Written Report Analysis for Group 1

This written analysis report aims to provide insights into the correlation of various risk factors and prevalent coronary heart disease. The dataset itself is a longitudinal study conducted from 1956 to 1968 on 4,434 participants. Our analysis will be obtained using Period 1 (beginning of study) and Period 3 (end of study). The dataset itself is a longitudinal study conducted from 1956 to 1968.

I. General Data Trends

In both period 1 and 3 data most of the participants smoked 0 cigs per day indicating that the majority did not smoke. For the majority of participants, risk factors such as total cholesterol and BMI were like their values from the beginning of the study.

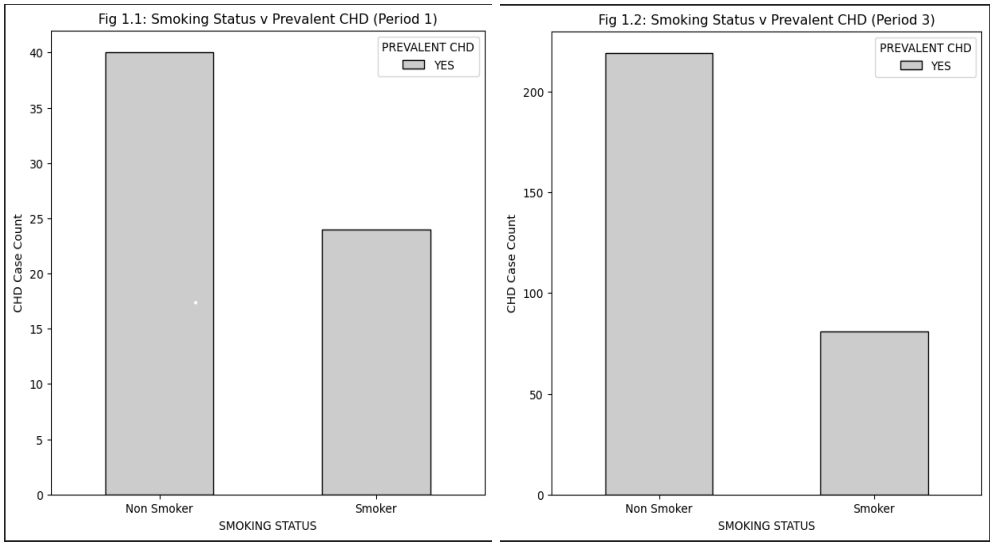
2754 participants were included in the final analysis. All completed 3 periods of measurements. 1546 participants were female and 1208 were male.

The number of participants with CHD rose from 64 in Period 1 to 300 in Period 3, indicating 236 additional participants experienced a CHD episode by Period 3.

Around 85% of participants in both Periods fell into the Healthy weight and Overweight BMI category. In general, the number of Overweight and Healthy Weight participants was very similar in both Period 1 & 3 (1230 healthy weight and 1168 overweight in period 1, 1175 healthy weight and 1167 overweight in period 3)

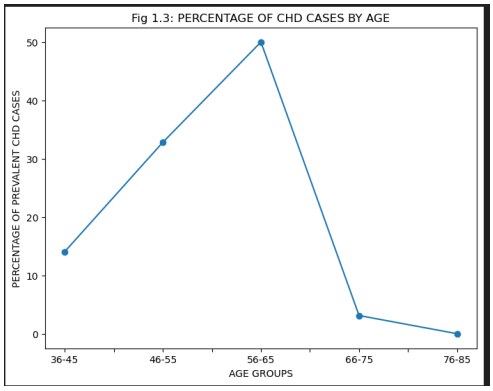
II. Smoking vs Prev CHD: Is there a correlation between participants' smoking status and CHD?

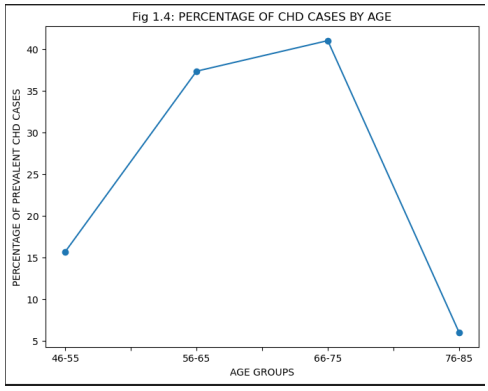
In our analysis, the null hypothesis was that there is no correlation between smoking status and CHD. To study this correlation we first plotted the number of smokers and non-smokers with CHD. From this bar graph, we can see that in both periods of data there is a larger number of non-smokers vs smokers. Additionally, the number of participants with CHD rose from period 1 to period 3. This indicates that there was an increase in CHD cases from period 1 to 3 independent of the participants smoking status. This was further confirmed by our chi square test. In our test for both periods, we got a chi-squared value of 0 with a critical value of 3.84 when the degree of freedom is 0. This proved our null hypothesis correct; there is no correlation between smoking status and CHD. For healthcare professionals, this indicates that smoking status by itself does not contribute to CHD. It may be worth looking at other risk factors as contributors or smoking status in combination with other risk factors. This will help them determine the right group of treatment/help to minimize/treat CHD cases.



III. <u>Age vs CHD: What is the distribution of age among prevalent CHD cases and what does this mean?</u>

As part of our analysis, we chose to look at the spread of prevalent CHD cases across various age groups for both time periods. In period 1, the age group 56-65 has the highest number of individuals with CHD which was followed by the 45-55 age group. Now, since this was a longitudinal study we can expect a shift in the graph towards to the right. In other words we expected to see an increase in the 56-65, 66-75 age group as it's the same population from period 1 after an (10-12 years). This was indeed the case as these age groups accounted for the majority of the CHD cases. Interestingly we also did notice an increase in the number of participants in period 3. This was due to the fact that more participants were diagnosed with CHD by the end of the study. This could be possible due to a number of reasons, one being the fact that aging itself could be a potential contributor of CHD. ANother notable factor is the change in diet associated with the time period of the study as more processed/refined food is being eaten. It is also possible the lack of technology to detect CHD early on. CHD was also defined as heart attacks that were silent or asymptomatic, so it's possible that some participants never noted it or sought medical care. This could mean that a participant could have had CHD for a while but the medical equipment could have never picked it up causing us to see false data trends. This analysis has helped identify possible factors such as age, diet, technology that healthcare professionals should take into consideration when diagnosing/treating CHD.





IV. What is the distribution of total cholesterol amongst participants with CHD? Does it vary across sexes? Smoking status?

Sex and Smoking

In Fig 2.1, we clearly see that females account for a larger sum of the data, but men account for more cases of CHD in Period 1, although the combined total is significantly lower than the total population. When we compare the results with the same filters for Period 3, as seen in Fig 2.2, we find an increase in CHD cases over the periods. To find some contributing factors to the rise in CHD cases, we filter by smoking status of a person in Period 1 versus Period 3 to view any correlation. For both Periods 1 and 3, non-smokers take the lead with least cases of CHD involved, however interestingly enough there is also a drop in smoker count from Period 1 to Period 3, by 365 participants (28%), but also an increase in CHD cases among smokers by 337% (from 24 to 81), with a respective increase 95% by females, and remainder 242% by males in Period 3. Smoking males however, still accounted for the largest count of CHD cases.

Sex, Smoking and Cholesterol

The distribution of cholesterol levels revealed that most CHD participants fell under the ‘abnormally high’ category. No participants fell under the lowest category ‘abnormally low’. Upon further inspection, the cholesterol category was filtered by sex (male and female) in the same period to see the distribution. As seen in the figure ‘Analysis of Cholesterol Level in Participants with CHD by Sex in Period 1’, the distribution revealed that nearly half of all female participants made up for the ‘healthy’ and ‘at risk’.

In Fig 2b.1, half of all participants with CHD have healthy or at risk level cholesterol, while the remainder identify with abnormally high cholesterol, and no participants with abnormally low cholesterol. This is important to acknowledge as cholesterol has been a known factor in affecting the development of CHD in the medical field. Further in Fig 2.2, we find men

with CHD make up larger sums of abnormally high cholesterol, however, there also were more men with CHD in comparison to females.

No trend of significance was shown when viewing CHD by smoking status and cholesterol level, only that non-smokers made up the largest count in the 'abnormally high' category.

Plotting a line graph of the cholesterol distribution across total population, population with CHD and without, revealed that the cholesterol levels tend to peak around the same range of 200-260, which is closely related to the 'at risk' range. This is of significance as it implies that cholesterol may actually not be a big factor in CHD likelihood as per that logic, more individuals would've been expected to develop CHD by the third Period. However, CHD may also require a longer wait time to come to surface, as it takes decades for CHD to develop and the time gap between Period 1 and Period 3 was a little over a decade (12 years).

V. <u>What is the BMI distribution of participants with CHD? Does it vary across sexes? Smoking status?</u>

The last analysis addresses the BMI factor and how it relates to CHD. BMI was provided in the dataset as numerical values in kg/m^2. We found it most appropriate to bin the values into the commonly known categories: Underweight, Healthy Weight, Overweight, and Obesity. The bins were as <18.5, 18.5-24.9, 25.0-29.9, >30.0 respectively. Binning the values allowed us to represent the distribution of the participants with CHD based on their BMI category. This is seen in Figures 3.1 and 3.6. In both periods, the majority of participants with CHD fall into the Healthy weight and Overweight category. In Period 1, more CHD participants were Overweight than Healthy weight by 6.2%, the difference is much smaller in Period 3 of only 0.4%. The percentage of participants with CHD in the Obesity category decreased by 9.7% from Period 1 to Period 3 (25.0% to 15.3%), indicating a general decrease in BMI in participants with CHD from Period 1 to Period 3.

The purpose of studying the distribution of BMI in CHD participants is to bring to light any trends between the two variables that a healthcare professional such as a cardiologist may benefit from.

Analyzing this further, we have looked into the distribution of CHD vs BMI across the two sexes and smoking status.

a. Sex

Looking at the distribution of sex in participants with CHD, male participants, in both periods, had a higher rate of CHD - leading by 20.9% in Period 1 and 15.4% in Period 3. The

rate of female participants with CHD rose by 3.1 by Period 3 indicating more female participants had a CHD episode.

In both periods, male participants had a higher rate of CHD in all BMI categories except Underweight in Period 1 (where no participants had CHD) and Obesity in both Periods. In Period 1, the number of male participants who had CHD was almost double the number of female participants who had CHD in the Healthy weight and Overweight category (7 vs 15 in healthy weight and 9 vs 17 in overweight). In Period 3, the number male participants with CHD was also double that of female participants with CHD in the Overweight category, but only by a small margin in the Healthy weight category (40 vs 85 in Overweight and 61 vs 63 in Healthy weight). Interestingly, Obesity is the only category where the number of female participants with CHD was higher than male participants, in both periods (9 vs 7 in Period 1 and 24 vs 22 in Period 3). Despite the difference being small, it's the only anomaly.

The purpose of analyzing BMI trends in CHD for both sexes is for healthcare professionals such as cardiologists to use it in their counseling of their heart disease patients.

b. Smoking

Looking at the distribution of smoking habits in participants with CHD, non-smokers had a higher rate of CHD than non-smokers in both periods. They led by 25% in Period 1 and 46% in Period 3. The rate of smokers with CHD dropped by 10.5% from Period 1 to 3, indicating a portion of participants gave up smoking.

In both periods, non-smokers participants had a significantly higher rate of CHD across all BMI categories, except Healthy weight in Period 1. In Period 1, the number of non-smoking participants with CHD was double the number of smoking participants with CHD in the Overweight category and triple the number of smoking participants with CHD in the Obesity category. In Period 3, the number of non-smoking participants with CHD is double, triple, and quadruple the number of smoking participants with CHD in the Healthy weight, Overweight and Obesity categories. It appears that the difference between smokers and non-smoker rate of CHD is emphasized in the Obesity category in both Periods. Healthy weight in Period 1 is the only case where the number of participants with CHD is equal for smokers and non-smokers.

Such trends can be helpful for healthcare professionals such as cardiologists and also in smoking cessation programs when considering the patient's risk factors.