The Framingham Heart Study

Muhammad Affan Mansoor Malik

Marquette University

COSC 5610: Data Mining

10/08/2020

Introduction

Cardiovascular disease is the most common cause of mortality in developed countries. It is considered to be one of the serious illnesses and is the leading cause of death for men, women, and people of most racial and ethnic groups in the United States [1][4]. To understand the common factors that contribute to the development of CVD, "Framingham Heart Study" was initiated and on October 11th, 2020, this study will celebrate 72 years since the examination of its first participant in 1948 [2]. During this period, the study has provided substantial insight into the epidemiology of CVD and its risk factors. Said factors, if diagnosed early can aid in making crucial decisions and change high-risk patients' lifestyle as well as reduce complications and help in prevention [3]. This research will aid in finding out high risk factors and most relevant factors of a heart disease, as well as predict the overall risk using Logistic regression.

About the Dataset

The dataset is from a CVD study on the residents of Framingham, MA, which is publicly available on Kaggle. This dataset provides information of the patient, includes demographic, behavioral, medical history, and risk factors from the first physical examination of the patient. Each attribute is a potential risk factor. The dataset is a .csv file which has over 4000 records and 15 attributes populated.
It can be accessed through this link: https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset

Below is an overview of the dataset and its variables:

Demographic:

sex: male or female.

age: age of the patient.

education: level of education the patient has.

Behavioral:

currentSmoker: whether the patient is a current smoker.

cigsPerDay: the number of cigarettes that the person smoked on average in one day.

Medical:

BPMeds: whether the patient was on blood pressure medication.

prevalentStroke: whether the patient had previously had a stroke.

prevalentHyp: whether the patient was hypertensive.

diabetes: whether the patient had diabetes.

totChol: total cholesterol level.

sysBP: systolic blood pressure.

diaBP: diastolic blood pressure.

BMI: Body Mass Index.

heartRate: heart rate.

glucose: glucose level.

Predict variable (desired target):

10-year risk of coronary heart disease CHD

Research Question & Goal

Will a patient have a 10-year risk of developing a cardiovascular disease? The goal is to predict whether a patient has 10-year risk of future coronary heart disease (CHD).

Conclusions

Based on the variables there are a few questions that come to mind, such as, if demographic variables come into play? Who is more at risk, male or female? Or if behavioral variables like current smoker and cigarettes per day affect the outcome? Or what if the patient is not a current smoker but had a stroke in the past, are they in risk of future CHD? There are medical variables such as heart rate, blood pressure, diabetes which bring up question that could diabetes be a high-risk factor for heart rate failure [5]? Whether or not all these variables combined form an outcome much higher than their individual counterparts or are they dependent on each other.

It is too early to predict based on the given dataset & its variables; some questions & analysis/output might change in the future of this research as more information is revealed.

References

[1] Centers for Disease Control and Prevention. Underlying Cause of Death, 1999–2018. CDC WONDER Online Database. Atlanta, GA: Centers for Disease Control and Prevention; 2018. Accessed March 12, 2020. https://www.cdc.gov/heartdisease/facts.htm

[2] Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999-1008. doi:10.1016/S0140-6736(13)61752-3

[3] Buttar HS, Li T, Ravi N. Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation. Exp Clin Cardiol. 2005;10(4):229-249.

[4] Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2013 update: a report from the American Heart Association. Circulation. 2013;127:e6–e245.

[5] Rosano GM, Vitale C, Seferovic P. Heart Failure in Patients with Diabetes Mellitus. Card Fail Rev. 2017;3(1):52-55. doi:10.15420/cfr.2016:20:2