UCLA College | Physical Sciences Statistics

Factors Affecting Mortality in Heart Disease Patients

Alex Clare, Maya Hirsch, Henry Johnson, Lior Levy Meruk and Mary McSweeney

Table of contents

1 Abstract

This study evaluates factors influencing mortality in heart disease patients using data from Faisalabad Hospital in Pakistan. We chose 4 variables to explore: age, cholesterol, diabetes, and gender. Using two sample t-tests, a Chi-square goodness-of-fit test, and a logistic regression model, we explored which factors are significant in predicting mortality in heart disease patients. The results of our tests revealed that all four variables were statistically significant.

2 Background

This dataset is centered around heart disease patients collected from hospital records in Pakistan. Originally the data set consisted of 60 variables and 368 rows, one for each patient. The dataset offers a valuable opportunity to explore certain factors that could influence whether heart disease is lethal or not, which could lead to better risk assessment and treatment plans. Using statistical analyses, this project seeks to reveal insights that can uncover heart disease mortality trends.

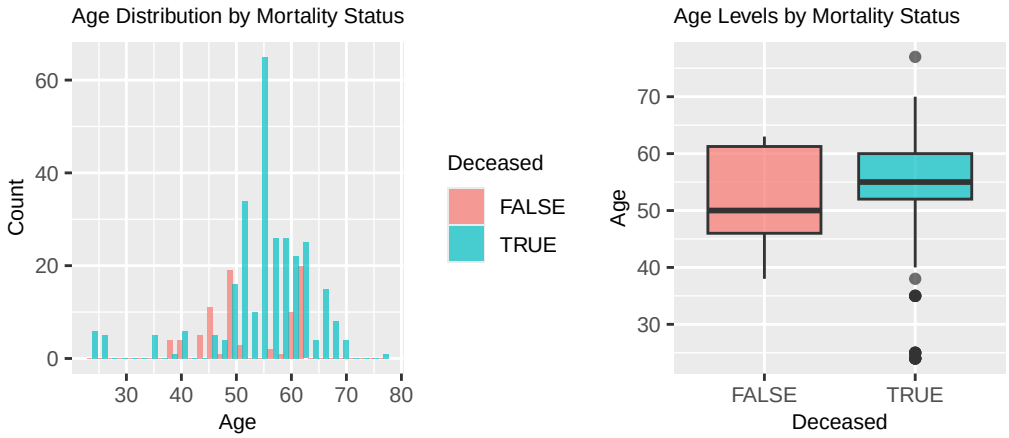
3 Exploratory Data Analysis

We conducted Exploratory Data Analysis (EDA) on the data set focusing on the factors age, cholesterol, diabetes, and gender, which were several key variables we were interested in. The EDA was performed focusing on the primary outcome variable mortality. We hoped to investigate potential patterns and relationships between variables.

3.1 Age vs. Mortality

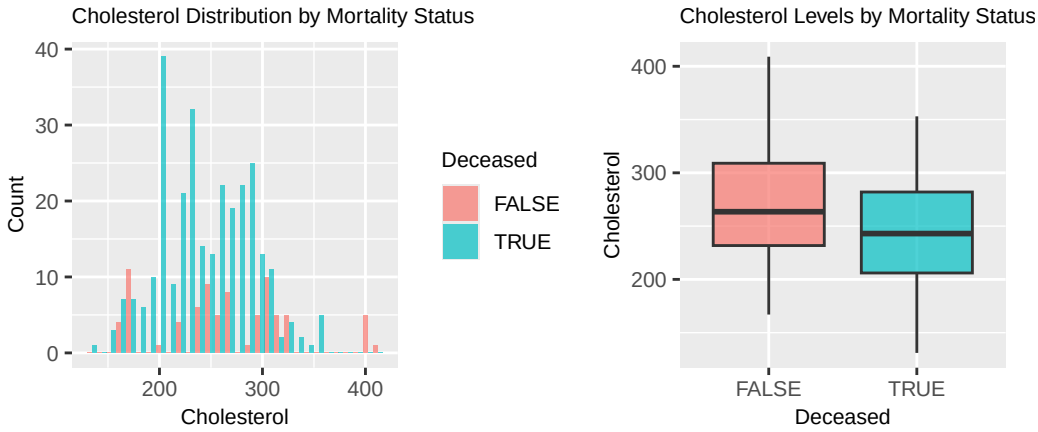
Our EDA revealed that deceased patients were approximately normally distributed throughout different ages, with some outliers around 15 years old. We see the largest amount of deceased patients is concentrated around 55 years old, which happens to be the mean of deceased patients. We chose to look into Age as a factor because this study from Memorial Hermann found a relationship between age and mortality from heart disease patients: "About 82% of people who die of coronary heart disease are 65 or older" ("Heart Disease and Age" 2024).

Patients who were not deceased were less common in general and were concentrated around the 60's and upper 40's. Oddly, non-deceased patients were mostly absent from the 50's age group. The mean of non-deceased patients lies around 50 years old.



3.2 Cholesterol vs. Mortality

The cholesterol by mortality status visual does not seem to follow any particular distribution, deceased patients in our data are relatively right-skewed. In general, it seems that deceased patients have lower levels of cholesterol with the highest concentration of patients occurring at around 200 mg/dl. *The average cholesterol of deceased patients is around 240 mg/dL. Non-deceased patients were roughly uniformly distributed throughout, and their average cholesterol was around 270 mg/dL. High cholesterol levels were found to contribute to mortality from heart disease "Serum total cholesterol and LDL-C level is associated with increased CVD mortality" (Jung et al. 2024)* A total cholesterol level of 240 mg/dL (6.21 mmol/L) or greater is considered high. ("Lipid Panel" 2024)



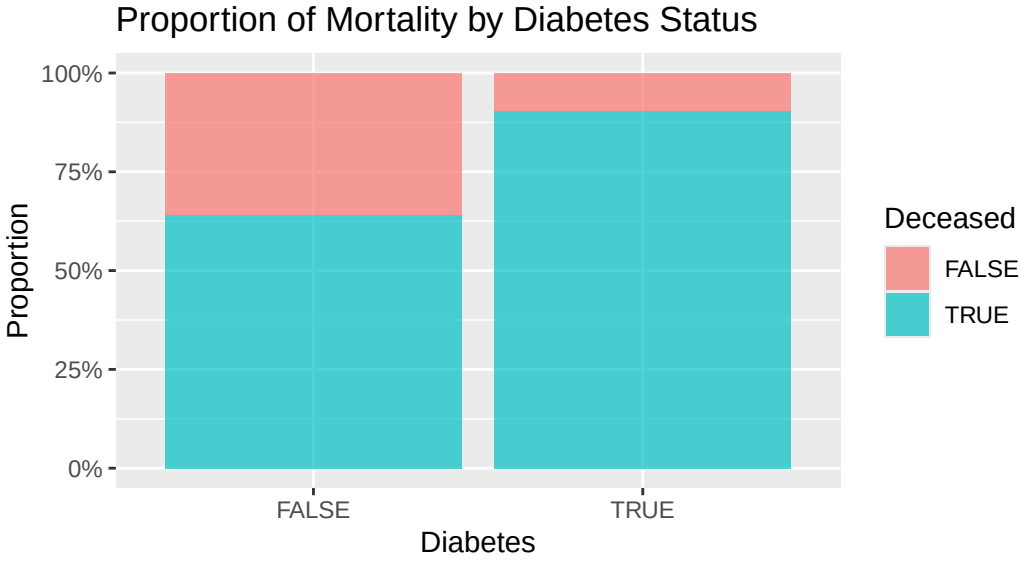
3.3 Diabetes vs. Mortality

Patients with diabetes had a significantly higher mortality rate (approximately 85%) compared to those without diabetes (approximately 62%). We chose to investigate

cholesterol levels in relation to mortality due to a known relationship between patients with diabetes and those who die from heart disease, as this study from the American Diabetes Association found: "Cardiovascular disease (CVD), where the heart and blood vessels are negatively impacted, is the number one cause of death in people living with diabetes, resulting in 2/3 of deaths in people with type 2 diabetes" ("Diabetes Complications - Cardiovascular Disease" 1995-2024) The dataset contained a nearly equal number of patients with and without diabetes.

Diabetes	Count
FALSE	170
TRUE	198

Diabetes	Alive	Deceased
FALSE	61	109
TRUE	19	179

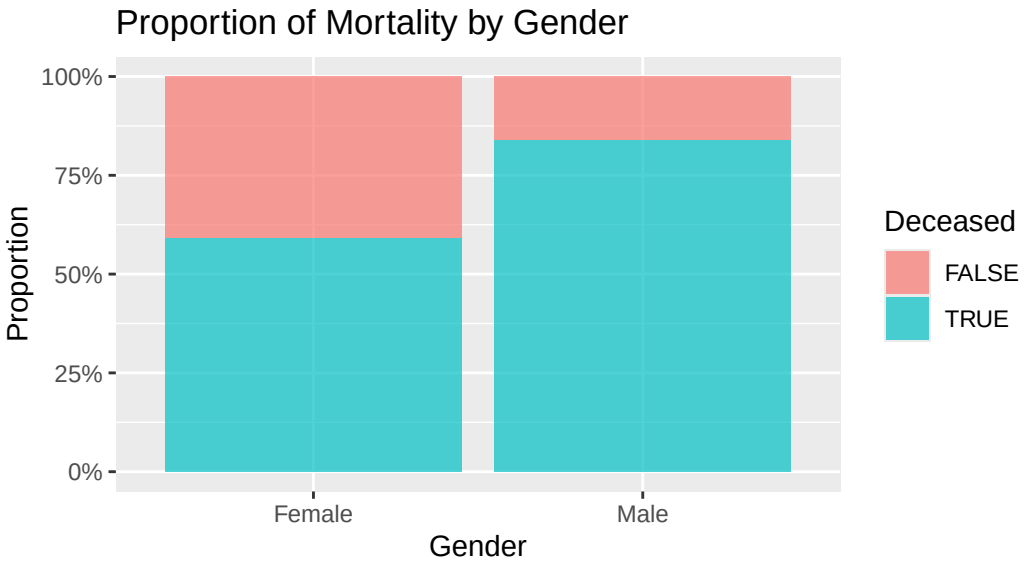


3.4 Gender vs. Mortality

When comparing males and females, male patients appear to have a higher mortality rate (around 80%) than females (around 60%). This means heart disease is more likely to be fatal in males than females. This aligns with this study from the National Institute of Health: "It was observed that men exhibited a higher risk of all-cause and cardiovascular mortality compared to women, within both the general population and the population afflicted by CVD" (Lv et al. 2024). Our dataset had a much higher Male patient count than Female. While we cannot say with certainty why that is, it may affect our results and analysis down the line.

Gender	Count
Female	83
Male	285

Gender	Alive	Deceased
Female	34	49
Male	46	239



3.5 Research Question

We chose to focus on these variables because they had easy interpretability and a fairly good distribution compared to other variables. Additionally, we were able to find outside studies that led us to believe there is an existing relationship between these variables and mortality from heart failure.

Our main research questions are:

- What are the significant factors impacting mortality in heart disease patients?

More specifically:

- Does diabetes have an effect on mortality in heart disease patients?
- Does high cholesterol have an effect on mortality in heart disease patients?
- Does a person's gender have an effect on mortality in heart disease patients?
- Does a person's age have an effect on mortality in heart disease patients?

4 Methods

4.1 Data Cleaning

We found our dataset on Kaggle, a public data science platform, so no data collection was needed. We verified no NA values were present in our data and proceeded to select only a few variables to focus on, both for ease and clarity of our project and analysis. As a last step, to improve clarity during analysis, we transformed the mortality variable (0 - Alive, 1 - Deceased) into a new variable, Deceased (0 - No, 1 - Yes). We also switched the values in the Diabetes variables. Originally, it was 0 - False 1 - True; we switched it to 0 - True 1 - False. We could now move on to test our hypotheses.

4.2 Inferential Tests

Two-sample t-tests were performed to compare the means of continuous variables, such as age and cholesterol levels, between patients who survived and those who did not. This method helps determine if the differences in these variables are statistically significant, giving insight into how they might affect mortality. Significant differences revealed by the t-test, indicate that these continuous variables could play an important role in influencing patient outcomes.

Chi-square tests were used to assess the relationship between categorical variables, such as gender and diabetes status, and patient outcomes (survived or died). This test determines whether the differences between groups are simply due to chance or indicate a meaningful association. Significant results from the chi-square test suggest that these categorical predictors may be connected to patient survival or mortality.

4.3 Logistic Regression

Logistic regression is a widely used statistical method for binary classification. It was employed to analyze how factors such as age, cholesterol levels, diabetes status, and gender influence patient outcomes. This statistical method predicts binary outcomes—in this case, whether a patient survived or died—by estimating the effect of each predictor variable while controlling for the influence of others. Logistic regression calculates an odds ratios, which quantify the strength and direction of the relationship between each predictor and the outcome. This model provides an understanding of how individual factors contribute to heart disease mortality.

5 Results

The chi-squared tests we performed revealed that there was a statistically significant association between gender and diabetes with mortality. Both variables had a p-value below 0.05. These two independent tests tell us that both gender and diabetes are independently related to mortality risk.

The t-tests we ran showed there is a difference in mean cholesterol levels and age between deceased and non-deceased groups. Both variables had a p-value below 0.05. The tests suggested there is a significant association with mortality when unadjusted for other predictors.

Table 1: Hypothesis Test Results

Hypothesis	Test Type	P-Value
Gender is associated with mortality	Chi-square test	2.955121e-06
Diabetes is associated with mortality	Chi-square test	2.398713e-09
Age is associated with mortality	T-test	4.704103e-02
Cholesterol is associated with mortality	T-test	6.93415e-03
All of above are predictors of mortality	Logistic regression	All predictors $p < 0.01$

We then utilized a logistic regression model to assess the independent effects of our variables on mortality. All four predictors remain significant with p-values below 0.05 in the adjusted model. Suggesting that these variables are important in contributing to increased risk of mortality.

The odds ratios indicate the strength of association for each predictor with mortality. The 95% Confidence intervals provide a range of plausible values for the predictors. Confidence intervals that exclude 1 are statistically significant. For cholesterol, a 1-unit increase in cholesterol (mg/dL) is linked to a 0.7% reduction in the odds of mortality. This means that according to our results, a lower cholesterol level leads to a lower risk of mortality from heart disease. For gender, males face a 3.35 times higher likelihood of mortality compared to females, this points to gender being a crucial factor in mortality. For Diabetes, having diabetes increases the odds of mortality by 5.52 times relative to non-diabetic individuals. For age, each additional year of age raises the odds of mortality by 9.7%, reflecting the association between aging and mortality.

Table 2: Odds Ratios and Confidence Intervals

Predictor	Odds Ratio	95% CI Lower	95% CI Upper	P-value
Cholesterol	0.9922668	0.9867026	0.9976216	0.0054860
Gender (Male)	3.3452444	1.6187533	7.0247149	0.0012268
Diabetes (True)	5.5156324	2.8232324	11.2760525	0.0000012
Age	1.0965299	1.0583131	1.1381847	0.0000006

To test for multicollinearity we calculated the Variance Inflation Factor (VIF) for all predictors. The results show that the VIF values for all predictors are below 5 which means each predictor contributes unique information about mortality in heart disease patients.

Table 3: VIF Values

	Cholesterol	Gender	Diabetes	Age
VIF	1.061287	1.600561	1.299088	1.645194

6 Limitations

Although we are confident in our methods and results, this study did not come without its limitations. First, we are limited to the scope of our dataset. Since our data came from a Pakistani hospital, we can only generalize our results to future Pakistani heart disease patients, not necessarily to the world population. Furthermore, we noticed while exploring the dataset that there were certain biases within the data. For example, 95% of patients in the dataset have been diagnosed with depression. We do not have any explanation for this bias. We also noticed that there were significantly more men in the dataset than women, which we can attribute to the fact that men are more likely to have heart disease than women (Lv et al. 2024). This means that we can not make conclusions about gender as confidently. Finally, while we did check for multicollinearity, we did not perform extensive tests on the interaction effects of the predictors.

7 Conclusion

The goal of this study was to see what factors impact the chance of mortality for heart disease patients. Through statistical tests, we found that age, cholesterol, diabetes, and gender are all significantly associated with mortality in heart disease patients. These results align with previous studies on heart disease, with the notable exception of cholesterol, which we found has a negative association with mortality. We would be interested in looking further into this variable in future studies, as this was our most interesting result.

References

"Diabetes Complications - Cardiovascular Disease." 1995-2024. American Diabetes Association. https://diabetes.org/about-diabetes/complications/cardiovascular-

disease#:~:text=Cardiovascular%20disease%20(CVD)%2C%20where,people%20with%20type%202%20diabetes.

"Heart Disease and Age." 2024. Memorial Hermann. https://memorialhermann.org/services/specialties/heart-and-vascular/healthy-living/education/heart-disease-and-age#:~:text=Some%20risk%20factors%20may%20be%20out%20of%20your%20control%3A&text=About%2082%25%20of%20people%20who,disease%20are%2065%20or%20older.

Jung, E., S. Y. Kong, Y. S. Ro, H. H. Ryu, and S. D. Shin. 2024. "Serum Cholesterol Levels and Risk of Cardiovascular Death: A Systematic Review and a Dose-Response Meta-Analysis of Prospective Cohort Studies." *MDPI (Multidisciplinary Digital Publishing Institute)*. https://pmc.ncbi.nlm.nih.gov/articles/PMC9316578/.

"Lipid Panel." 2024. John Hopkins Medicine. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lipid-panel.

Lv, Y., X. Cao, K. Yu, J. Pu, Z. Tang, N. Wei, J. Wang, F. Liu, and S. Li. 2024. "Gender Differences in All-Cause and Cardiovascular Mortality Among US Adults: From NHANES 2005-2018." *Frontiers in Cardiovascular Medicine* 11: 1283132. https://doi.org/10.3389/fcvm.2024.1283132.