# Exploring the Impact of Lifestyle Factors on the Development and Progression of Heart Disease

ALBAYRAK, Melike Gülşen*

**Abstract**

Heart disease remains a significant global health concern, prompting me to delve into the relationship between lifestyle choices and the development and progression of this condition. Utilizing the Heart Disease Dataset ,which contains detailed clinical and demographic information on patients with several variables related to heart disease, such as age, sex, blood pressure, cholesterol levels, and presence of heart disease, sourced from the UCI Machine Learning Repository, I analyzed a collection of 303 patient records featuring 14 distinct variables. According to this dataset, i designed the research question driving this study as: "What is the impact of lifestyle factors (such as smoking, physical activity, diet, and alcohol consumption) on the development and progression of heart disease?" For the introductory section i seeked to provide an outline of the importance of heart disease as a leading cause of illness and death worldwide. The section basically emphasized the need to understand the role of lifestyle factors in the development and management of heart disease. Then, in the literature review i combined relevant studies. During the "data" part, in order to ensure data accuracy, I performed initial data cleaning procedures to address missing information. Subsequently, I examined several key variables, including age, sex, chest pain type, resting blood pressure, cholesterol levels, and more, aiming to reveal their connections to heart disease. Using regression analysis techniques, I isolated the independent impacts of various lifestyle factors while controlling for other variables. This method facilitated a comprehensive understanding of how specific lifestyle choices influenced the risk of heart disease. To enhance data visualization, I also generated scatter plots, visualizing the correlation between age and resting blood pressure, as well as the link between heart disease (indicated by the target variable) and fasting blood sugar levels. These visualizations served to comprehension of potential relationships between lifestyle factors and heart disease. Notably, older people tended to exhibit higher resting blood pressure, plausibly attributed to age-related changes in blood vessels and diminished physical activity. Moreover, individuals with heart disease exhibited elevated fasting blood sugar levels, which could be linked to smoking and unhealthy dietary patterns. Specifically, nicotine consumption can heighten blood sugar levels by stimulating stress hormone release, whereas diets that has harmful beverages and foods. Consequently, these dietary choices may cause to bad cardivascular situations. These findings align

*21080562, Github Repo

with existing research emphasizing the pivotal role of a healthy lifestyle in preventing and managing heart disease. By renouncing smoking, engaging in regular physical activity, adopting a balanced diet, and practicing moderate alcohol consumption, individuals, especially elder population, can significantly reduce their heart disease risks. So, my research underscores the significance of lifestyle factors in heart diseas., By getting more into healthier habits and choices, we can foster improved heart health. Nonetheless, it is still essential to understand the limitations inherent in my study. The dataset employed may not encompass all other determinants influencing heart disease, such as genetics and socioeconomic factors etc. Consequently, future investigations should aim to explore these factors more comprehensively to garner a more nuanced understanding. To sum up, my project represents a comprehensive analysis of the influence of lifestyle factors on heart disease using the Heart Disease Dataset. It is my hope that these findings will contribute to public health initiatives, inform policy decisions, and empower individuals to proactively safeguard their cardiovascular well-being.

# 1 Introduction

Heart disease is still a major public health problem worldwide, causing a large burden of morbidity and mortality. Lifestyle habits such as smoking, physical exercises, diet and alcohol consuming have long been accepted as important changeable factors that can affect the occurrence and progression of cardiovascular heart diseases. Searching the impact of these lifestyle factors on heart health is essential for effective prevention and treatment strategies. The purpose of this project is to investigate the interplay between lifestyle factors and the presence or severity of heart diseases, using the Heart Disease Dataset from the UCI Machine Learning Repository which is a website that i found from my searchings about the most helpful dataset providers to find suitable datasets for researchs. After my investigations for finding a good dataset, I chose the Heart Disease Dataset that has 303 observations with a set of 14 clinical and demographic variables including categorical (sex, chest pain type (cp), fasting blood sugar $> 120$ mg/dl (fbs), resting electrocardiographic results (restecg), exercise induced angina (exang), slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), thalassemia (thal), target), discrete numeric (maximum heart rate achieved (thalach)) and continous numeric (age, resting blood pressure (trestbps, in mm Hg), serum cholesterol (chol, in mg/dl), ST depression induced by exercise relative to rest (oldpeak)) variables. After chosing the dataset, i pursued on looking for articles related to heart diseases and i saw that how specific lifestyle behaviours contribute to the risk and progression of cardiovascular heart problems. So at last, i designed my question as "What is the impact of lifestyle factors (such as smoking, physical activity, diet, and alcohol consumption) on the development and progression of heart disease?". Then, i just focused more specifically on the articles related to my research question for the project.

## 1.1 Literature Review

The impact of lifestyle choices on the development and progression of heart disease has been a subject of many studies. This paper focuses on reviewing the results of a sample of studies on smoking, physical activity, diet and alcohol consumption in relation to cardiovascular health. One of the studies focused on the effects of smoking and smoking cessation on cardiovascular events older individuals. The results showed a clear association between smoking and increased cardiovascular risk, while quitting was linked to a reduced risk of mortality (Mons et al., 2015). In terms of physical activity, Carnethon (2009) looked at how much physical activity is enough to be beneficial for our hearts. The evidence suggests that moderate-intensity physical activity with at least 150 minutes per week produces important cardiovascular benefits. Also the role of diet in preventing heart disease was investigated by Bechthold et al. (2019) through systematic searches and dose-response meta-analyses looking at food categories and their association with coronary heart diseases, heart failure and stroke They found that higher intakes of fruit, vegetables, whole grains etc. were associated with a lower risk of these coronary outcomes. In relation to alcohol consumption, Larsson et al. (2020) conducted a Mendelian randomisation study to investigate the association between alcohol consumption and cardiovascular disease. This report demonstrates a causative relationship between higher alcohol consumption and an increased risk of peripheral arterial blockage and stroke. But this report demonstrates a causative relationship between higher alcohol consumption and an increased risk of peripheral arterial blockage and stroke. But, the findings also showed that there is no certain genetically related links between alcohol consumption and cardiovascular disease. Collectively, the results of these studies support the importance of a healthy lifestyle, including avoiding smoking, regular physical activity, a balanced diet and moderate alcohol consumption, in the screening and managing of heart disease.

# 2 Data

The dataset used in this study is the UCI Heart Disease Dataset, a publicly available dataset of 303 observations and 14 variables curated specifically for heart disease research. The dataset is made up of a wide range of medical and demographic-based datas about patients suspected of having heart disease. Before performing the analysis, a few edits were applied on the original dataset. Firstly, I carried out a general check on the dataseta to ensure if the datas are suitable. And it was found that a subset of five observations had insufficient data, making them unsuitable for analysis. These observations were therefore excluded from the dataset. Summary statistics were calculated to provide an overview of the dataset. Descriptive statistics, including measures such as mean, median, standard deviation and range, were calculated for relevant variables associated with lifestyle determinants, heart disease indicators and other relevant covariables. These summary statistics give an indication of the central tendencies, variability and distributions of the data set, aiding subsequent analysis and interpretation of the results.

Table 1: Summary Statistics

| | Mean | Std.Dev | Min | Median | Max |
|---|---|---|---|---|---|
| age | 54.54 | 9.05 | 29.00 | 56.00 | 77.00 |
| ca | 0.68 | 0.94 | 0.00 | 0.00 | 3.00 |
| chol | 247.35 | 52.00 | 126.00 | 243.00 | 564.00 |
| cp | 3.16 | 0.96 | 1.00 | 3.00 | 4.00 |
| exang | 0.33 | 0.47 | 0.00 | 0.00 | 1.00 |
| fbs | 0.14 | 0.35 | 0.00 | 0.00 | 1.00 |
| oldpeak | 8.80 | 11.07 | 0.00 | 4.00 | 62.00 |
| restecg | 1.00 | 0.99 | 0.00 | 1.00 | 2.00 |
| sex | 0.68 | 0.47 | 0.00 | 1.00 | 1.00 |
| target | 0.95 | 1.23 | 0.00 | 0.00 | 4.00 |
| thal | 4.73 | 1.94 | 3.00 | 3.00 | 7.00 |
| thalach | 149.60 | 22.94 | 71.00 | 153.00 | 202.00 |
| trestbps | 131.69 | 17.76 | 94.00 | 130.00 | 200.00 |

# 3 Methods and Data Analysis

Shortly, in this project, the analysis involved several key steps. To explain them briefly, i conducted a data preprocessing to handle missing values and exclude observations with incomplete data. Then, i added summary statistics to describe the dataset that provides an overview of the distribution of variables. Next, i conducted a regression analysis with code chunks to identify the independent affects of each lifestyle factor, while controlling by confounding variables. The regression models helped us interpret the impact of the specific factors on the risk of heart disease. Furthermore, i wanted to support the results with visual representations, including scatter plots, which allowed for a graphical description. In all, the methodology used a mix of statistical analyses and plots to explore the impact of lifestyle factors on heart diseases. With the Figure 1 below, I basically indicated that older adults tend to have higher resting blood pressure, possibly due to age-related changes in blood vessels and lower levels of physical activity. Also, with the support of previous researchs, we can make an interpretation that alcohol has been shown to increase blood pressure acutely, and excessive alcohol drinking can lead to sustained increases in blood pressure over time.

The figure 2 below, suggests that individuals with heart disease (0 = no, 1, 2, 3, 4 = yes and the levels of it) tend to have higher fasting blood sugar levels, specifically represented as if the level is higher than 120 mg/dl, 1 = true; 0 = false, which can be attributed to factors such as smoking and poor diet with other determinants. Especially nicotine can increase blood sugar levels by stimulating the release of stress hormones, which raise blood glucose levels. A diet high in refined carbohydrates, sugary beverages and foods, and processed foods may cause to higher blood sugar levels as well. At last, we can say that these dietary choices can lead to insulin resistance, impaired glucose tolerance, and ultimately higher fasting blood sugar levels.

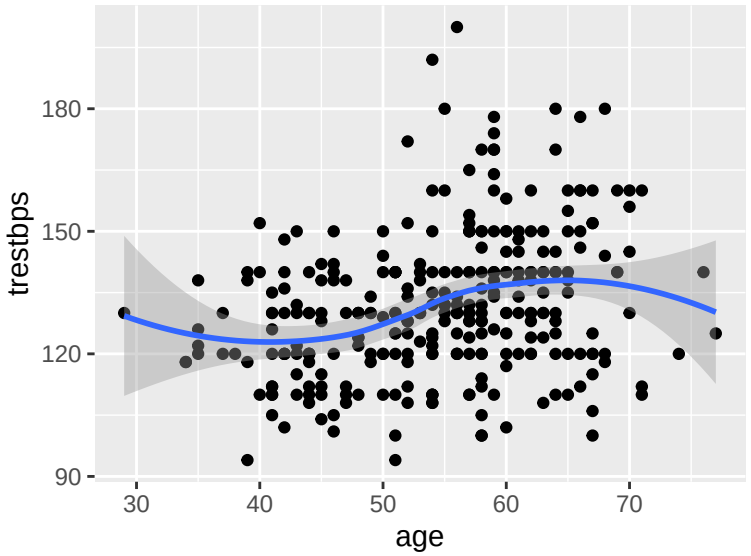


Figure 1: Age and Trestbps

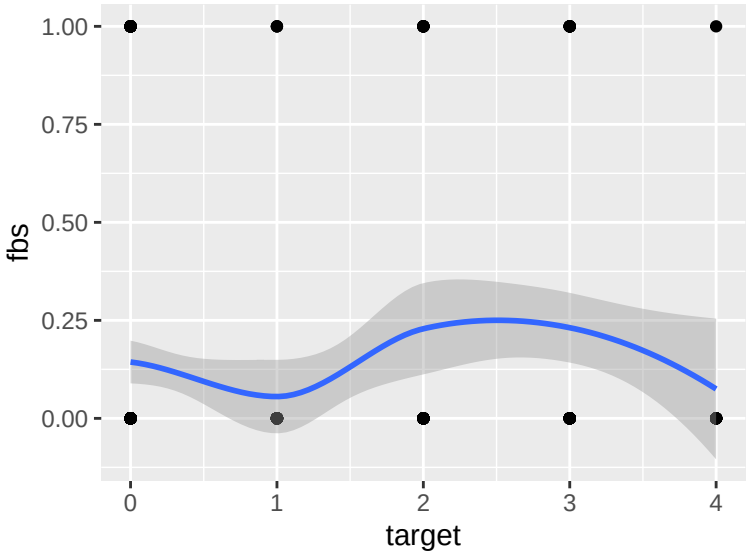


Figure 2: Target and Fbs

# 4 Conclusion

By this project, I wanted to test the impact of lifestyle factors such as physical activity, smoking, alcohol consuming and dietary habits on the progression and development of heart diseases. The analysis of the UCI Repository Heart Disease Dataset gave important informations about the relationship between these issues and their effect on heart diseases. While it provides valuable informations, it is also essential to acknowledge the dataset may have inherent biases and limitations. In addition, this project is focused on a specific set of lifestyle factors, and other variables such as genetics, socioeconomic situation etc. were not included. Because of that, future research should seek to address these gaps and adopt a more professional approach to explore the multifaceted nature of heart diseases. In conclusion, the findings of this project support the importance of avoiding smoking, regular physical activity, a healthy diet and responsible alcohol consuming in reducing the risks of heart disease. Also i believe that with my findings from this research, we can also inform public health initiatives, policy decisions and individualised approaches to cardiovascular health management. Still, further research and collective efforts are needed to develop our understanding and improve targeted interventions to effectively avoid cardiovascular heart diseases.

# 5 References


Bechthold, A., Boeing, H., Schwedhelm, C., Hoffmann, G., Knüppel, S., Iqbal, K., De Henauw, S., Michels, N., Devleesschauwer, B., Schlesinger, S., et al. (2019). Food groups and risk of coronary heart disease, stroke and heart failure: A systematic review and dose-response meta-analysis of prospective studies. *Critical Reviews in Food Science and Nutrition*, *59*(7), 1071–1090.

Carnethon, M. R. (2009). Physical activity and cardiovascular disease: How much is enough? *American Journal of Lifestyle Medicine*, *3*(1_suppl), 44S–49S.

Larsson, S. C., Burgess, S., Mason, A. M., & Michaëlsson, K. (2020). Alcohol consumption and cardiovascular disease: A mendelian randomization study. *Circulation: Genomic and Precision Medicine*, *13*(3), e002814.

Mons, U., Müezzinler, A., Gellert, C., Schöttker, B., Abnet, C. C., Bobak, M., Groot, L. de, Freedman, N. D., Jansen, E., Kee, F., et al. (2015). Impact of smoking and smoking cessation on cardiovascular events and mortality among older adults: Meta-analysis of individual participant data from prospective cohort studies of the CHANCES consortium. *Bmj*, *350*.