Exploratory data analysis of the BRFSS 2015 Heart Disease Health Indicators Dataset

Afonso D. Carreira[1], Marta F. Carvalho[1], Rita S. Marques[1], Tomás V. Geraldes[1]

1 – Department of Medical Sciences, University of Aveiro; Campus Universitário de Santiago, Agra do Crasto, Edifício 30, 3810-193, Aveiro

Abstract

Background Cardiovascular diseases (CVD) remain the leading cause of death globally, with incidence significantly influenced by both lifestyle behaviours and socioeconomic disparities. This study aims to explore the association between CVD risk and various risk factors within the U.S. population, with an emphasis on the indirect risk factors of income and education, utilising data from the 2015 Behavioural Risk Factor Surveillance System (BRFSS).

Methods The analysis and processing of the data were conducted using *RStudio*, with a cross-sectional analysis conducted on a cohort of 253,680 respondents, assessing factors such as a history of heart disease, hypertension, cholesterol levels, physical activity, and alcohol consumption, alongside demographic variables including age, sex, income, and educational attainment. Logistic regression models were employed to estimate the probability of CVD based on these determinants.

Results The findings revealed that CVD prevalence increases with age, and that lower income and education levels are significantly associated with elevated CVD risk, prone to the effects of unequal access to healthcare and differences in health-related behaviours. Moreover, the study acknowledged gender-specific risk patterns, with men showing a higher risk in early adulthood, and women exhibiting increased susceptibility post-menopause. Physical inactivity, smoking, and hypertension were recognised as key modifiable risk factors.

Conclusion The study underscores the critical role of socioeconomic and lifestyle factors in influencing CVD risk, emphasising the need for targeted public health interventions, particularly among disadvantaged populations, to mitigate these modifiable risks.

Keywords: Cardiovascular Diseases, Risk Factors, Socioeconomic Status, Education Status

1. Introduction

1.1. Cardiovascular Disease

Cardiovascular disease (CVD) is a broad term that includes various conditions affecting the heart and blood vessels, such as high blood pressure, atherosclerosis, heart failure, strokes, arrhythmias, valvular heart disease, among others [1]. According to World Health Organization (WHO), CVDs are the leading cause of death worldwide and approximately 17.9 million people die from them each year.

1.2. Risk factors

Factors that can increase the risk of heart disease are known as risk factors and can be associated with aspects such as ageing-related structural and functional abnormalities in the heart and blood vessels, sex, and genetic heredity. These are such examples of uncontrollable risk factors, meaning that the prevention of the disease cannot be controlled or influenced in any way. However, some of the risk factors are controllable by the subject, such as the person's diet, smoking habits, stress, sedentary lifestyle and obesity, alongside many others [2]: when done correctly, the controlling of these risk factors translates into a reduction of the chances of contracting CVDs. Educational level and income are two others risk factors that may be related to heart disease: however, these risk factors are more difficult to categorize into modifiable and non-modifiable.

1.3. Behavioural Risk Factor Surveillance System

The Behavioural Risk Factor Surveillance System (BRFSS) is an American system of health-related surveys. The annual BRFSS questionnaires collect state data about U.S. residents about their health-related risk behaviours [3]. Their main objective is to assess, based on citizens' responses, the risk of developing heart disease, taking into account factors such as age, a non-modifiable risk factor, metabolism, which slows down over time, and the thickening of both blood vessels and arteries, which can lead to the development of hypertension – a major indicator of CVD risk [2].

There are three types of factors that are incorporated in this questionnaire: a) risk factors, previously mentioned, that identify features that are notable risk factors associated with heart disease; b) health behaviours, which comprehend factors such as physical activity and fruit and vegetable consumption; c) demographic factors, such as age, gender, education and income, and others, which can be pointed out as the primary source of diversity in the questionnaire answers.

Two indicators of heart disease risk also present in the BRFSS questionnaire data are education and income. It is essential to evaluate not only direct risk factors, such as hypertension, smoking, and cholesterol, but also indirect factors, which are often overlooked. For instance, socioeconomic status is expected to play a significant role in CVD prevention, especially since the U.S. healthcare system is an extremely expensive service, with many people relying on health insurance, almost as equally expensive, to manage costs [4]. Education level can also play a key role in preventing heart disease, as it provides greater awareness of the prevalence of these diseases and the associated risk factors.

However, it is important to underline that BRFSS relies on information reported directly by the respondent, so it may be subject to several sources of possible error. How questions are worded may elicit responses in a certain way and can result in what is called "measurement error". Similarly, the ability to accurately recall details varies by person and how much time has passed since the event they are trying to recall, which leads to the so-called "response error". It is also possible that the people who choose to take part are different than those who do not. Interviews are conducted only in English and Spanish in Washington State, so adults who are not able to be interviewed in English or Spanish are not included. Households without telephones are not contacted. Thus, BRFSS findings can only be generalized to English and Spanish speaking adults living in households with telephones, which translates to a “selection bias”. There is no reason to believe that these sources of bias change significantly from year to year: therefore, even if the results are not completely accurate, a comparison of such over

time is possible, which can allow for the determination of the increase or decrease of a given condition.

1.4. Objectives of this research project

Heart disease is the leading cause of death in the United States [5]. Therefore, a thorough study of the BRFSS survey answers is key to understanding patterns involved in the development of cardiovascular diseases in the sample at study, with this study focusing mainly on the influence of indirect risk factors age, sex, economic level/income, and education on cardiovascular disease, with a slight focus on other factors such as high blood pressure and cholesterol levels, diabetes, and others, all of which are considered direct risk factors and therefore have a larger contribution to cardiovascular disease. In addition to this, a statistical analysis will be conducted to evaluate correlation and/or significant association between variables, with the goal of possibly predicting instances of heart diseases and/or developing a predictive model of cardiovascular disease according to risk factors, health behaviours or demographic factors.

2. Materials and Methods

As previously stated, all data shown in this report is provided from the 2015 BRFSS questionnaire [6], and all data processing was done using the software *RStudio* (version 2024.04.2, *R* version 4.4.1).

All information about the questions posed to the participants in the questionnaire can be found in **Annex I**. The data was supplied in comma-separated values (CSV) form, which was then imported onto *R* and set as a data frame. Viewing the imported data in more detail, all columns of the data frame are of the numeric type, with the greater content of such columns being presented as simple “Yes” or “No” answers converted into numerical categories and being classified as binary variables.

2.1. Data set loading and understanding

Data treatment began as we set every numeric column that needed its respective treatment as a column of the factor type, with as many levels as groups implemented by the research peers, to guarantee full comprehension of the data frame. Having done all data reformulation, it is taken that the dataset holds 253,680 records and 22 columns, which correspond to the sample size and the answers to the questions posed to each participant.

2.2. Exploratory Data Analysis

Firstly, a broader data analysis was taken by the usage of exploratory data analysis techniques, to infer and to help understand the data distribution, patterns and correlations, or any type of anomalous events.

Having this in mind, a brief comparison between these variables and other factors such as the consumption of drugs and/or alcohol, physical and mental health, and others, can be made, with all data comparisons being present in **Annex II, Figure 2**. All multivariate analysis charts were made using the *ggplot* command in *R*, part of the *ggplot2* package.

In order to develop a method of prediction of heart disease based in each patient's answers, we used logistic regression to evaluate the probability of risk of heart disease, which is known for being an algorithm that finds success/failures probabilities of a given event when its dependent variable is binary (in our case, "Yes" and "No"). Therefore, we used the binomial logistic regression as a model to predict cardiovascular disease based in the questionnaire answers. To evaluate the probability of cardiovascular disease, we used the fitted values for our dataset, followed by a ranking of the values of such probabilities, by increasing order.

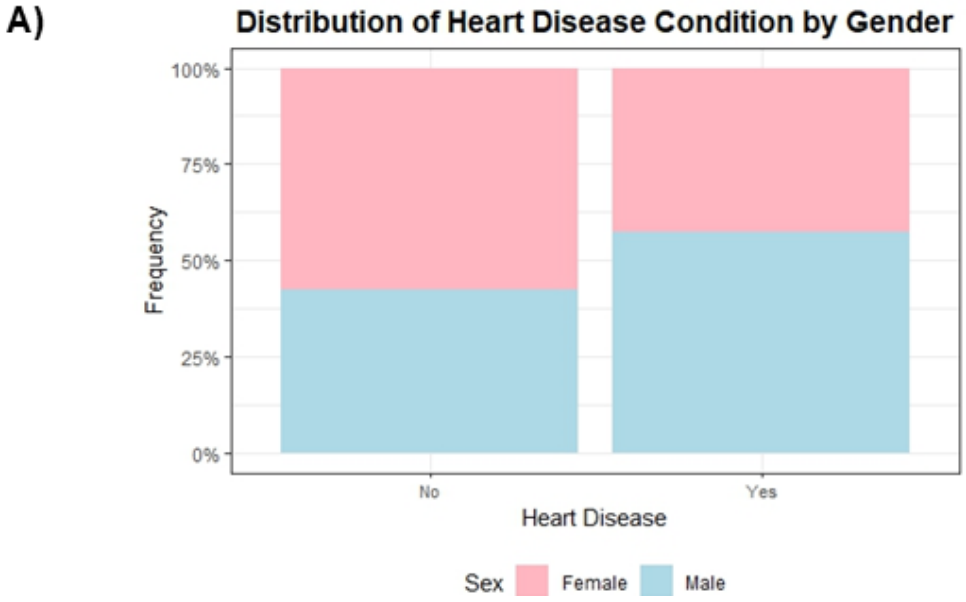
Two data partitions, *training* and *testing* were created, for training and if validated, plotting purposes. The *training* dataset contained 70% of the sample values (n=177577), a value we found reasonable for training purposes, due to an exceptionally large sample size. Next, we performed a likelihood ratio test, since a logistic regression is said to provide a better fit to our data if it demonstrates an improvement over a model with fewer predictors. In our

case, two fitting models were created, with the aim of testing whether the observed differences in the model fit are statistically significant from the reduced model fit, with fewer predictors. The first model plotted the heart disease risk against all variables in our data set, and the second model only incorporated the most statistically relevant variables in our data set, being considered a reduced model. A chi-squared ANOVA test of comparison with the models allowed us to reject the null hypothesis – that is, that given that H_0 states that the reduced model is a better fit for our data set (p-value: 2.283e-06). Therefore, we used the non-reduced model for the implementation of the confusion matrix.

3. Results and Discussion

3.1. Graphical analysis

The charts presented in **Figure 1A and 1B** provide an overview of the heart disease or attack in the sample distributed by age and gender, both of which are key factors in analysing the prevalence of cardiovascular diseases. These demographic factors are linked to cardiovascular risk and are essential for understanding the epidemiology of these conditions.



B)

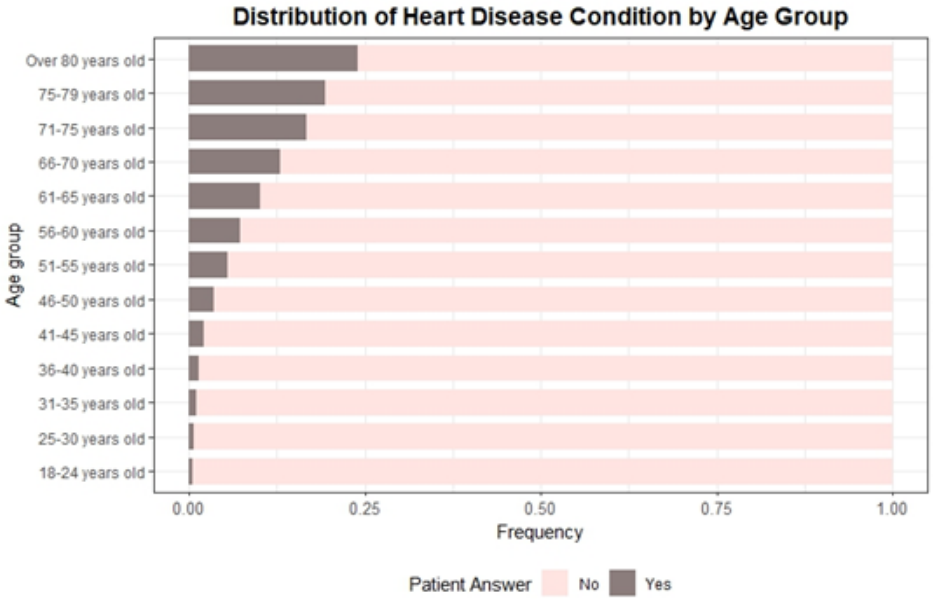


Figure 1: Panel of A) distribution of heart disease conditions by gender and B) by age group.

The age distribution presented in **Annex A, Figure 1N** reveals a higher concentration in the older age groups, with the highest percentages in the 61-65 years (13.1%), 66-70 years (12.7%), and 56-60 years (12.2%) categories. Aging is a major risk factor for cardiovascular diseases, as it is associated with physiological changes such as arterial stiffness and endothelial dysfunction. The cumulative effect of risk factors, such as hypertension and diabetes, further increases the risk in older populations. This is illustrated in the graphic, which shows that individuals over 80 years of age exhibit the highest prevalence of poor cardiac conditions. In contrast, younger age groups, such as 18-24 years and 25-30 years, exhibit a significantly lower prevalence of cardiovascular diseases, except in cases of genetic predisposition or unhealthy behaviours such as smoking.

Regarding the gender distribution, present in **Annex I, Figure 1M**, 56.0% of the sample is female, while 44.0% is male. According to the graphic, cardiovascular diseases affect both sexes differently, regardless of age, with males exhibiting a higher prevalence of a history of heart problems compared to females. In men, the risk is higher, specially at younger ages, often linked to elevated LDL cholesterol levels and lifestyle factors like smoking and alcohol consumption. However, in women, the risk of cardiovascular disease increases significantly

after menopause due to the decline in oestrogen levels, which previously provided cardiovascular protection by maintaining vascular elasticity and promoting a favourable lipid profile [2].

Figure 2 depicts the income distribution, in relative frequency, of all participants of the questionnaire. The income distribution depicted highlights the significant impact of socio-economic status on cardiovascular disease risk. Income level is a key determinant, influencing healthcare access, lifestyle factors, and overall cardiovascular burden.

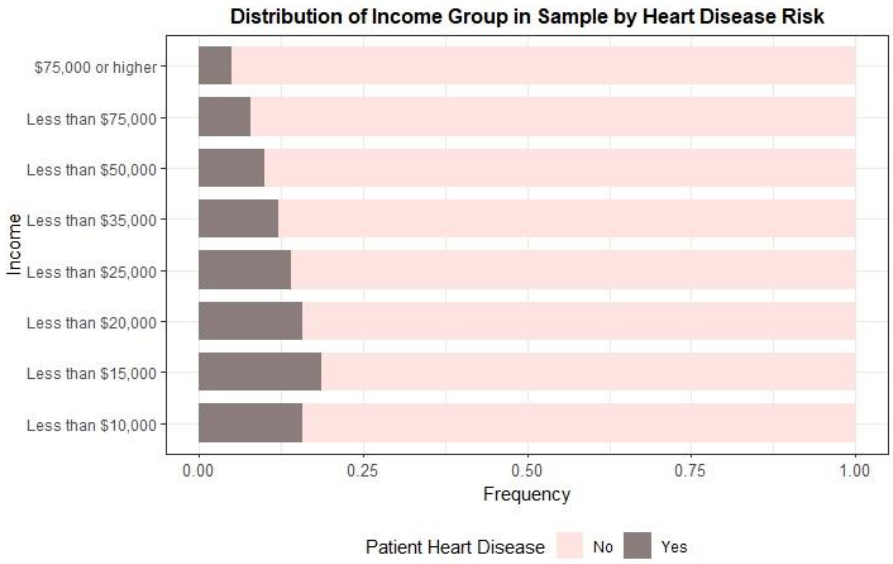


Figure 2: Proportion of cardiovascular disease risk grouped by income distribution of all participants of the questionnaire.

In our sample, most of the respondents without history of heart disease or attack fall within the > $75,000 income group, suggesting better access to healthcare and preventive services. Higher income is generally associated with healthier lifestyle choices, such as balanced diets and regular exercise, which lower cardiovascular risk. Conversely, individuals in lower income brackets (< $10,000 to $25,000), though fewer in this cohort, typically face higher cardiovascular risk due to limited healthcare access, poorer nutrition, and increased psychosocial stress, all contributing to conditions like hypertension and atherosclerosis.

Moreover, disparities in healthcare lead to delayed diagnoses and suboptimal treatment in lower-income groups, exacerbating cardiovascular morbidity and mortality. Despite the

sample's skew toward higher income levels, socio-economic disparities remain a critical factor in cardiovascular health, with lower income groups disproportionately affected.

Figure 3 depicts the education status distribution of all respondents of the questionnaire.

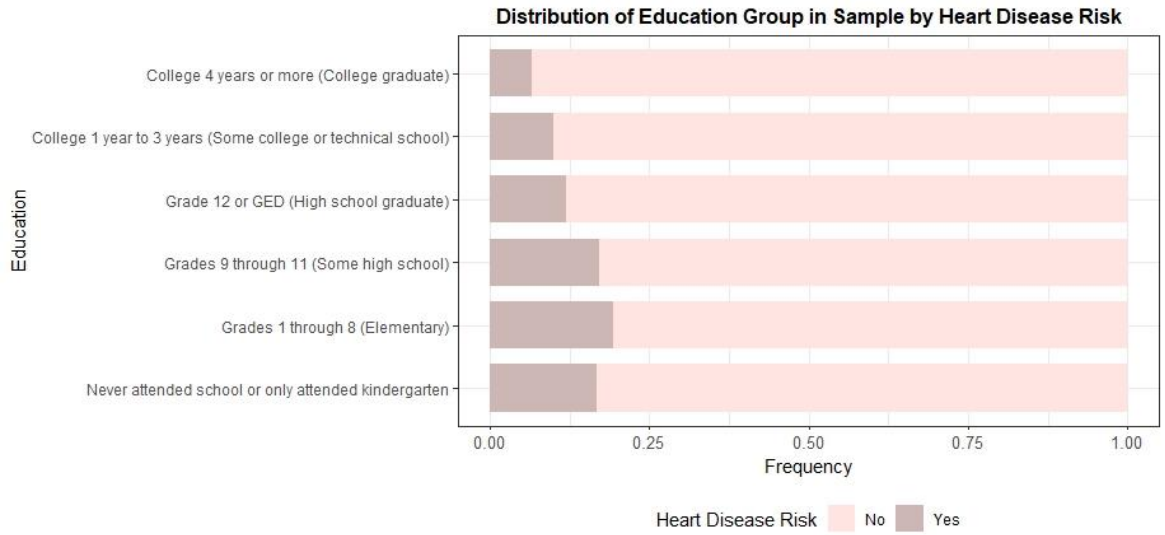


Figure 3: Proportion of education status distribution grouped by heart disease risk.

The education status distribution displayed in the graph highlights the relationship between educational attainment and cardiovascular disease (CVD) risk, an often-underestimated determinant of health. Education is strongly correlated with health outcomes, influencing lifestyle choices, awareness of preventive measures, and access to healthcare resources.

Most of the sample in this cohort has achieved a college degree (4 years or more), followed by those with 1 to 3 years of college education and high school graduates. Higher education is strongly associated with better health literacy, which promotes healthier behaviours, such as adherence to balanced diets, regular physical activity, and the avoidance of smoking and excessive alcohol consumption. Individuals with higher education levels are also more likely to engage in preventive healthcare practices, including regular check-ups and early detection screenings for cardiovascular conditions.

Conversely, the sample with lower education levels, particularly those with only elementary or partial high school education, is notably smaller. Lower educational attainment is associated with reduced access to health information and a lower likelihood of engaging in health-promoting behaviours. Individuals with minimal education are more likely to experience socioeconomic disadvantages, contributing to poorer nutrition, higher stress levels, and limited access to healthcare services, all of which exacerbate CVD risk.

Additionally, education status also influences occupational opportunities and income, further affecting one's ability to afford and access quality healthcare. Those with higher education levels are more financially secure, allowing for better healthcare access and adherence to treatment plans, while lower-educated groups are more vulnerable to cardiovascular risk factors due to financial instability and reduced healthcare engagement.

Regarding other variables other than education and income, the multivariable data analysis allows us to infer those factors such as high blood pressure and cholesterol levels, smoking habits, diabetes and patient healthcare access and doctor affordability (**Annex II, Figures 1A-1D and 1G-1H**) are related to a larger prevalence of cardiovascular disease. Surprisingly, on a first look, there does not seem to be a correlation between alcohol consumption and cardiovascular disease (**Annex II, Figure 1F**): for instance, factors such as the coined French Paradox [7], which correlated the consumption of alcoholic beverages such as red wine, high in antioxidants with a reduced disease of CVD, even on a high cholesterol diet. Granted, not all alcoholic beverages possess these antioxidants, but one can make assumptions that most patients that drink alcohol probably drink red wine, especially since the publication of this research, red wine consumption levels rose by over 40% in the United States [8]. However, one can only make assumptions about the lack of correlation between alcohol consumption and CVD.

3.2. Measuring the probability of Cardiovascular Heart Disease

As previously stated, we used the binomial logistic regression model to predict the risk of cardiovascular disease based on the patient. **Figure 4** depicts the generalized binomial linear model using all our patients' data.

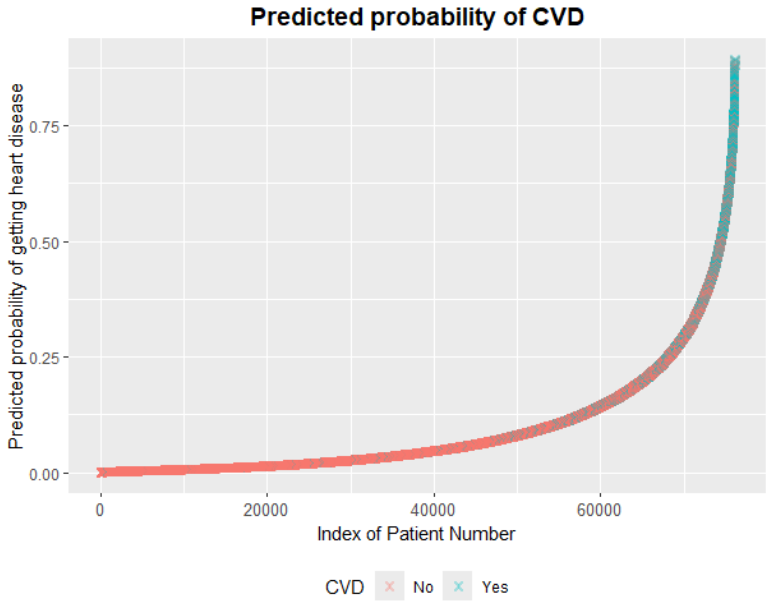


Figure 4: Logistic regression of the predicted probability of cardiovascular disease in the test set of patients.

By analysing the figure, it is noticeable that most patients have in fact a diagnosis that is appropriate for their probability of heart disease; however, some cases of patients that were diagnosed with heart disease but have a low probability of such are noticeable.

Thia can be justified by the results of the validation of the predicted values: on a deeper analysis, this validation was carried out using a classification rate method with recursion to a confusion matrix presented in **Annex III, Table 1**. The confusion matrix presented an accuracy of 0.9076 and a sensitivity of 0.9889, which are acceptable results. However, the specificity is of a mere 0.1256 and the negative predictive value is of 0.5416, which indicate that there is an exceptionally low proportion of negative predictions (predicted patients that do not have heart disease) that are negative. This is countered by the extremely high positive predictive value of 0.9158, which means that the threshold needs to be changed to prevent so many false

negatives. Taking all of this into account, the balanced accuracy is of 0.5576. A Receiver Operating Characteristic (ROC) curve can be an answer to this problem, but such analysis is out of the scope of this work. Also, these false positives values could be due to the overall design of the questionnaire and its bias, alongside it not having questions that do cover all the patient's medical history nor has in account factors such as genetics, for instance. However, the binomial linear regression is a powerful and useful tool to help aid in predicting heart disease, but it needs to be coupled with a data source that is dependable and comprehensive and needs further treatment, out of the scope of this work.

Key Points:

- Although subjected to bias, the Behavioural Risk Factor Surveillance System (BRFSS) is a respectable tool for analysis of the relationship between various lifestyle factors and the likelihood of developing heart diseases over time, which in turn can translate into a better understanding of patterns and predictions of heart disease in the years to come.
- In our Exploratory Data Analysis (EDA), we conclude that (insert factors) are the factors that can be the most correlated with Cardiovascular Heart Disease (CVD).
- With the data of the questionnaire, we were able to predict the probability of heart disease in patients having into account their answers to the questionnaire, allowing to develop predictive heart disease models, however, with some source of bias.

References

1. Vasatova M, Pudil R, Horacek JM, Buchler T. Current Applications of Cardiac Troponin T for the Diagnosis of Myocardial Damage. In: 2013: 33–65.

2. Balakumar P, Maung-U K, Jagadeesh G. Prevalence and prevention of cardiovascular disease and diabetes mellitus. Pharmacol Res 2016; 113: 600–609.

3. U.S. Department of Health & Human Services Accessibility External Links Privacy Policies No Fear Act FOIA Nondiscrimination OIG Vulnerability Disclosure Policy Public Health Publications USA gov. Behavioral Risk Factor Surveillance System. .

4. Leiyu Shi DAS. Essentials of the U.S. Health Care System. 6th ed. .

5. Prevalence of Uncontrolled Risk Factors for Cardiovascular Disease: United ... - Cheryl D. Fryar, Te-Ching Chen, Xianfen Li - Google Livros. https://books.google.pt/books?hl=pt-PT&lr=&id=MQmDxjrBdrYC&oi=fnd&pg=PA4&dq=cardiovascular+disease+united+states&ots=t3u05xDItu&sig=9ykvmFyd4Vu9TE1aKPYB6qO4IeA&redir_esc=y#v=onepage&q=cardiovascular%20disease%20united%20states&f=false (October 23, 2024, date last accessed).

6. Heart Disease Health Indicators Dataset. https://www.kaggle.com/datasets/alexteboul/heart-disease-health-indicators-dataset (October 21, 2024, date last accessed).

7. Langlais F, Kerboull M, Sedel L, Ling RSM. The “French paradox.” Journal of Bone and Joint Surgery - Series B 2003; 85: 17–20.

8. Phillips Roderick. Alcohol : a history. 2014; 655.

9. Weir CB, Jan A. BMI Classification Percentile And Cut Off Points. StatPearls 2023;

Annex I: Questionnaire

In the variables under study, responses to the “Yes/No” questions were registered using a binary code, with 0 for “No” and 1 for “Yes”, with the exception of the *Diabetes* and *Sex* categories, with the former having codes 0, 1 and 2 for absence or presence of pre-diabetes and diabetes, and the latter having answer 0 as “Female” and 1 as “Male”. The correspondence between the numbers and the answers are in accordance with the 2015 BRFSS questionnaire and present in **Table 1**. All variables in Table 1 were converted to factors with their corresponding levels in *RStudio*.

Table 1: Distribution of questionnaire category answers.

Heart disease attack High blood pressure High cholesterol Cholesterol check Smoker Stroke Physical activity Fruits 1+ times per day Veggies 1+ times per day Heavy alcohol consumption Any healthcare No doctor visit because of cost Difficulties walking	0 = No 1 = Yes
Age	1 = 18-24 years old 2 = 25-30 years old 3 = 31-35 years old 4 = 36-40 years old 5 = 41-45 years old 6 = 46-50 years old 7 = 51-55 years old 8 = 56-60 years old 9 = 61-65 years old 10 = 66-70 years old 11 = 71-75 years old 12 = 76-80 years old 13 = Over 80 years old
Income	1 = Less than $10.000 2 = $10.000 - $15.000 3 = $15.000 - $20.000 4 = $20.000 - $25.000 5 = $25.000 - $35.000 6 = $35.000 - $50.000 7 = $50.000 - $75.000 8 = $75.000 or more
General health	1 = Excellent 2 = Very good 3 = Good 4 = Fair 5 = Poor
Education	1 = Never attended school or only attended kindergarten 2 = Grades 1 through 8 (Elementary) 3 = Grades 9 through 11 (Some high school) 4 = Grade 12 or GED (High school graduate) 5 = College 1 year to 3 years (Some college or technical school) 6 = College 4 years or more (College graduate)
Diabetes	0 = No 1 = Pre-diabetic 2 = Diabetic
Sex	0 = Female 1 = Male

The responses to the variables “Mental health” and “Physical health” represent the number of the days within a month in which a person reported feeling mentally and physically unwell, respectively. These values range from ‘0’ days (indicating no days feeling unwell) to ‘30’ days (indicating feeling unwell every day).

According to NCBI [9], body mass index (BMI) is a statistical measure that estimates body fat based on a person’s weight and height, applicable to all genders of all ages. In our work, the body mass indexes were not separated into categories, but according to NCBI, these follow the classification of:

a) < 16.5 (kg/m^2): Severely underweight;

b) 16.5 ≤ BMI < 18.5: Underweight;

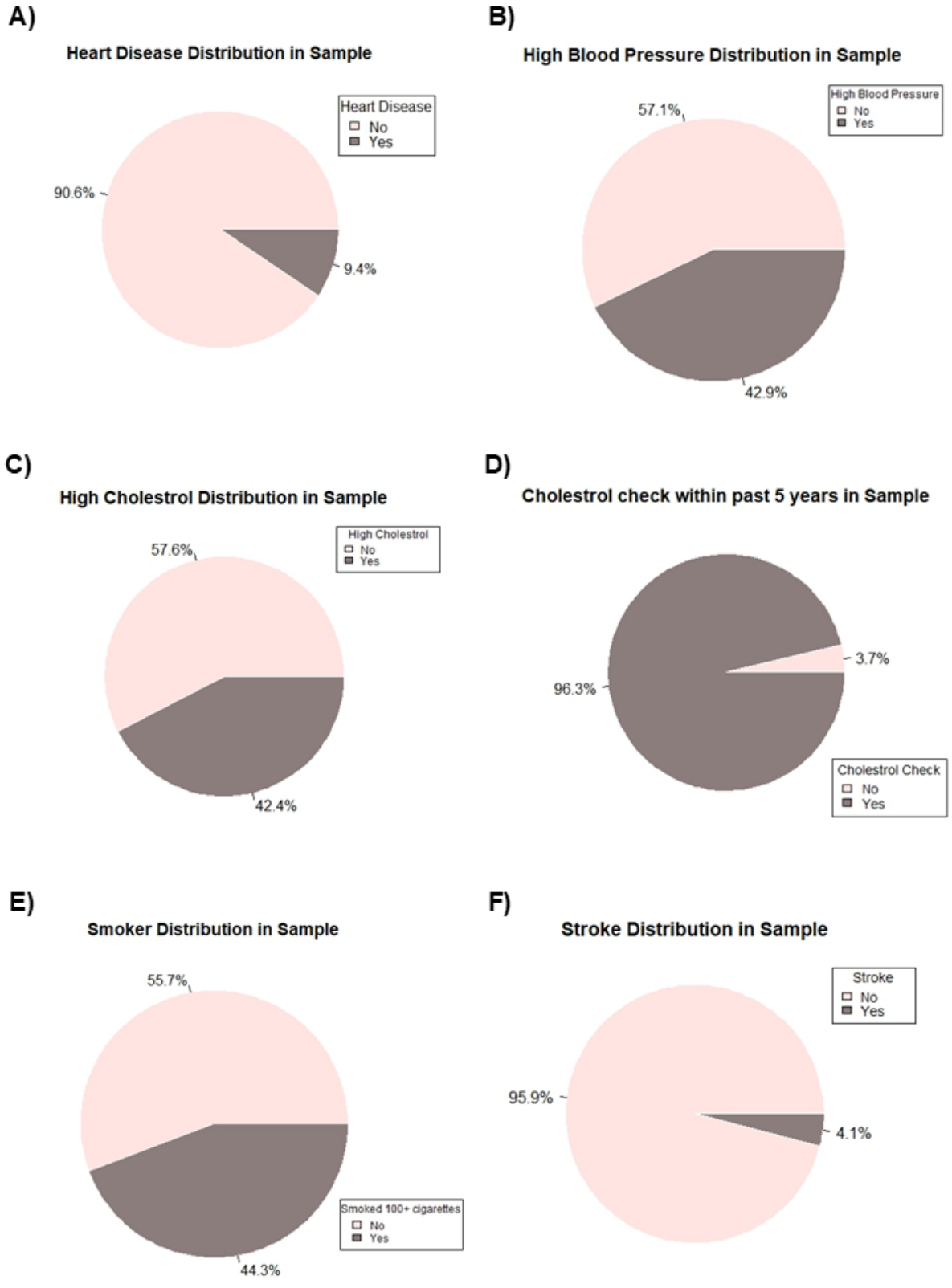
c) 18.5 ≤ BMI ≤ 24.9: Normal weight;

d) 25 ≤ BMI ≤ 29.9: Overweight;

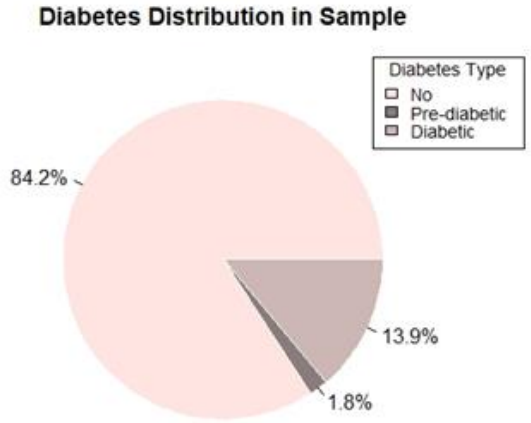
e) ≥ 30 kg/m^2 : Obesity (with 3 obesity classes).

Annex II: Supplementary figures

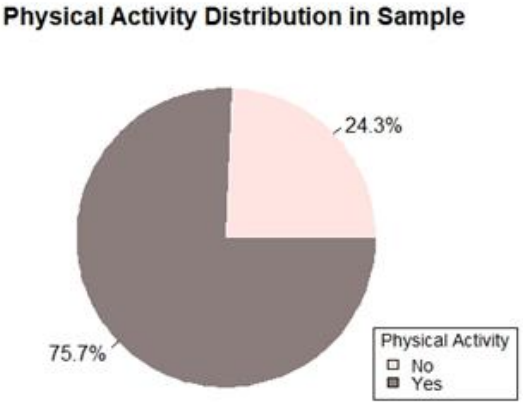
Figure 1: Panel of pie charts of distributions of patient answers for every questionnaire category.



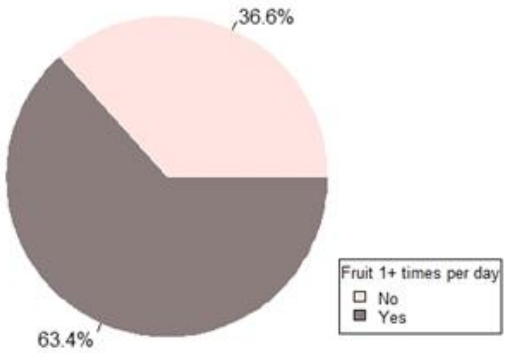
G)


Diabetes Distribution in Sample
Diabetes Type
No
Pre-diabetic
Diabetic
84.2%
13.9%
1.8%

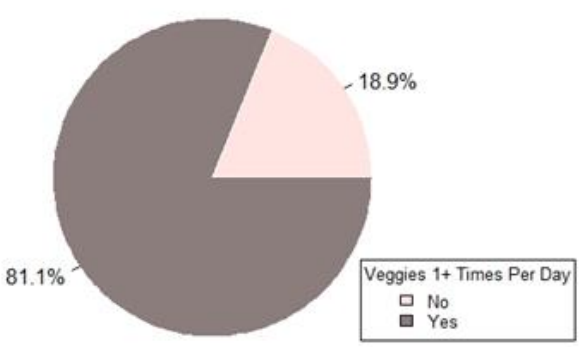
H)


Physical Activity Distribution in Sample
24.3%
75.7%
Physical Activity
No
Yes

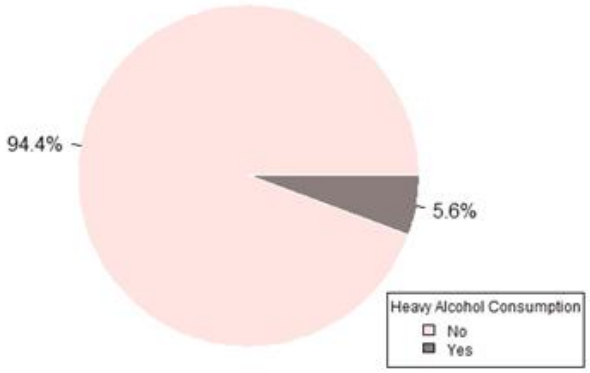
I)


Fruit Consumption in Sample
36.6%
63.4%
Fruit 1+ times per day
No
Yes

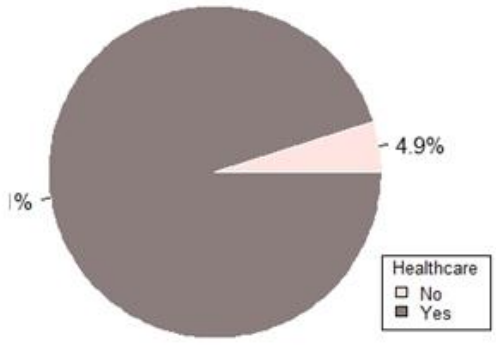
J)


Veggies Consumption in Sample
18.9%
81.1%
Veggies 1+ Times Per Day
No
Yes

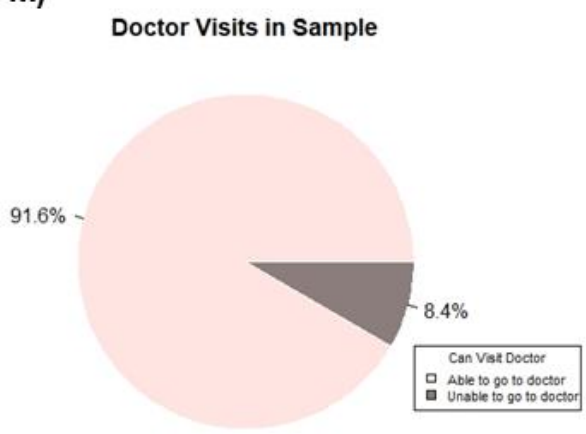
K)


Heavy Alcohol Consumption in Sample
94.4%
5.6%
Heavy Alcohol Consumption
No
Yes

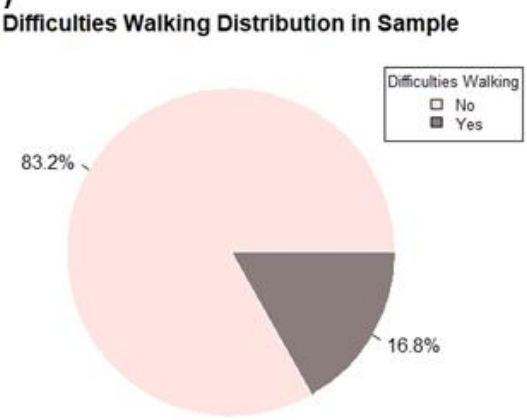
L)


Healthcare Access in Sample
4.9%
%
Healthcare
No
Yes

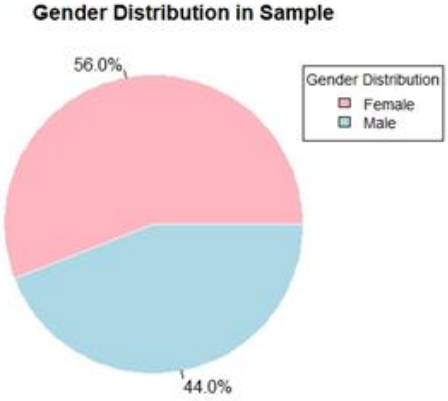
M)


Doctor Visits in Sample
91.6%
8.4%
Can Visit Doctor
Able to go to doctor
Unable to go to doctor

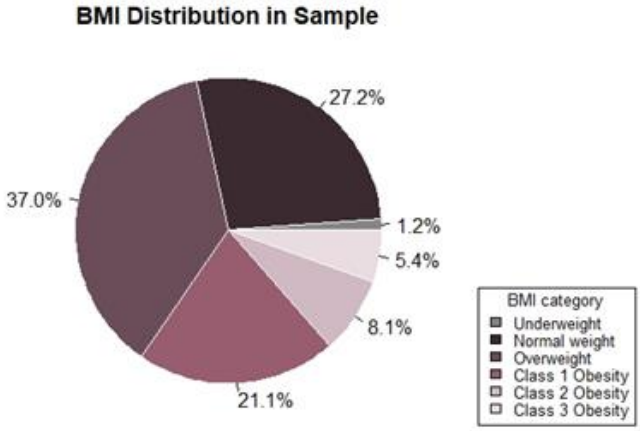
N)


Difficulties Walking Distribution in Sample
Difficulties Walking
No
Yes
83.2%
16.8%

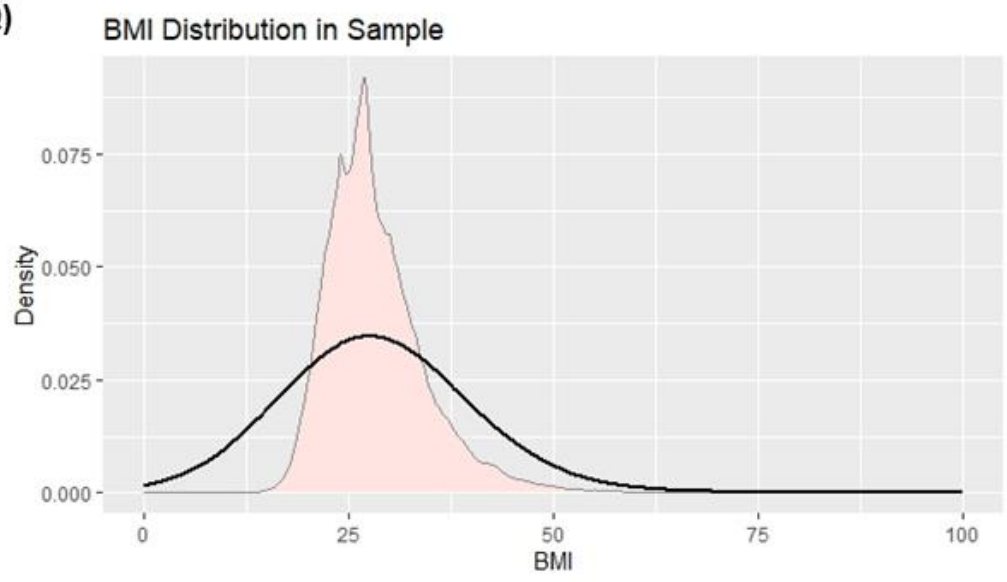
O)


Gender Distribution in Sample
56.0%
44.0%
Gender Distribution
Female
Male

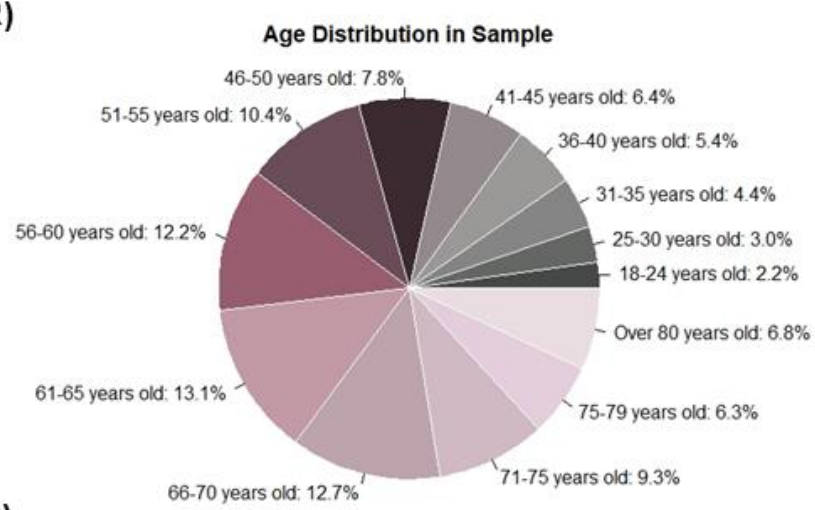
P)


BMI Distribution in Sample
27.2%
37.0%
1.2%
5.4%
8.1%
21.1%
BMI category
Underweight
Normal weight
Overweight
Class 1 Obesity
Class 2 Obesity
Class 3 Obesity

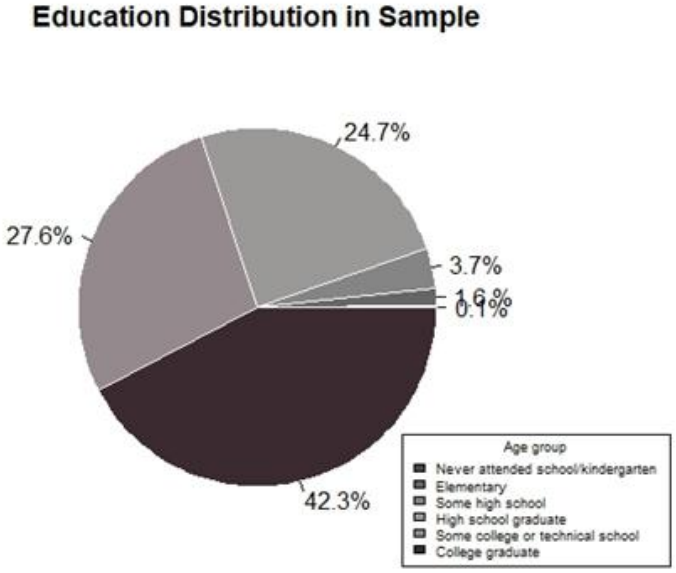
Q)


BMI Distribution in Sample
Density
0.075
0.050
0.025
0.000
0
25
50
75
100
BMI

R)


Age Distribution in Sample
46-50 years old: 7.8%
41-45 years old: 6.4%
51-55 years old: 10.4%
36-40 years old: 5.4%
31-35 years old: 4.4%
56-60 years old: 12.2%
25-30 years old: 3.0%
18-24 years old: 2.2%
Over 80 years old: 6.8%
61-65 years old: 13.1%
75-79 years old: 6.3%
71-75 years old: 9.3%
66-70 years old: 12.7%

S)


Education Distribution in Sample
24.7%
27.6%
3.7%
1.6%
0.1%
42.3%
Age group
Never attended school/kindergarten
Elementary
Some high school
High school graduate
Some college or technical school
College graduate

T)

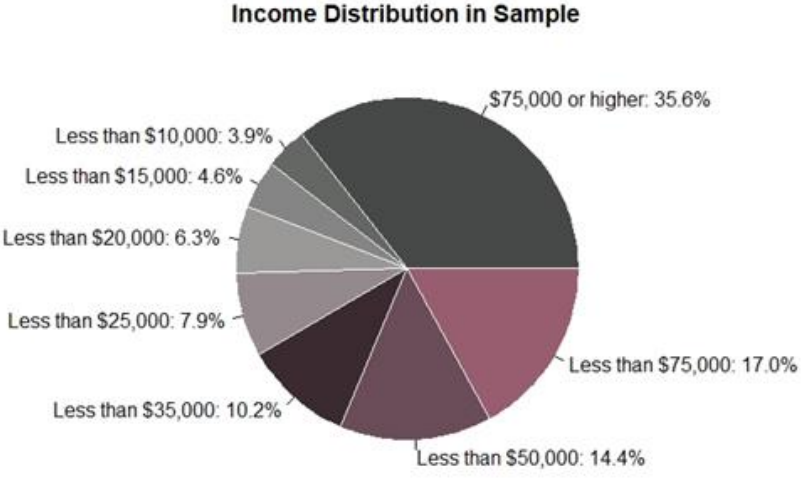
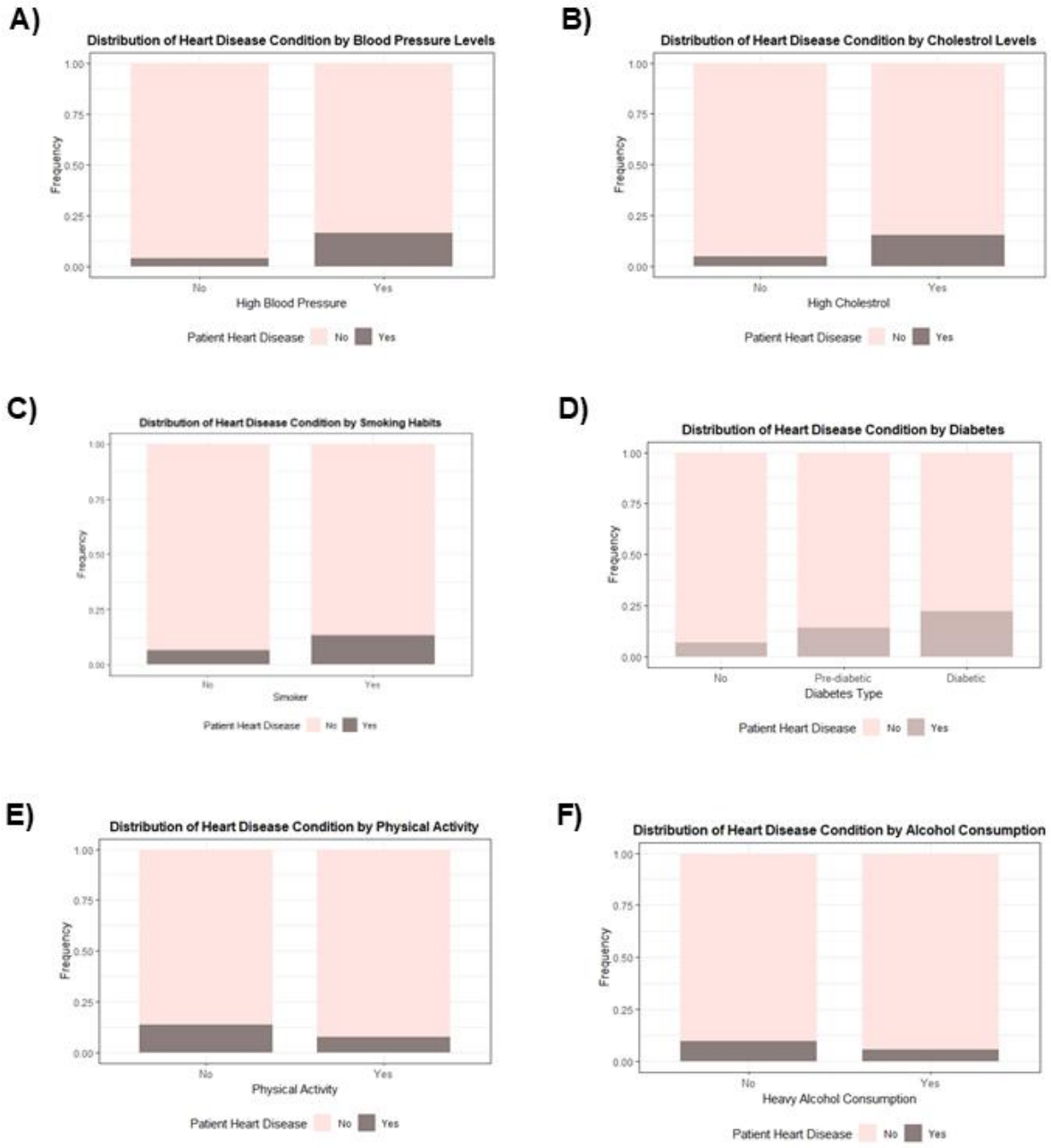
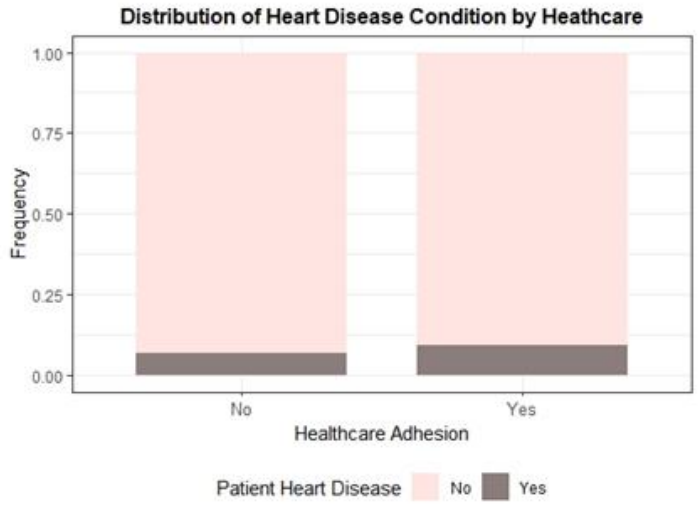

Income Distribution in Sample
$75,000 or higher: 35.6%
Less than $10,000: 3.9%
Less than $15,000: 4.6%
Less than $20,000: 6.3%
Less than $25,000: 7.9%
Less than $75,000: 17.0%
Less than $35,000: 10.2%
Less than $50,000: 14.4%

Figure 2: Multivariate analysis bar plots, with questionnaire category fill.



G)


Distribution of Heart Disease Condition by Heathcare
1.00
0.75
0.50
0.25
0.00
Frequency
No
Yes
Healthcare Adhesion
Patient Heart Disease
No
Yes

H)

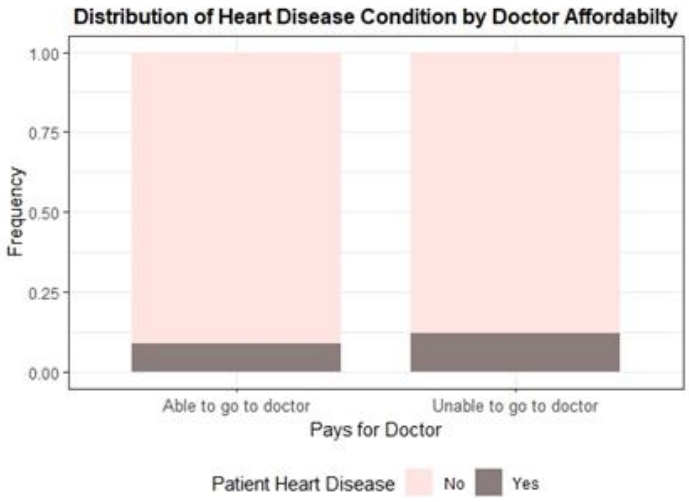
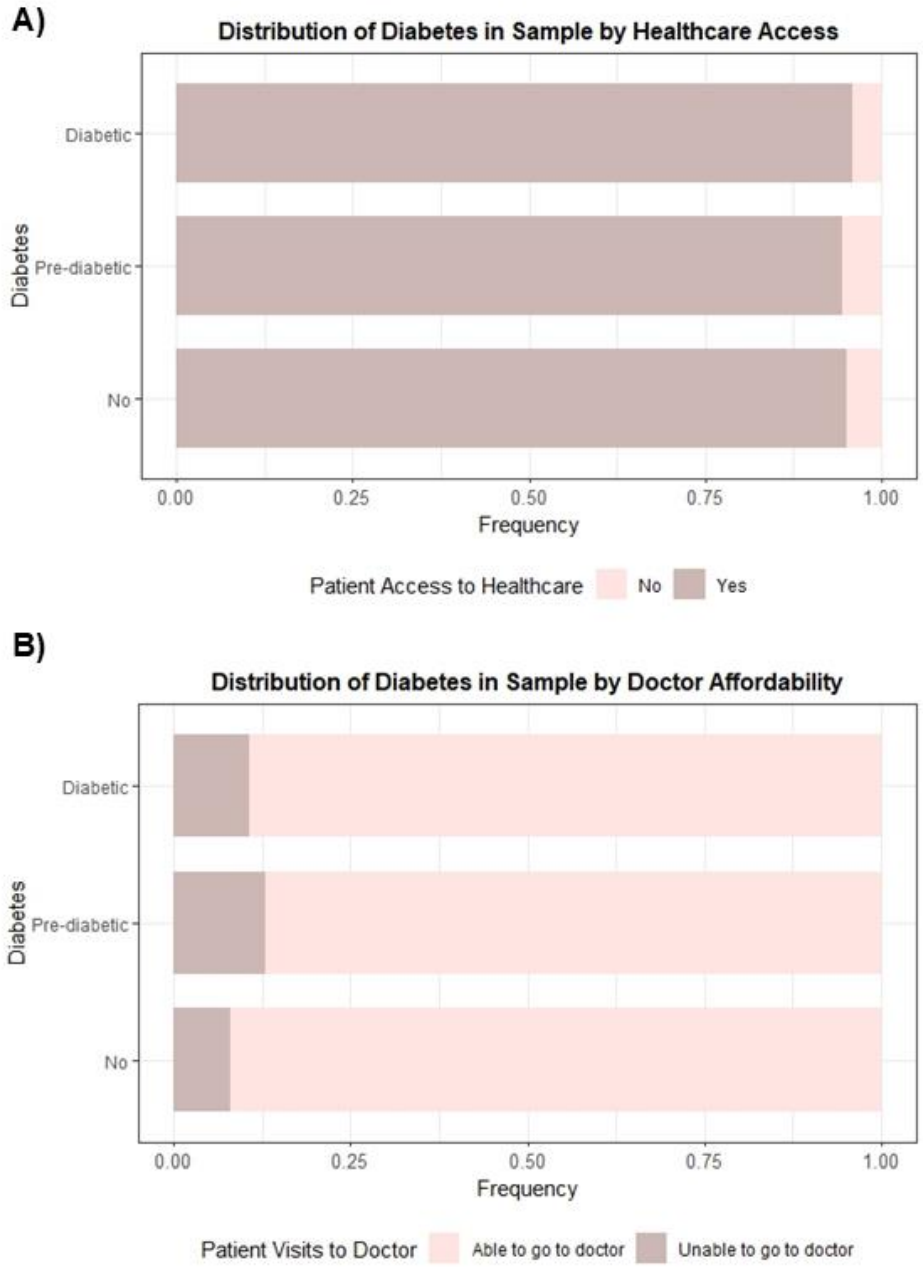

Distribution of Heart Disease Condition by Doctor Affordabilty
1.00
0.75
0.50
0.25
0.00
Frequency
Able to go to doctor
Unable to go to doctor
Pays for Doctor
Patient Heart Disease
No
Yes

Figure 3: Proportion of A) healthcare and B) doctor visits by types of diabetes.



Annex III: Confusion Matrix

The confusion matrix was calculated using the base *RStudio* package.

Table 2: Confusion Matrix of the training data partition.

		Reference	
		No	Yes
Prediction	No	159073	14625
	Yes	1778	2101

Parameters obtained:

- Accuracy : 0.9076
- 95% CI : (0.9063, 0.9090)
- No Information Rate : 0.9058
- P-Value [Acc > NIR] : 0.004312
- Kappa : 0.17547
- Mcnemar's Test P-Value : < 2.2e-16
- Sensitivity : 0.9889
- Specificity : 0.1256
- Pos Pred Value : 0.9158
- Neg Pred Value : 0.5416
- Prevalence : 0.9058
- Detection Rate : 0.8958
- Detection Prevalence : 0.9782
- Balanced Accuracy : 0.55763
- 'Positive' Class : No