

Analysis of Risk Factors for Hypertension

Chengjing Zhang^{1,a,*}

¹*School of Mathematical Sciences, South China Normal University, Guangzhou, 510000, China*

a. 1811030514@stu.hrbust.edu

**corresponding author*

Abstract: With the increasing incidence of hypertension, it is particularly important to understand and prevent hypertension. To prevent hypertension, we need to know what factors may lead to its occurrence. This article aims to explore the influence of smoking, diabetes, BMI and other factors on hypertension. Through the analysis of the data provided by Kaggle website, the single factor and multi factor logistic regression analysis methods were used to study the impact of eight variables on hypertension, including gender, age, BMI, recent smoking status, daily smoking volume, diabetes, cholesterol level and heart rate. The results showed that in the univariate analysis, except for gender, all other factors were significantly correlated with hypertension. In this study, smoking was negatively correlated with hypertension, while other factors were positively correlated. However, in the multivariate analysis, smoking status, daily smoking and diabetes decreased significantly. In conclusion, diabetes and smoking still have an impact on hypertension, but this effect may not be so important. People should pay more attention to weight control, but smoking and diabetes also affect hypertension, and we should also pay attention to the impact of smoking and diabetes. For example, smoking will affect cardiovascular disease, and diabetes is still positively related to hypertension in this study.

Keywords: hypertension, diabetes, smoke, age, BMI.

1. Introduction

Hypertension is a common and highly harmful disease, and there is a close causal relationship between blood pressure levels and the risk of cardiovascular and cerebrovascular diseases and death, posing a serious threat to the health of the global population. The incidence of hypertension is on the rise in China, which has attracted widespread attention from the medical and public health sectors [1]. Hypertension is not only one of the main risk factors for many cardiovascular diseases, but also leads to other complications including heart failure, left ventricular hypertrophy, atrial fibrillation, and end-stage renal disease, seriously affecting the quality of life and socio-economic development of patients [2]. The harm caused by hypertension is multifaceted. Therefore, understanding the influencing factors of hypertension and taking preventive measures is crucial.

Many factors are risk factors of hypertension, such as obesity, smoking, age, diabetes, etc. [3,4]. In particular, the relationship between BMI, obesity and hypertension shows that there is a significant correlation between the increase of BMI, obesity and the incidence rate of hypertension in the population [5]. Research has shown the specific mechanism by which obesity affects hypertension [6]. Some studies also discussed the influence of age on the incidence rate of hypertension [7,8]. As

far as smoking is concerned, people have found that smoking can temporarily increase blood pressure and affect the effectiveness of antihypertensive drugs [9], and smoking can increase the risk of cardiovascular disease [10]. Many articles have revealed the high correlation between hypertension, diabetes and smoking [11]. However, most articles did not discuss whether diabetes and smoking were still one of the risk factors of hypertension after introducing multiple variables. Therefore, it is necessary to further explore the risk factors of hypertension.

In relevant research fields, research shows that the relationship between smoking intensity and the incidence rate of hypertension is still one of the risk factors of hypertension, even after controlling for gender, age, obesity and other confounding factors [11]. Some articles use Cox risk models to explore the mutual effects of obesity, smoking, and familial diseases on hypertension [12], but do not introduce age variables. But some studies have also found that there is no correlation between smoking and hypertension in the population after separating variables [13]. There are still some controversies and research gaps regarding the comprehensive effects of smoking on hypertension. Other studies found that diabetes is a risk factor for hypertension by using logistic regression analysis [11-13], but age was not included in the important variable.

Therefore, this paper will use Logistic regression analysis to analyze various risk factors of hypertension, and further explore whether smoking and diabetes are significantly associated with hypertension, considering age and BMI. By revealing these impacts, this study aims to provide theoretical support for the prevention and control of hypertension in high-risk populations.

This article mentioned the relationship between hypertension and age, BMI, diabetes, smoking and other risk factors, which improved the awareness of patients to prevent hypertension, so that people can better prevent hypertension. The next part of this study is as follows. The second part is Method, which introduces the data source of this article, Variable selection, Method introduction. The third part is Results and Discussion, which presents the findings of this study and points out similarities and differences with previous research. The fourth part, Conclusion, summarizes the research findings of this article.

2. Method

2.1. Data source

The data in this literature was collected from the Kaggle website, compiled by MD RAIHAN KHAN and updated for publication in April 2024. In this dataset, 13 important factors associated with hypertension were collected, and this study selected 8 variables with research significance to study the risk factors of hypertension.

2.2. Variable selection

This study used 4440 samples from the dataset for statistical purposes. Eight variables were selected as gender, age, BMI, recent smoking status, number of cigarettes per day, whether diabetes is ill, cholesterol level, and heart rate. Use symbols such as male, age, BMI, currentSmoker, cigsPerDay, diamonds, totalChol, heartRate, etc. to represent these eight variables. When measuring blood pressure, $SBP \geq 140$ mmHg and/or $DBP \geq 90$ mmHg are considered to be at risk of hypertension in this study, with a risk of 1 to facilitate subsequent logical analysis.

2.3. Method introduction

This study used SPSS 27.0 to analyze the research. Use binary logistic regression to analyze the influencing factors of each variable on hypertension. Then, use multiple logistic regression analysis to analyze the impact of multiple factors on hypertension. In this way, we can judge and compare

whether smoking and diabetes are factors affecting hypertension. Then, the collinearity test is used to analyze whether there is interference between the factors. And further through hierarchical analysis, after controlling age and BMI, whether the influence of smoking and diabetes on hypertension remains significant.

3. Results and Discussion

Table 1: Single factor Logistic regression analysis

Element	B	SE	Wald χ^2	P	Exp(B)
male	0.026	0.067	0.145	0.703	1.026
	-0.808	0.044	337.105	<0.001	0.446
age	0.081	0.004	369.418	<0.001	1.085
	-4.906	0.221	495.003	<0.001	0.007
BMI	0.169	0.009	330.141	<0.001	1.184
	-5.216	0.248	442.654	<0.001	0.005
currentSmoker	-0.451	0.067	45.325	<0.001	0.637
	-0.584	0.045	168.075	<0.001	0.558
cigsPerDay	-0.013	0.003	18.582	<0.001	0.987
	-0.686	0.041	277.033	<0.001	0.504
diabetes	0.953	0.195	23.921	<0.001	2.594
	-0.825	0.034	595.302	<0.001	0.438
totChol	0.008	0.001	106.703	<0.001	1.008
	-2.705	0.190	203.679	<0.001	0.067
heartRate	0.026	0.003	88.677	<0.001	1.026
	-2.799	0.217	166.777	<0.001	0.061

In Table 1, except for gender, each factor has a significant association with hypertension. There is a negative correlation between recent smoking status and daily smoking. Age, BMI, diabetes, cholesterol level, heart rate are positively correlated

Table 2: Multivariate Logistic regression analysis

Element	B	SE	Wald χ^2	P	Exp(B)
male	0.137	0.081	2.837	0.092	1.147
age	0.077	0.005	257.555	<0.001	1.080
BMI	0.157	0.010	242.372	<0.001	1.169
currentSmoker	-0.143	0.121	1.391	0.238	0.867
cigsPerDay	0.004	0.005	0.479	0.489	1.004
diabetes	0.215	0.222	0.938	0.333	1.240
totChol	0.004	0.001	21.078	<0.001	1.004
heartRate	0.028	0.003	79.696	<0.001	1.029
	-11.934	0.496	579.093	<0.001	0.000

In Table 2, the smoking status, the number of cigarettes per day, and the prevalence of diabetes are no longer significant. Other factors remain significant.

Table 3: Multicollinearity test

Element	B	SE	Wald χ^2	t	P	1/VIF	VIF
	-1.612	0.071		-22.780	0.000		
male	0.014	0.014	0.015	0.996	0.319	0.923	1.084
age	0.014	0.001	0.254	16.945	0.000	0.875	1.143
currentSmoker	-0.014	0.014	-0.016	-1.048	0.295	0.881	1.135
diabetes	0.051	0.042	0.017	1.236	0.217	0.980	1.020
totChol	0.001	0.000	0.056	3.838	0.000	0.910	1.099
BMI	0.028	0.002	0.248	17.072	0.000	0.927	1.078
heartRate	0.005	0.001	0.130	9.075	0.000	0.961	1.040

Based on the multiple logistic regression analysis above, this study considers whether there is an exact correlation between variables, and conducts a multicollinearity test. Through the test results, it is found that there is no exact correlation between the various factors in this study.

Table 4: Multivariate logistic regression analysis including age, BMI, currentSmoker, and diamonds

Element	B	SE	Wald χ^2	P	Exp(B)
age	0.077	0.005	290.775	0.000	1.080
currentSmoker	0.001	0.075	0.000	0.995	1.001
BMI	0.161	0.010	273.017	0.000	1.175
diabetes	0.322	0.219	2.149	0.143	1.379

In the table 4, after introducing age and BMI, the association between currentSmoker, diamonds, and hypertension was not significant, Age and BMI are important factors affecting the association between current Smoker, diabetes, and hypertension. Therefore, we will control for these two factors separately and use univariate logistic regression analysis to explore the association between current Smoker, diamonds, and hypertension.

Table 5: Logistic regression analysis of diabetes after BMI grouping

Element	B	SE	Wald χ^2	P	Exp(B)
BMI<25					
diabetes	0.594	0.403	2.176	0.140	1.811
	-1.441	0.058	609.722	0.000	0.237
currentSmoker	-0.494	0.116	18.155	0.000	0.610
	-1.163	0.082	200.101	0.000	0.312
25=<BMI<30					
diabetes	0.782	0.294	7.073	0.008	2.186
	-0.615	0.051	147.080	<0.001	0.541
currentSmoker	-0.174	0.101	2.994	0.084	0.840
	-0.515	0.066	60.161	0.000	0.597
BMI>=30					
diabetes	0.673	0.425	2.506	0.113	1.960
	0.292	0.090	10.661	<0.001	1.339
currentSmoker	-0.159	0.182	0.756	0.385	0.853
	0.382	0.109	12.296	0.000	1.465

In order to better understand the complex relationship between diabetes, BMI and hypertension, further hierarchical analysis was conducted: the data were classified by BMI, and the impact of diabetes on hypertension was analyzed in different BMI groups. In these three groups, the sample size is 1927,1755,593 people.

In the table 5, classify BMI. In the group with BMI<25, diabetes was no longer significant, and smoking was still significant. In the group with 25≤BMI<30, diabetes was still significant, but significantly decreased, and whether smoking was no longer significant. In the group with BMI≥30, diabetes and smoking were no longer significant.

Table 6: Logistic regression analysis of diabetes after further grouping of BMI

Element	B	SE	Wald χ^2	P	Exp(B)
25<BMI<27.5					
diabetes	0.836	0.464	3.246	0.072	2.308
	-0.731	0.066	121.557	<0.001	0.481
27.5<BMI<30					
diabetes	0.651	0.382	2.905	0.088	1.917
	-0.443	0.079	31.300	<0.001	0.642

In the table 6, further divided into two groups, with a BMI between 25 and 30, consisting of 1056,699 individuals. After further grouping 25≤BMI<30, we found that diabetes was no longer significantly associated with hypertension.

Table 7: Single factor logistic regression analysis after age grouping

Element	B	SE	Wald χ^2	P	Exp(B)
30<age<45					
diabetes	0.745	0.681	1.196	0.274	2.106
	-1.726	0.074	541.027	0.000	0.178
currentSmoker	-0.115	0.150	0.586	0.444	0.892
	-1.650	0.114	207.932	<0.001	0.192
45≤age<60					
diabetes	0.753	0.261	8.358	0.004	2.124
	-0.655	0.046	198.949	0.000	0.520
currentSmoker	-0.290	0.092	10.030	0.002	0.748
	-0.495	0.062	63.744	<0.001	0.609
60<age<75					
diabetes	0.414	0.348	1.414	.234	1.512
	0.083	0.078	1.116	.291	1.086
currentSmoker	-0.233	0.165	1.998	0.158	0.792
	0.177	0.092	3.696	0.055	1.194

Classify the data by age and analyze the impact of recent smoking on hypertension in different age groups. In these three groups, the sample size is 1427,2123,690

In table 8, the 30<age<45 and 60<age<75 groups, smoking and diabetes were no longer significantly related to hypertension. However, at 45≤age<60, these two are still significantly associated with hypertension.

Table 8: Single factor logistic regression analysis after further age grouping

Element	B	SE	Wald χ^2	P	Exp(B)
45= \leq age= \leq 52					
diabetes	0.908	0.359	6.380	0.012	2.478
	-0.908	0.064	201.266	0.000	0.404
currentSmoker	-0.194	0.126	2.374	0.123	0.824
	-0.779	0.090	74.407	<0.001	0.459
52<age<60					
diabetes	0.513	0.356	2.083	0.149	1.671
	-0.331	0.065	25.992	0.000	0.718
currentSmoker	-0.263	0.139	3.606	0.058	0.769
	-0.214	0.087	6.050	0.014	0.807

After stratified analysis, it was found that smoking was no longer significant in the age groups of 30-45 and 60-75. Further stratify and divide the age groups between 45 and 60 into two groups with a total of 1225,898 people.

In table 8, the significance of smoking and diabetes decreased. Except for diabetes, which is still significant in the group of 45= \leq age= \leq 52, the others are no longer statistically significant.

In conclusion, in this study, smoking is negatively related to hypertension, This is different from previous studies that have shown a positive correlation between smoking and hypertension [3,4,12],and diabetes is positively related to hypertension. In the multivariate logistic regression analysis of this study, the significance of smoking and diabetes disappeared, and the other five factors still maintained significant correlation. In collinearity testing, there is no exact correlation between the eight variables. This indicates that the multivariate logistic regression analysis was not disrupted. And after controlling age and BMI, the correlation significance of smoking, diabetes and hypertension has decreased. We infer that they may still have an impact on hypertension, but the effect is relatively small, not significant enough, not big enough to be a risk factor. Of course, this may also be due to a small sample size, which increases the uncertainty of estimation, leading to a decrease in significance and an increase in p-value.

4. Conclusion

Based on the above research results, this paper aims to analyze the impact of smoking, diabetes and other factors on hypertension, and explore their performance in multivariate logistic regression analysis. By analyzing the data provided by Kaggle website, this paper uses single factor and multi factor logistic regression analysis methods to study the impact of eight variables on hypertension, including gender, age, BMI, smoking, diabetes, cholesterol level and heart rate.

In univariate logistic regression analysis, except for gender, all other factors were significantly correlated with hypertension. In particular, smoking status was negatively correlated with daily smoking volume, while age, BMI, diabetes, cholesterol level and heart rate were positively correlated. However, in the multivariate logistic regression analysis, the significance of smoking status, daily smoking and diabetes decreased, while other factors remained significantly correlated.

Further multicollinearity testing showed that there was no significant correlation between the eight variables in the study, indicating the reliability of the results of multiple logistic regression analysis. In addition, after controlling age and BMI, the effects of smoking and diabetes on hypertension were significantly reduced. Therefore, it can be inferred that smoking and diabetes may still affect hypertension, but their effects are relatively small, not enough to be significant risk factors.

In conclusion, this study reveals the potential role of smoking and diabetes in the pathogenesis of hypertension. However, these effects may not be significant due to the small sample size. Future research can further expand the sample size to reduce the uncertainty of estimation, so as to more accurately assess the actual impact of smoking and diabetes on hypertension.

Although smoking did not show any correlation with hypertension in this study, smoking increases the probability of cardiovascular disease and should be prevented. Patients should reduce smoking. Diabetes has no obvious significance in this study, but it is still positively related to hypertension. diabetes patients should pay more attention to living habits and prevent hypertension.

References

- [1] *Summary of China Cardiovascular Health and Disease Report 2020 [J]. Chinese Journal of Circulation, 2021,36 (06): 521-545*
- [2] *Chinese Guidelines for the Prevention and Treatment of Hypertension (Revised 2018 Edition) [J]. Chinese Journal of Cardiology, 2019,24 (01): 24-56*
- [3] *Yan Ruiping, Wang Xiliang, Yao Fenxia, et al. Comparison of the effectiveness of decision tree model and logistic regression analysis model in identifying risk factors for hypertension [J]. Chinese Journal of Disease Control and Prevention, 2022, 26 (02): 218-222. DOI: 10.16462/j.cnki. zhjbkz. 2022.02.016*
- [4] *Gao Jinchai, Peng Wei. Research progress on the etiology, risk factors, and treatment of primary hypertension [J]. Journal of Hebei North University (Natural Science Edition), 2016,32 (02): 44-53*
- [5] *Huang Yali, Pan Yan, Lin Daiqian, et al. The relationship between different obesity indicators and hypertension in adults in Fujian Province [J]. Chinese Journal of Hypertension (Chinese and English), 2024,32 (02): 171-179. DOI: 10.16439/j.issn.1673-7245.2024.02.02.012*
- [6] *Zhou Chuanfei, Fu Xiuying, Zhou Xingye, et al. Research progress on the mechanism of obesity induced hypertension [J]. Journal of Gannan Medical College, 2022, 42 (12): 1268-1271+1278*
- [7] *Zhu Huijuan. Analysis of the interaction between overweight, obesity, and age on hypertension [J]. Jiangsu Preventive Medicine, 2023,34 (01): 33-35. DOI: 10.13668/j.issn.1006-9070.2023.01.008*
- [8] *Wang Hui, Qi Mengmeng, Guo Xin. Analysis of the current situation and influencing factors of pre hypertension in middle-aged and young people undergoing physical examinations in Zhengzhou City [J]. South China Preventive Medicine, 2021,47 (04): 457-460*
- [9] *Ling Yaqin. Research progress on the impact of smoking on hypertensive patients [J]. General Nursing, 2015,13 (30): 3001-3003*
- [10] *Liu Fangyuan, Ren Mingfen. Research progress on smoking induced cardiovascular diseases [J]. Journal of Xinxiang Medical College, 2023,40 (05): 497-500*
- [11] *Su Yun, Tian Qingfeng, Liu Beibei, et al. The relationship between smoking intensity and hypertension and diabetes among adult residents in Henan Province [J]. Journal of Zhengzhou University (Medical Edition), 2024,59 (01): 109-112. DOI: 10.13705/j.issn.1671-6825.2022.09.057*
- [12] *Yu Shuping, Fan Qiongling, Yang Fei, et al. The impact of family history, overweight/obesity, and smoking interaction on the incidence of hypertension among herdsmen in the Nanshan pastoral area of Xinjiang [J]. Journal of Xinjiang Medical University, 2021,44 (06): 753-758*
- [13] *Xing Huiyuan, Liu Yan, Zhao Ai, et al. The effects of smoking, alcohol consumption, and dietary factors on hypertension in individuals with hyperuricemia [J]. Chinese Journal of Disease Control, 2018,22 (06): 555-559. DOI: 10.16462/j.cnki. zhjbkz. 2018.06.004*