Adam Lang

9/13/20

IBM Data Science-Final Project

**Introduction/Business Problem**

Cardiovascular Disease (CVD) remains the leading cause of morbidity and mortality despite 4 decades of declining mortality rates in the United States. CVD remains the leading cause of morbidity and mortality for women in the United States (Cho et al. 2020). The age-adjusted death rate attributable to CVD, based on 2017 data, is 219.4 per 100,000. On average, someone dies of CVD every 37 seconds in the U.S. There are 2,353 deaths from CVD each day, based on 2017 data (AHA, 2020). Around 17.5 million people die each year from cardiovascular diseases (CVDs), an estimated 31% of all deaths worldwide. This statistic is expected to grow to more than 23.6 million by 2030 (Hajar, 2016).

CVDs are a group of disorders of the heart and blood vessels, and they include coronary heart disease, cerebrovascular disease, peripheral arterial disease, and rheumatic heart disease, congenital heart disease, deep vein thrombosis, and pulmonary embolism. The cause of CVD is usually the presence of a combination of risk factors, such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol, hypertension, diabetes, and hyperlipidemia (Hajar, 2016). Collectively, CVDs and their risk factors are a gateway to other diseases. It is therefore vital to understand and research CVD risk factors and how they predict the development of heart disease.

Risk Factors for Heart disease are broken down into “Modifiable” and “Nonmodifiable”. They are as follows:

**Modifiable**

* **Smoking or exposure to environmental tobacco smoke:** The most preventable risk factor. Smokers have more than twice the risk of developing cardiovascular disease. On average, smoking costs 13 years of life to a male smoker and 14 years to a female smoker. Exposure to smoke — secondhand smoking — increases the risk even for non-smokers.
* **Obesity:** One of the highest risk factors for CVD. Obesity is defined as having a body mass index (BMI) ≥30 kg/m2. Obesity is often subdivided into classes (Class I: BMI=30.0–34.9, Class II: BMI=35.0–39.9, Class III: BMI ≥40.0) to further stratify health risk. Using BMI-based diagnostic criteria, 39.8% of the US population meets the definition of obesity with 7.7% having class III obesity or severe obesity, defined as a BMI≥40.0 kg/m2 Obesity is a strong independent predictor of CVD even in the absence of other risk factors, however, interestingly after onset of CVD the relationship between higher BMI and clinical outcomes is not linear. BMI and obesity should however be carefully considered with respect to an individuals’ amount of lean mass and fat mass (Carbone et al. 2019).
* **Sedentary lifestyle (not enough physical activity):** Sedentary behavior and physical inactivity are among the leading modifiable risk factors worldwide for CVD and all-cause mortality. Although the American Heart Association, the American College of Cardiology, and the American College of Sports Medicine, among other leading organizations, have emphasized that sedentary behavior (SB) and physical inactivity (PI) are major modifiable cardiovascular disease (CVD) risk factors, a sizable percentage of the United States and worldwide population still present with high levels of SB/PI and low levels of physical activity (PA). Recently, a major emphasis has been directed at making health promotion a priority, including the promotion of PA and exercise training (ET) and improving levels of cardiorespiratory fitness (CRF) in the United States and worldwide in efforts to prevent chronic diseases, especially CVD (Lavie et al. 2019).
* **Diabetes:** Whether Type 1 or Type 2 the risk for CVD is the same. A fasting plasma glucose above 125mg/dL is at risk for Diabetes. Diabetes and CVD share similar risk factors – high cholesterol, hypertension, and obesity.
* **High cholesterol or abnormal blood lipids (fats):** Controlling levels of LDL (bad cholesterol), HDL (good cholesterol), total cholesterol and triglycerides (most common body fat) will reduce risk of CVD. Total cholesterol should be less than 200mg/dL. Triglycerides should be less than 150mg/dL. HDL levels above 60mg/dL and LDL levels below 70mg/dL are ideal to prevent CVD.
* **Hypertension (high blood pressure):** For persons older than 50, systolic blood pressure is more important than diastolic blood pressure as a cardiovascular disease risk factor. Starting at 115/75 mmHg, cardiovascular disease risk doubles with each increment of 20/10 mmHg throughout the blood pressure range. Normal blood pressure is 120/80 and for each increment of systolic pressure over 20 and diastolic pressure over 10 it increases your stage of developing hypertension.

**Nonmodifiable**

* **Gender:** Women tend to develop CVD 10 years later in life than Men, but outcome is often worse.
* **Age older than 50 years:** 80% of people who die from CVD are 65 years or older.
* **Family history of heart disease:** A primary risk is a relative who developed heart disease before age 55.

One controversial risk factor not discussed above is alcohol intake. Some studies have shown that moderate alcohol intake can reduce your risk of developing and even dying from CVD. There is some evidence that moderate alcohol intake will increase good levels of HDL cholesterol. It is important to understand that “moderate drinking” is defined as an average of one drink per day for women and one or two for men. A drink might be less than you think: 12 ounces of beer, 4 ounces of wine or 1.5 ounces of 80-proof spirits. It is however when excessive alcohol intake can lead to hypertension, heart failure or even stroke. Excessive drinking can also lead to cardiomyopathy a dangerous disorder that enlarges the heart muscle leading to many issues among the CVDs (Johns Hopkins, 2020).

Heart disease continues to be studied worldwide. In the United States our understanding of heart disease is due to the work of the Framingham Heat Study (FHS). The study began in 1948 with 5209 adult subjects from Framingham and is now on its third generation of participants. Much of our appreciation of the pathophysiology of heart disease came from the results of studies from the FHS. It established the traditional risk factors, such as high blood pressure, diabetes, and cigarette smoking for coronary heart disease. Framingham also spearheaded the study of chronic noninfectious diseases in the USA and introduced preventive medicine.

It is obvious that CVD is not only a major cause of morbidity and mortality in our society but influences the development of other acute and chronic diseases and how we continue to evolve our approach to preventative medicine and lifestyle improvements. It is therefore prudent to continue to work with cardiovascular risk factor data and try to predict what leads to CVD developing as risk factors and people evolve. The goal of this project is to perform exploratory data analysis on cardiovascular risk factor data and use machine learning algorithms to try and predict which risk factors lead to the development of heart disease.

**References**

AHA (2020) Heart Disease and Stroke Statistics – 2020 Update. Retrieved from: <https://professional.heart.org>.

Carbone et al. (2019) Obesity paradox in cardiovascular disease: where do we stand? Vasc Health Risk Manag. Vol. 15:89-100.

Cho et al. (2020) Summary of Updated Recommendations for Primary Prevention of Cardiovascular Disease in Women. JACC State of the Art Review. Vol. 75(20): DOI: 10.1016/j.jacc.2020.03.060

Hajar, R. (2016) Framingham Contribution to Cardiovascular Disease. Heart Views. Vol. 17(2):78-81.

Johns Hopkins (2020) Alcohol and Heart Health: Separating Fact from Fiction. Retrieved from: hopkinsmedicine.org.

Lavie et al. (2019) Sedentary Behavior, Exercise, and Cardiovascular Health. Circulation Research. Vol. 124(5). Retrieved from: <https://ahajournals.org>.

UCSF Health (2020) Understanding Your Risk for Heart Disease. Retrieved from: ucsfhealth.org.

Torpy et al. (2003) Risk Factors for Heart Disease. JAMA. Vol. 290(7):980.