing that patients who have T2DM combined with NAFLD may have more severe obesity and triglyceride metabolism disorders. BMI is a risk predictor of NAFLD. The BMI of the T2DM with NAFLD group is greater than that of the T2DM alone group, P<0.01. Obesity leads to excessive accumulation of triacylglycerol and cholesterol in hepatocytes, which in turn induces liver steatosis and inflammation, and increases the risk of more serious diseases such as cirrhosis [8, 9].

The pathogenesis of T2DM is mainly insufficient secretion of IR and islet B cells, the body's inability to effectively utilize insulin, or insufficient insulin secretion is insufficient, resulting in a continuous rise in blood sugar levels, which in turn leads to lesions in blood vessels and nerve tissue, coronary heart disease, and serious complications such as liver and kidney problems. Studies find that the hepatocytes which in patients with insulin resistance gradually degenerate under long-term fat accumulation, and the risk of developing NAFLD is higher [10]. NAFLD is a multisystem disease related to metabolic disorders, although it is a benign disease; if the disease progresses, it may lead to metabolic steatohepatitis, cirrhosis, liver failure, and even hepatocellular carcinoma. There is growing evidence that the harm of NAFLD is not limited to the high morbidity and mortality

associated with liver disease, but is also associated with disability and death rate in patients with T2DM. The results of a 2019 Meta-analysis studies show that nearly 16,000 people with diabetes evaluated the relationship between NAFLD and T2DM and found that people with NAFLD were 2.22 times more likely to develop diabetes than people without NAFLD [11]. In recent years, the incidence of T2DM combined with NAFLD has increased more and more rapidly, and persistent hyperglycemia can have a certain impact on liver detoxification and metabolic capacity, thereby reducing the quality of life of patients [12]. Therefore, it is particularly important to prevent and treat NAFLD based on blood sugar control [13].

5. Conclusion

Based on logistic regression analysis, this study identified key influencing factors for T2DM complicated with NAFLD using clinical data from 602 patients. The results showed that BMI, triglyceride (TG) levels, and gender were independent predictors of T2DM combined with NAFLD. However, the study has several limitations. First, its cross-sectional design prevents the exploration of causality. Second, the sample was sourced from a single hospital, and the data are somewhat outdated, which may limit generalizability. Third, potential confounding factors such as diet, physical activity, and genetic background were not considered. Additionally, the high standard deviation in TG values suggests possible data skewness, which may affect model accuracy.

Future research should adopt multi-center prospective designs with larger and more diverse samples. Including additional variables such as lifestyle factors and genetic markers could improve predictive performance. Longitudinal studies are also needed to validate the long-term predictive power of the model and to explore the dynamic interplay between T2DM and NAFLD.

References

[1] Qun L, Xuemei L. Analysis of risk factors for non-alcoholic fatty liver disease in Tibetan population with type 2 diabetes. Tibetan Medicine, 2025, 46(02): 42-43.

[2] Kumar N, Tan J, Chen Z, et al. Intraoperative cell-salvaged autologous blood transfusion is safe in metastatic spine tumour surgery: early outcomes of prospective clinical study. Eur Spine J, 2023, 32(7): 2493-2502.

[3] Zhou J. A review of the application of autologous blood transfusion. Braz J Med Biol Res, 2016, 49(9): e5493.

[4] Guolan C, Yanyan L, Yanqing H, et al. Analysis of influencing factors of newly diagnosed type 2 diabetes mellitus combined with non-alcoholic fatty liver disease. Clinical Medical Research and Practice, 2024, 9(18): 1-4.

[5] Khneizer G, Rizvi S, Gawrieh S. Non-alcoholic fatty liver disease and diabetes mellitus. Adv Exp Med Biol, 2021, 1307: 417-440.

[6] Powell E E, Wong V W, Rinella M. Non-alcoholic fatty liver disease. Lancet, 2021, 397(10290): 2212-2224.

[7] Fatty Liver and Alcoholic Liver Disease Group, Hepatology Branch of the Chinese Medical Association; Expert Committee on Fatty Liver Disease of the Chinese Medical Doctor Association. Guidelines for the prevention and treatment of non-alcoholic fatty liver disease (2018 updated version). Chinese Journal of Hepatology, 2018, 26(3): 195-203.

[8] Polyzos S A, Kountouras J, Mantzoros C S. Obesity and nonalcoholic fatty liver disease: from pathophysiology to therapeutics. Metabolism, 2019, 92: 82-97.

[9] Angelico F, Baratta F, Pastori D, et al. Statins and non-alcoholic fatty liver disease. Liver Int, 2019, 39(9): 1787-1795.

[10] Taneja V. Letter to the editor: intrahepatic lipid content after insulin glargine addition to metformin in type II diabetes mellitus with nonalcoholic fatty liver disease. Hepatology, 2020, 71(3): 1128.

[11] Mantovani A, Byrne C D, Bonora E, et al. Nonalcoholic fatty liver disease and risk of incident type 2 diabetes: a meta-analysis. Diabetes Care, 2018, 41(2): 372-382.

[12] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2024, 74(3): 229-263.

[13] Zachou M, Flevari P, Nasiri-Ansari N, et al. The role of anti-diabetic drugs in NAFLD: Have we found the Holy Grail? A narrative review. Eur J Clin Pharmacol, 2024, 80(1): 127-150.