Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/42/2024CH0210

The research of influence factors that possibly lead to cardiovascular disease

Yuchen Dou

Ivy Experimental High School, Zhang Jiagang, 215600, China

fupiao@ldy.edu.rs

Abstract. The emergence of cardiac disease is influenced by numerous variables. Every year a large percentage of people around the world die from heart attacks. This study was to look at how important clinical variables are connected to the chance of having CVD. The analysis used a Kaggle dataset with details from 70,000 patients and 10 different variables. The study examined how factors such as age, gender and changes in cardiac activity during exercise affect overall heart health. Multiple linear regression models were used to analyze these effects. The results show a significant correlation between these factors and an increased risk of heart disease. This emphasizes the importance of these predictors in clinical assessments. It can be concluded from this study that regular medical check-ups, early prevention and treatment should be carried out for these vulnerable groups. The general prevalence of heart disease in the nation can be decreased by implementing these strategies. According to the study, personalized health strategies have the potential to improve CVD outcomes and strengthen preventive measures. All of the experimental results suggest that continued documentation and study of these pathogenic factors in medical diagnostics and experiments could aid in drug development as well as improve medical technology and help more patients with cardiovascular disease recover.

Keywords: Cardiovascular disease, risk factors, multiple linear regression.

1. Introduction

Throughout history, one of the leading causes of death globally has been cardiovascular disease. It takes the lives of at least 17.9 million people every year. This represents 32% of the entire global death population, with 85% of these deaths caused by heart attacks and strokes [1]. The large impact of CVD on public health requires continuous research to understand its causes, how to prevent it, and how to treat it. Research has demonstrated the complexity of cardiovascular disease. Numerous factors influence its etiology. These risk factors are usually divided into two groups: modifiable and non-modifiable [2]. The way these factors interact with each other makes preventing and treating CVD more complicated, which is why healthcare professionals and researchers focus a lot on them.

Diet is a major modifiable factor that affects how the cardiovascular system works and has a big impact on heart health. Eating a lot of saturated fats, trans fats, and cholesterol is closely linked to a higher chance of developing atherosclerosis. The accumulation of fat in the arteries, known as atherosclerosis, can lower blood flow and raise the possibility of a heart attack or stroke. [3]. On the other hand, eating a diet rich in fruits, vegetables, whole grains, and proteins can lower the risk of CVD

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/42/2024CH0210

[4]. The Mediterranean diet is especially good for the heart because it has a lot of unsaturated fats, fiber, and antioxidants [5].

Physical activity is also very important for preventing cardiovascular disease. Frequent exercise increases cholesterol levels, insulin sensitivity, blood pressure, and helps people maintain a healthy weight [6]. The American Heart Association recommends at least 150 minutes of moderate-intensity aerobic exercise per week to keep the heart healthy [7]. Conversely, leading a sedentary lifestyle raises the chance of developing cardiovascular illnesses. This implies that physical activity should be a regular part of a person's life [8].

One well-known cause of cardiovascular disease is smoking. The risk of developing cardiovascular disease is two to four times higher in smokers than in non-smokers, according to studies [9]. Smoking damages the blood vessels, causes inflammation, and increases the chance of blood clots, which can lead to atherosclerosis [10]. As a result, giving up smoking is crucial to preventing CVD, and numerous studies have demonstrated that doing so dramatically lowers a person's chance of developing CVD [11].

Although lifestyle modifications are a major component in lowering the risk of CVD, an individual's risk of contracting CVD is also greatly influenced by unchangeable factors such as gender, and heredity. In addition to genetic variables that increase the chance of acquiring the disease, a person is more likely to have CVD if there is a family record of the condition [12]. Another significant component is age, as the accumulation of daily stress on the arteries with aging raises the risk of cardiovascular disease [13]. Furthermore, there are gender differences in CVD risk. Men are more likely to get CVD at earlier ages than women, whose risk rises dramatically after menstruation [14].

In conclusion, understanding the many factors that contribute to CVD is necessary to create effective strategies for prevention and treatment. This paper will review the different modifiable and non-modifiable risk factors associated with CVD, with a special focus on diet, physical activity, smoking, and genetic susceptibility. This research attempts to contribute to the information required to address the worldwide burden of CVD by identifying these factors.

2. Methodology

2.1. Data source

Jocelyn Dumlao's "Cardiovascular Disease Dataset" on Kaggle is the source of the dataset used in this study. This heart disease dataset, sourced from a reputable multispecialty hospital in India, comprises a rich array of information encompassing 14 common features, making it a valuable resource for cardiovascular research. With a cohort of 1000 subjects and 12 distinct features, this dataset serves as a pivotal tool for developing early-stage heart disease detection methods and constructing predictive machine-learning models.

2.2. Variable selection

The dataset consists of data from 70,000 patients and includes variables about various personal and lifestyle factors. An extensive examination of the numerous risk variables connected to cardiovascular disease (CVD) is made possible by this large dataset.

Table 1 presents a thorough synopsis of 14 typical features linked to cardiovascular disease (CVD). These features include critical clinical indications including maximal heart rate, serum cholesterol levels, resting blood pressure, and fasting blood glucose, in addition to demographic aspects like age and gender. These traits were essential in this study's research of cardiovascular health outcomes since the medical community acknowledges their significance in assessing a person's risk of developing cardiovascular disease (CVD).

Table 1. Cardiovascular Disease Dataset Description

Attribute	Unit	Type of data
Patientid	Numeric	Number
Age	Numeric	In Years

DOI: 10.54254/2753-8818/42/2024CH0210

Table 1. (continued).

Gender	Binary	0 (female) / 1 (male)
Resting BP	Numeric	94-200 (in mm HG)
serumcholestrol	Numeric	126-564 (in mg/dl)
Fastingbloodsugar	Binary	0(false), 1(true) > 120 mg/di
Chestpain	Nominal	0(typical angina), 1(atypical angina), 2(non-anginal pain), 3(asymptomatic)
Restingelectro	Nominal	0(normal), 1(ST-T wave abnormality), 2 (probable or definite left ventricular hypertrophy)
Maxheartrate	Numeric	71-202
Exerciseangina	Binary	0(no), 1(yes)
Oldpeak = ST	Numeric	0-6.2
Slope	Nominal	1(upsloping), 2(flat), 3(downsloping)
noofmajorvessels	Numeric	0,1,2,3
Target	Binary	0(Absence of Heart Disease), 1(Presence of Heart Disease)

2.3. Method introduction

This paper attempts to examine the correlation between different risk variables and CVD using multiple logistic regression models. Since the dependent variable (presence of CVD) is a dichotomous variable, logistic regression is applicable to this analysis.

The probability that a person has CVD was determined using the logistic regression model in relation to specific variables. The model provides a coefficient for each predictor to quantify its relationship to the likelihood of developing CVD.

3. Results and discussion

3.1. Histograms for numerical variables

Histograms were employed to examine the distribution patterns of various numerical variables, while specific variables such as age, resting blood pressure, maximum heart rate, and peak heart rate were analyzed for their correlations (Figure 1). The age distribution generally followed a normal curve but showed a slight skewness towards older age groups. The majority of the patients fell within the 40 to 70-year range, with a pronounced concentration around the 55 to 65-year age bracket.

Resting blood pressure values exhibited a right-skewed distribution, with most patients recording levels between 120 and 150 mmHg. A smaller subset of patients had significantly higher blood pressure readings, which contributed to the skewness. Similarly, serum cholesterol levels also demonstrated a right-skewed pattern. The bulk of the patients had cholesterol levels ranging between 200 and 300 mg/dL, with only a few cases exceeding the 300 mg/dL mark.

In contrast, the distribution of maximum heart rate was left-skewed, indicating that most patients had a maximum heart rate between 120 and 170 beats per minute, with the highest frequency observed in the 140 to 160 beats per minute range. This distribution suggests a tendency for patients to achieve higher heart rates during peak exercise.

Additionally, the distribution of exercise-induced ST-segment depression displayed right-skewness when compared to resting levels. Most values were clustered between 0 and 2, suggesting that mild ST-segment depression was prevalent among the majority of patients. This skewness points to a general trend of mild to moderate exercise-induced ischemia in the study population.

Overall, these distribution patterns offer valuable insights into the typical ranges and outliers for key cardiovascular risk factors in the patient cohort.

DOI: 10.54254/2753-8818/42/2024CH0210

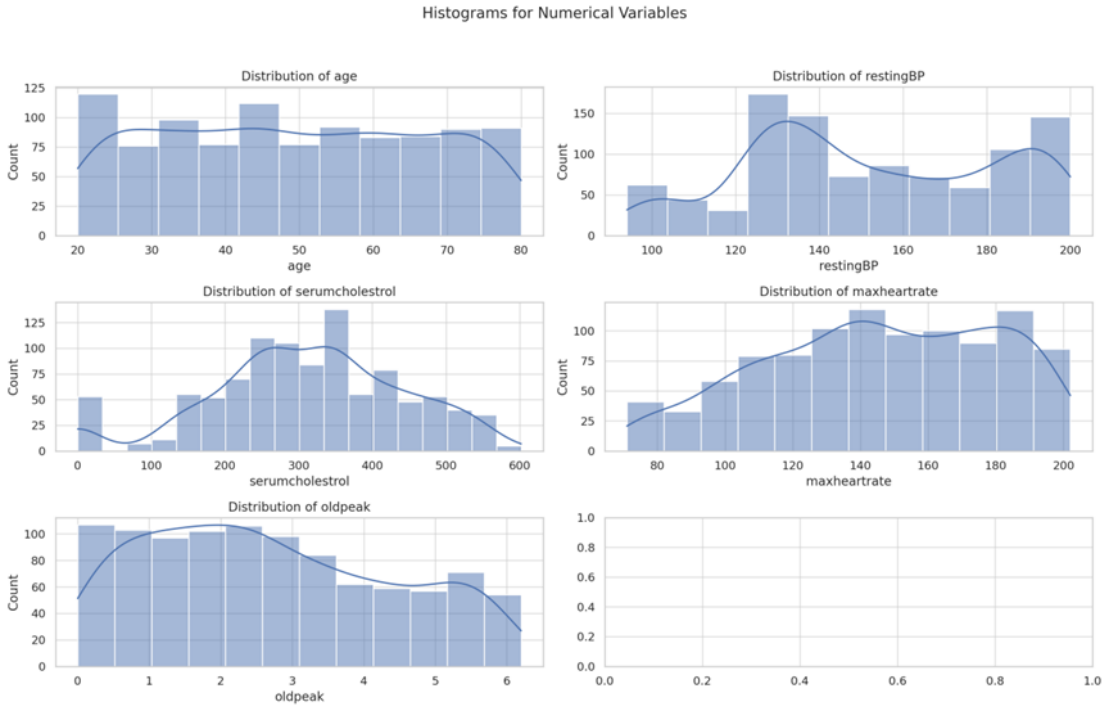

Figure 1. Histograms for numerical variables

3.2. Boxplot for variables

Resting blood pressure, serum cholesterol, and oldpeak show several outliers. These outliers might indicate patients with unusually high blood pressure, cholesterol levels, or ST depression. Age and maximum heart rate show fewer outliers, suggesting more consistent distributions. The presence of outliers in resting blood pressure, serum cholesterol, and oldpeak may affect model performance (Figure 2).

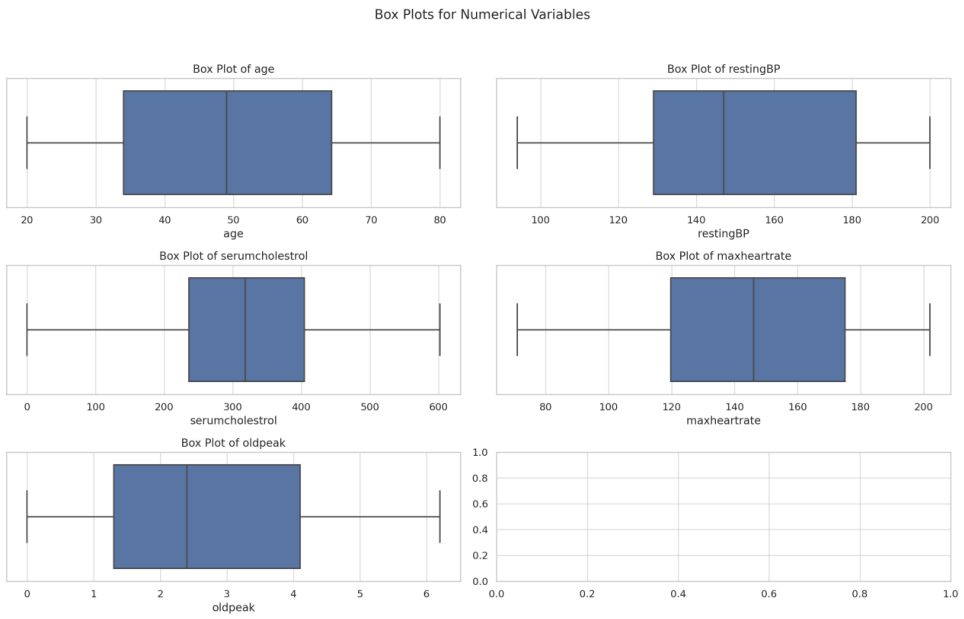

Figure 2. Box Plots for Numerical Variables

DOI: 10.54254/2753-8818/42/2024CH0210

4. Conclusion

This study provides valuable insights into the various factors that influence cardiovascular disease (CVD) risk and highlights the central role of certain clinical and demographic variables. By applying multiple linear regression modeling to the combined dataset, this paper identified several important predictors of CVD risk, including age, resting blood pressure, maximum heart rate, and exercise-induced ST-segment depression. It was shown that there was a strong association between each of these factors and the chance of developing cardiovascular disease. The findings emphasize the need for regular monitoring and effective management of these parameters in the clinical setting, which is essential for early identification and intervention in high-risk populations.

The significance of this study also extends to the field of public health, suggesting that focusing on and addressing these specific risk factors can significantly reduce the incidence and consequences of CVD. Future studies could improve the predictive accuracy of the model by expanding the analysis to include additional variables such as genetic predisposition, lifestyle choices, and psychosocial stressors. In addition, incorporating advanced machine learning techniques could further refine these models to provide more nuanced insights and facilitate the development of sophisticated predictive tools. All things considered, this work forms the basis of a growing corpus of research that supports the implementation of a comprehensive, personalized approach to the treatment and avoidance of cardiovascular disease.

References

[1] World Health Organization. (2021). Cardiovascular diseases (CVDs).
[2] Yusuf S, Hawken S and Ounpuu S 2004 Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries: case-control study. The Lancet, 364(9438), 937-952.
[3] Mozaffarian D, Wilson P W and Kannel W B 2008 Beyond established and novel risk factors: Lifestyle risk factors for cardiovascular disease. Circulation, 117(23), 3031-3038.
[4] Estruch R, Ros E, Salas-Salvadó J, Covas M I, Corella D, Arós F and Martínez-González M A 2013 Primary prevention of cardiovascular disease with a Mediterranean diet. New England Journal of Medicine, 368(14), 1279-1290.
[5] Tosti V, Bertozzi B and Fontana L 2018 Health benefits of the Mediterranean diet: metabolic and molecular mechanisms. The Journals of Gerontology: Series A, 73(3), 318-326.
[6] Mora S, Cook N, Buring J E, Ridker P M and Lee I M 2007 Physical activity and reduced risk of cardiovascular events: potential mediating mechanisms. Circulation, 116(19), 2110-2118.
[7] American Heart Association 2018 Recommendations for physical activity in adults and kids.
[8] Lee I M, Shiroma E J, Lobelo F, Puska P, Blair N, Katzmarzyk P T, et al. 2012 Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. Lancet 380, 219-229.
[9] Ambrose J A and Barua R S 2004 The pathophysiology of cigarette smoking and cardiovascular disease: an update. Journal of the American College of Cardiology, 43(10), 1731-1737.
[10] Benowitz N L 2004 Cigarette smoking and cardiovascular disease: pathophysiology and implications for treatment. Progress in Cardiovascular Diseases, 46(1), 91-111.
[11] Critchley J A and Capewell S 2003 Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. JAMA, 290(1), 86-97.
[12] Roberts R and Stewart A F R 2012 Genes and coronary artery disease: where are we? Journal of the American College of Cardiology, 60(18), 1715-1721.
[13] North B J and Sinclair D A 2012 The intersection between aging and cardiovascular disease. Circulation Research, 110(8), 1097-1108.
[14] Maas A H E M and Appelman Y E 2010 Gender differences in coronary heart disease. Netherlands Heart Journal, 18(12), 598-602.