|  |
| --- |
|  |
| Following the pain  A correlation analysis between chest pain and cardiovascular disease in hospitalized adults |
| |  |  |  | | --- | --- | --- | | Malcolm Smith Fraser | 11/14/20 | IDS 702 – Modeling and Representation of Data | |

Contents

[Summary 1](#_Toc56271007)

[Introduction 1](#_Toc56271008)

[Data 1](#_Toc56271009)

[Model 1](#_Toc56271010)

[Conclusions 1](#_Toc56271011)

# Summary

A few sentences describing the inferential question(s), the method used and the most important results.

This analysis will quantify the scale to which chest pain, exercise induced, and non-exercise induced, is an indicator of more serious underlying cardiovascular pathologies. It will also explore what other predictors, when looked at in conjunction with heart disease, improve increase the correlation between the chest pain a patient reports and a true diagnosis of cardiovascular disease.

The regression uses patient data from the Cleveland clinic foundation. The response variable, a patient’s angiographic heart disease status, is binary – referring to <50% arterial narrowing and >50% arterial narrowing. As such a logistic regression is used as our predictive model.

# Introduction

A more in-depth introduction to the inferential question(s) of interest.

For years now, cardiovascular diseases have been recognized as the leading causes of death both within the United States and globally. While there have been significant strides made towards developing metrics of identifying the progression of heart disease in individuals, the only predictor that is immediately recognized by a patient, without any tests, is their chest pain.

In exploring the correlation between chest pain and cardiovascular disease, I hope that an optimal course of action can be established for patients who feel experience chest pain. Also, to understand of the presence of chest pain

The goal of this analysis is not to find a single model that best predicts a patient’s heart disease status, rather it’s to determine what predictors, when observed in conjunction with chest pain are significant indicators of heart disease.

# Data

You should describe the data in this section: how you obtained the data, the variables included, dealing with missing/erroneous values, exploratory data analysis etc.

# Model

A detailed description of the model used, how you selected the model, how you selected the variables, model assessment, model validation, and presentation of the model results. What are your overall conclusions in context of the inferential problem(s)?

# Conclusions

In this section, you should present the importance of your findings, and describe any limitations of the study. You can also address future work here if there are extensions of your analysis you find interesting, especially those that may address some of the limitations already mentioned.

When interpreting the results, the primary metric of interest is sensitivity. Sensitivity is the proportion of true positives to the total number of individuals with heart disease. Chest pain is only an important metric in the early stages and as such we want to maximize the correlation between chest pain and heart disease. At these early stages we also care less about the potential for false positives as the next steps would merely be a more rigorous test as opposed to a treatment that could put healthy patients at risk.

**Is chest pain an accurate signal of heart disease? What if it is exercise induced?**

The proportion of patients with heart disease is .46

Chest pain alone achieves a sensitivity of 0.75, a specificity of 0.75, and an accuracy of 0.75. Exercise induced chest pain alone has a significantly lower sensitivity of 0.54 which tells us that exercise induced chest pain is not as closely correlated to cardiovascular disease as regular chest pain.

When looking at the model with both exercise-induced chest pain and non-exercised-induced chest pain, along with the interaction between those two predictors, the model performs identical to the model with chest pain alone.

Examining the model shows that every level of chest pain is significant at a significance level of 0. When compared to the chest pain baseline of asymptomatic, the odds of a patient with typical angina having a positive heart disease diagnosis is 83.43% lower, for patients with atypical angina it is 91.48% lower, and for patients with non-anginal pain it is 89.51% lower. These results are very counter intuitive as they signal that chest pain among these patients is less correlated with cardiovascular disease than having no chest pain at all (asymptomatic).

**Is the persistence of chest pain more significant if a patient has other confounding factors?**

Many of the other predictors have no effect of the models predictive ability, however, thalach – maximum (hear rate achieved), oldpeak (a measure of ST interval depression during exercise), ca (number of major blood vessels), and thal(thallium stress test results) all have an effect of the predicted outcomes when modeled with chest pain. Since the overall goal of this analysis is to find a predictors that, along with chest pain, can best identify heart disease I will only be interpreting the models that saw changes in the predictive metrics.

When thalach is included in the model with chest pain we see an improved sensitivity of 0.77, a specificity of 0.78, and an accuracy of 0.77. Thus, in terms of predictive power, heart rate and chest pain are combine for a model generate predictions that better distinguish between diseased and non-diseased patients. All predictors are significant at a significance level of 0. When compared the baseline of an asymptimatic patient with average thalach, the odds of a patient experiencing typical angina having a positive heart disease diagnosis is 77.53% lower, for patients experiencing atypical angina 85.93% lower, and for patients with non-anginal pain 86.22% lower. Also, for every one-unit increase in thalach, the odds of a patient having a positive diagnosis decreases by 3.29%. While these trends are consistent with what was seen when modeling chest pain alone, the magnitude of the effect that the various forms of chest pain have does appear to have decreased at every level. Modeling with thalach makes the chest pain predictors weaker in terms of how much they decrease the odds of having a heart disease diagnosis, and thus not more significant of predictors.

When oldpeak and its interaction with chest pain is included in the model, we see a decrease in sensitivity to 0.65, a specificity improvement of 0.83, and a similar accuracy accuracy of 0.75. In terms of predictive power, this model performs is worse than the baseline model at identifying chest pain since our primary metric is specificity. However, in terms of overall fit to the data the model does have an improved the deviance by 42.66. As in previous models, the coefficients for chest pain are all significant and signal similar trends. When compared to the baseline of an asymptomatic patient with average oldpeak, a one unit increase in oldpeak is correlated with an increased odds of heart disease diagnosis of 175%, however, given that the patient is experiencing typical angina that increase in diagnosis odds is only 4.35%. As with thalach, modeling with oldpeak also decreases the magnitude of the effect that the chest pain predictors have on the odds of a heart disease diagnosis.

When ca is included in the model with chest pain we see an improved sensitivity of 0.80, a specificity of 0.71, and an accuracy of 0.76. Thus, in terms of predictive power, heart rate and chest pain are combine for a model generate predictions that better distinguish between diseased and non-diseased patients. All predictors are significant at a significance level of 0. When compared the baseline of an asymptimatic patient with average thalach, the odds of a patient experiencing typical angina having a positive heart disease diagnosis is 77.53% lower, for patients experiencing atypical angina 85.93% lower, and for patients with non-anginal pain 86.22% lower. Also, for every one-unit increase in thalach, the odds of a patient having a positive diagnosis decreases by 3.29%. While these trends are consistent with what was seen when modeling chest pain alone, the magnitude of the effect that the various forms of chest pain have does appear to have decreased at every level. Modeling with thalach makes the chest pain predictors weaker in terms of how much they decrease the odds of having a heart disease diagnosis, and thus not more significant of predictors.

**How do these effects compare for both men and women?**

Exploring how sex effects the correlation between chest pain and cardiovascular disease does not find any significant results. It was found that the proportion of asymptomatic women diagnosed with heart disease is much higher than asymptomatic women. This, however, is not surprising. Chest pain in women is often not a heart attack symptom and it reasonable to assume that this correlation would be the same for signaling the progression of heart disease.