Heart of the Matter:
A Deep Dive into Germany's Battle with Cardiovascular Diseases

Vojtěch Sýkora [* 1] **Denis Kovačević** [* 2] **Nam Nguyen The** [* 3]

Abstract

In an era of advanced medicine, health risks persist. This paper provides a comprehensive analysis of the most common health concerns from 1990 to 2019 for Germany comparing them to the rest of the world, with a focus on more developed countries. Utilizing a combination of disease data and data about disease factors, such as healthcare systems and indicators of lifestyle, we show cardiovascular diseases being a major issue, with ischemic heart disease being their largest contributor, and we show their significance using a permutation test. We analyze the difference in lifestyle and healthcare and their effect on the death rate of ischemic heart disease using combination of a visual analysis and a random forest model. In the end, we show the importance of each factor for predicting the death rate. Code is available at https://github.com/sykoravojtech/IHD_germany_2024.

1. Introduction

Germany, one of the largest economies in the world with its advanced healthcare system, faces an intriguing paradox: its life expectancy lags behind other high-income countries. This discrepancy, as highlighted in the analysis by Jasilionis et al. (2023) in "*The underwhelming German life expectancy*," poses critical questions about the underlying factors contributing to this phenomenon.

The underlying reasons are still undefined. This paper continues the work of Jasilionis et al. (2023) focusing on cardiovascular diseases. In 2019, they were the leading cause

---

[*]Equal contribution [1]Matrikelnummer 6636502, vojtech.sykora@student.uni-tuebingen.de, MSc Machine Learning [2]Matrikelnummer 6707752, denis.kovacevic@student.uni-tuebingen.de, MSc Machine Learning [3]Matrikelnummer 6608479, nam.nguyen-the@student.uni-tuebingen.de, MSc Machine Learning.

of death in Germany, accounting for 38% of all deaths. We put a special emphasis on ischemic heart disease, which is the most common cardiovascular disease in Germany.

Firstly, we analyzed cardiovascular diseases in Germany from 1990 to 2019 and compared them to the rest of the world (Section 2.1). One of the biggest obstacles in our analysis was the lack of data that matched our data for diseases in terms of time period and countries (Section 2.2). We presented our model for the death rate of ischemic heart diseases and its interpretation (Section 3). Finally, we discussed our results and limitations of our analysis (Section 3).

2. Data and Methods

There is no single dataset that contains all the information we need. Thus, we had to combine data from multiple sources. In this section, we will describe the data we used and show exploratory analysis of the data. We will also explain the methods we used to analyze the data.

2.1. Cardiovascular Diseases data

The first step in our analysis was to find out whether cardiovascular diseases are an issue in Germany. We used the data from the Global Burden of Disease study (Institute for Health Metrics and Evaluation (IHME), 2020) to compare the incidence and death rate of cardiovascular diseases in Germany to other high-income countries and the world. The data also contains the incidence and death rate for specific age groups which we used to compare the age distribution of cardiovascular diseases in Germany to other countries. In Figure 1 we can see that Germany has very high incidence and death rate of cardiovascular diseases.

We performed a permutation test to check whether the incidence rate in Germany is statistically significant. In the test, we compared the average incidence rate from 1990 to 2019 of Germany to the average incidence rate of the world. With over a million permutations, we found that the p-value is ≈ 0.045. When using the standard threshold of 0.05, this p-value would be judged as statistically significant. In the plot on the right we can also see that the ratio of death rate to incidence rate is much more similar. There appears
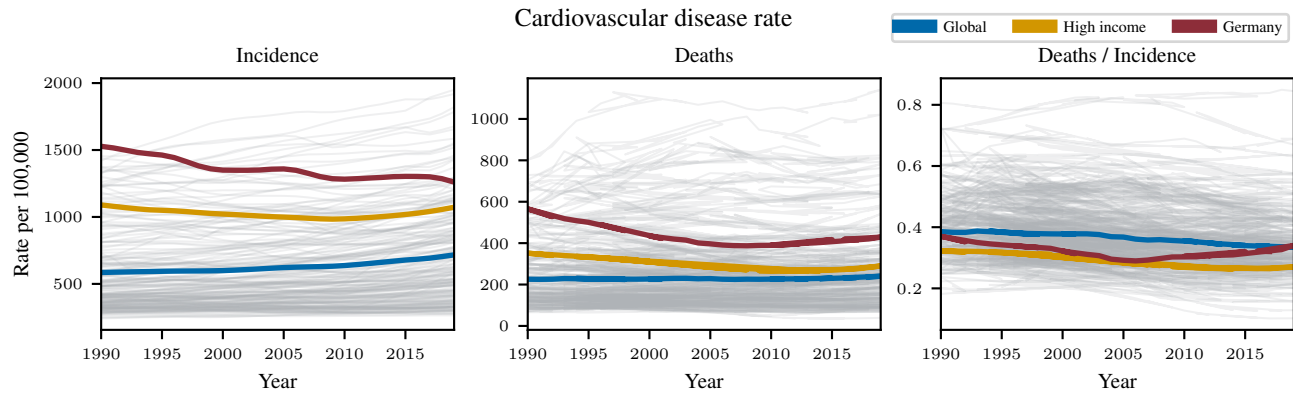
Figure 1. Cardiovascular diseases in the world over time. From left to right: incidence rate, death rate, and the ratio of death rate to incidence rate.

to be no major difference between the global average and the average of high-income countries, which should have a better healthcare system. This suggests that the effect of the quality of healthcare on the ratio is not very significant. In the entire report, we used the same grouping of high-income countries as the Global Burden of Disease study. It is based on gross national income per capita and is available at the world bank website.
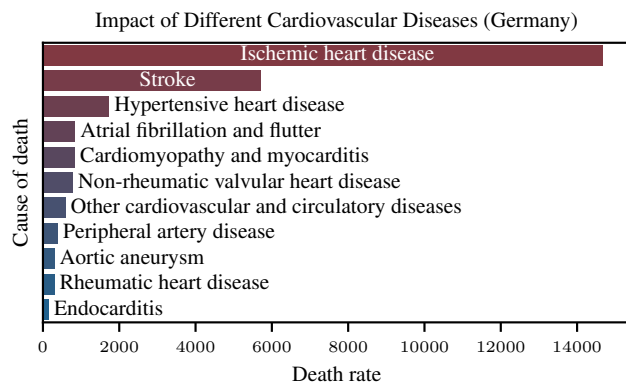

Figure 2. Death rate of different cardiovascular diseases only for Germany. Ischemic heart disease takes up majority of the deaths.

Since cardiovascular diseases are a group of diseases, we wanted to delve deeper and explored the most common cardiovascular diseases in Germany. The top two are ischemic heart disease and stroke (Figure 2). We found out that the incidence rate of ischemic heart disease in Germany is also very high, as shown in its plot. The incidence rate of stroke in Germany is way more similar to the global average, as shown in its plot. This suggests that ischemic heart disease is the main cause of the high incidence rate of cardiovascular diseases in Germany. We also made animations for the

change in disease rate over time. They are available at this link.

2.2. Data sources for factors

Our next step was to use different data sources to find out the underlying factors that have an effect on the incidence rate of ischemic heart disease. We looked at some of the most common risk factors: alcohol consumption and fat consumption. We also included health expenditure as a proxy for the quality of healthcare. Because of the distribution of incidence rate depending on age, as shown in Figure 3, we also included the median age of the population as a factor.
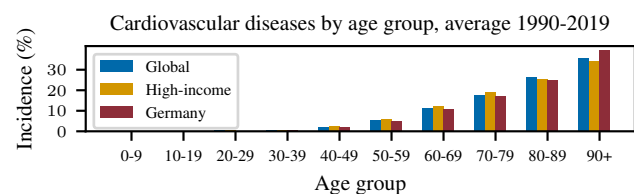

Figure 3. Average incidence rate of cardiovascular diseases for 10 year age groups. The incidence rate is much higher for older age groups, but always the highest in Germany.

The biggest obstacle was the lack of data that is available for all countries over a long time period.

For different factors we had to use different data sources. The data for health expenditure (as USD per capita) and alcohol consumption (liters of pure alcohol per person aged 15 and older) was taken from OECD (Organisation for Economic Co-operation and Development) (2024). The data for population age is from United Nations, World Population Prospects (2022), while the data for fat consumption (grams per person per day) is from Food and Agriculture Organi-
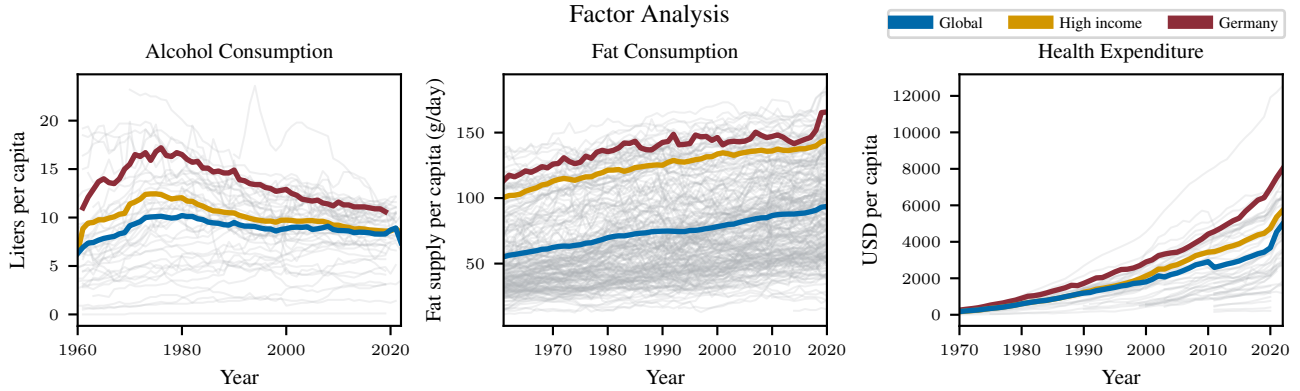
Figure 4. Factor analysis. From left to right: alcohol consumption, fat consumption, and health expenditure.

zation of the United Nations (2020). The data varies in the time period it covers, the number of countries it contains, and the number of missing values, as shown in Table 1. We used the data from 1990 to 2019, which is the time period that is available for the diseases.

Table 1. Overview of the data sources. The missing data is calculated for the time period from 1990 to 2019.

Data source	No. of countries	Missing data
Ischemic heart disease	206	0%
Health expenditure	50	0%
Fat consumption	194	0%
Alcohol consumption	49	0%
Population age	238	0%

For the rest of the project we merged the data sources into one dataset, with the goal of modeling the death rate of ischemic heart disease. Because the data sources did not only have different countries, but also different country names we had to do some data preprocessing. The main part was generating ISO-3 country codes for every dataset. We decided on not averaging the missing values, but instead filtering them out.

The factors we chose not only reflect the well-known factors of cardiovascular diseases, but also reflect Germany's culture. High alcohol consumption mainly in the form of beer and high fat consumption such as in a Bratwurst are what most of the world knows about this nation. But how out of the ordinary is this behaviour in Germany compared to the rest of the world and other high-income countries? Is Germany's culture pushing them towards a quicker death?

In Figure 4 in the leftmost plot we can see alcohol consumption per capita in liters. While it has been on a steady decline since the 1970s, Germany still stays far beyond the global and high-income averages. This decline in alcohol consumption is potentially indicative of successful public

health interventions, policy reforms, and changing social norms towards alcohol usage within the country. The middle plot presents the trends in fat supply per capita in grams per day. All three highlighted lines show that the trend is gradually increasing while keeping Germany on top. The rightmost plot illustrates health expenditure in USD per capita, where all three cohorts show a substantial increase over the five-decade span.

Alcohol consumption and health spending align with the overarching trend that Germany is trying to diminish the burden from cardiovascular diseases; however, fat consumption seems to go against it. Nevertheless, seeing a difference between two lines does not give a solid view on how each factor truly affects the incidence or death rate which is where the results section comes in.

3. Results

To model the death rate of ischemic heart disease we tried using a linear model with different transformations of the data, but the results were not satisfactory. Because of this, we decided to use random forest regression.
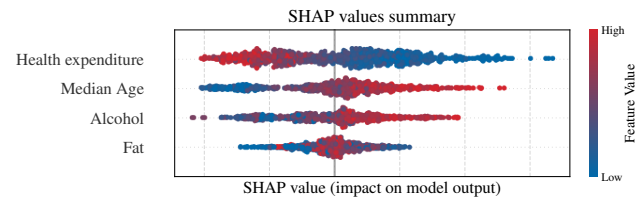

Figure 5. Summary of SHAP values for a predictive model of disease death rate. Each dot's color represents the feature's value, and its position represents the impact on the model's output.

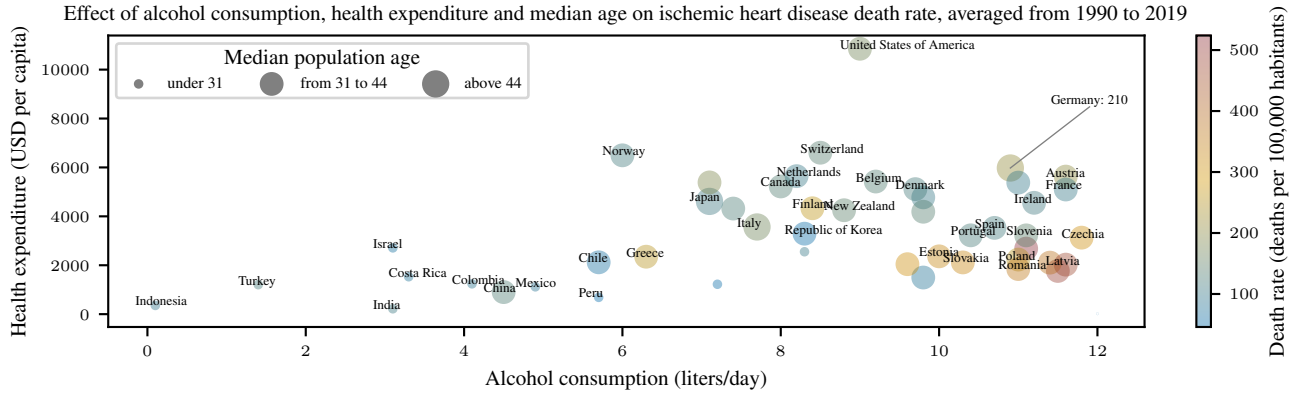We used the `RandomForestRegressor` from the

Figure 6. The combined effect of healthcare spending, alcohol consumption, and median age on the death rate of ischemic heart disease. Special emphasis to the comparison between Germany, high income countries, and the world.

scikit-learn library (Pedregosa et al., 2011) with grid search to find the best parameters. Our best model had the R^2 score of ≈ 0.82 on the test set, while the R^2 score of the linear model was ≈ 0.39. The value represents the proportion of the variance that is explained by the model. We interpret the model using the SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017) (see Figure 5). SHAP values show how the impact of a feature on the model's output.

We can see that the effect of fat consumption is mixed and not very easy to interpret. The effect of the other three features (Healthcare spending, alcohol consumption, and median age) is more clear. For that reason, we decided to use only those three features in our final plot (see Figure 6) in which we show the combined effect of the three features on the death rate. The median age is represented by the size of the bubble, healthcare spending by the y axis, alcohol consumption by the x axis, and the color represents the death rate. We see that countries more towards the right (higher alcohol consumption) have a higher death rate. It is also clear that the countries with the lowest death rate are the ones represented by the smallest bubbles (lowest median age).

4. Discussion & Conclusion

In this project, we analyzed cardiovascular diseases with a specific focus on Germany. We showed that the incidence rate of cardiovascular diseases in Germany is statistically significantly higher than the world average. We analyzed some of the possible factors that could influence the death rate of ischemic heart disease, which is the most common cardiovascular disease in Germany.

Our model showed that the death rate of ischemic heart disease is influenced by healthcare spending, alcohol consumption, and median age, while the effect of fat consumption is mixed and not clear for interpretation. The fact that the effect of fat consumption is not clear is surprising, since it is a well-known risk factor for cardiovascular diseases. We found that lower median age, lower alcohol consumption, and higher healthcare spending all lead to lower death rates, which was to be expected.

Taking into account Germany's high healthcare spending and acceptable death to incidence ratio for cardiovascular diseases, we do not believe that the cause of the problem lies in the healthcare system. It appears that it is actually a significant factor to death rate being lower than expected based on lifestyle. We also found that the German alcohol consumption is higher than the world average (more than two times higher), and that alcohol consumption in general plays a significant role in this issue.

The limitation of our analysis is that we didn't analyze all the possible factors that could influence the death rate of ischemic heart disease. Such factors could be smoking, vegetable consumption, or physical activity. We focused only on cardiovascular diseases, but it would be interesting to analyze less prevalent diseases as well. Another possible extension of this project would be to explore if there is any genetic predisposition for cardiovascular diseases in Germany, using a genome-wide association study (GWAS).

Contribution Statement

This project was done in a group of three. Because of the amount of different data sources, there is no clear division of work. All of us worked on the data collection and analysis. Building on that work, each of us also produced visualizations for their part of the analysis. Writing the report was a joint effort.

References

Food and Agriculture Organization of the United Nations. Daily per capita fat supply, 2020. https://ourworldindata.org/grapher/daily-per-capita-fat-supply, 2020. Accessed: Jan 2024.

Institute for Health Metrics and Evaluation (IHME). Global Burden of Disease Study 2019 (GBD 2019) Data Resources. http://ghdx.healthdata.org/gbd-2019, 2020. Accessed: Jan 2024.

Jasilionis, D., van Raalte, A. A., Klüsener, S., and Grigoriev, P. The underwhelming german life expectancy. *European Journal of Epidemiology*, 38(8): 839–850, 2023. ISSN 1573-7284. doi: 10.1007/s10654-023-00995-5. URL https://doi.org/10.1007/s10654-023-00995-5.

Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4765–4774. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

OECD (Organisation for Economic Co-operation and Development). Alcohol Consumption and Health Spending. Alcohol Consumption: https://data.oecd.org/healthrisk/alcohol-consumption.htm, Health Spending: https://data.oecd.org/healthres/health-spending.htm, 2024. Accessed: Jan 2024.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

United Nations, World Population Prospects. Median age. https://ourworldindata.org/grapher/median-age?tab=table&time=earliest..2020, 2022. Accessed: Jan 2024.