Risk Scoring for CHD

Between Two Arms

Reenu Rathi – A00258770 12/5/18 Data Analytics

Table of Contents

ABSTRACT

Purpose of this study to determine the coronary risk scores in adult male and female of control and treatment groups. The study comprised a sample of people (29-91 years, n=547) and the entire elderly population (33-84 years, n=444). The risk score will be analysed based on sex, age, smoking, diabetes s, systolic blood pressure, total cholesterol, and HDL. The risk profile based on age and sex one group will be compared with the other group so find the risk score difference between Group. Based on these results, in the groups studied, the risk of coronary artery disease risk score may be not differing between the control and treatment group whoever differs when compared based on Gender and Age and Smoking History

Key words: coronary artery disease, risk score, population-based study

---

Introduction

Background:

Cardiovascular diseases account for 18 million deaths per year in the world, coronary artery diseases and cerebrovascular diseases being responsible for two thirds of these deaths and for approximately 22% of the 55 million deaths due to all causes. Estimates on mortality due to cardiovascular diseases according to the region indicate that developing countries contribute with a greater part of the overall burden of mortality due to the disease than developed countries, with a relative excess of 70%.

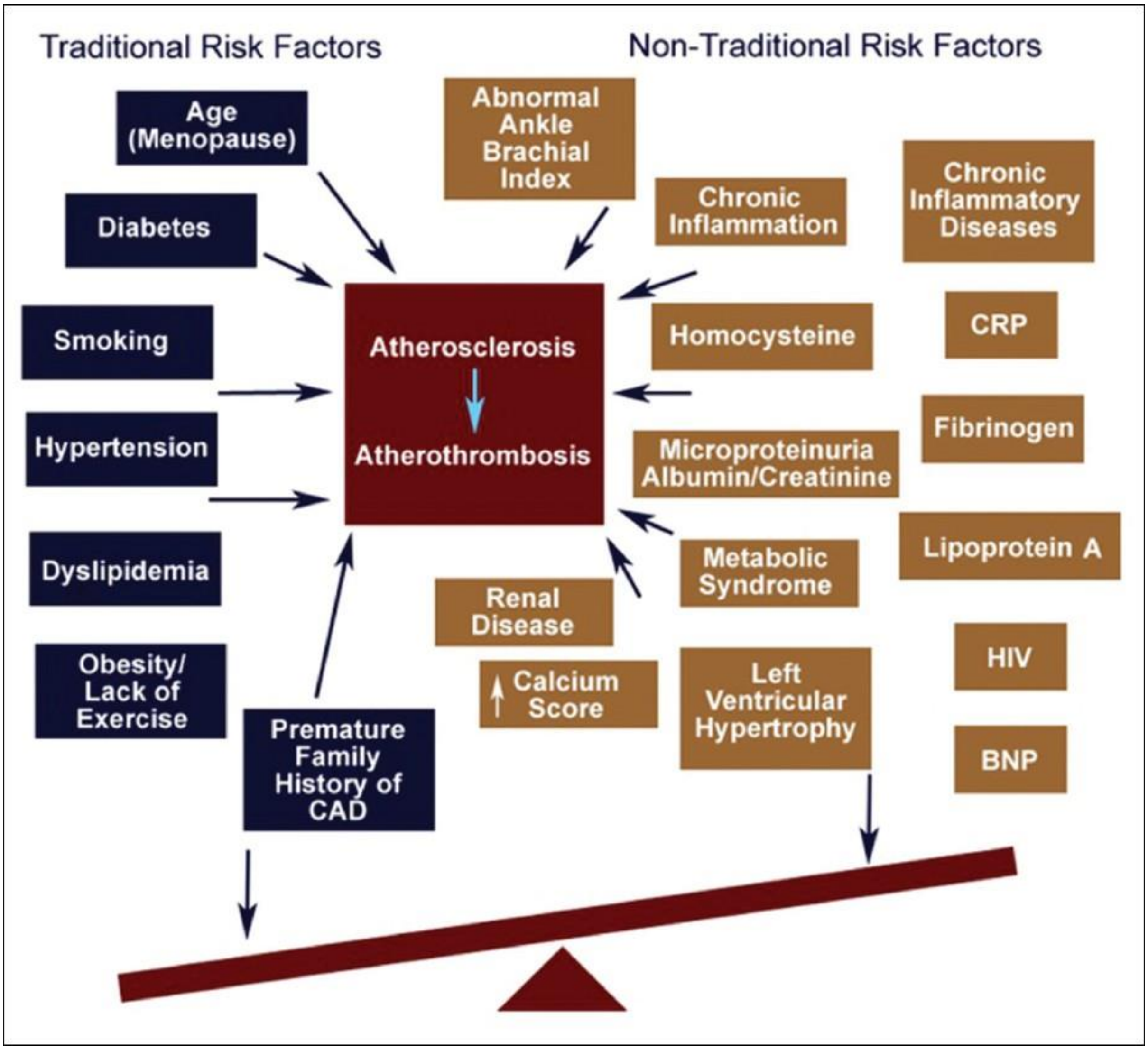
Despite this evidence, epidemiological studies have shown that cardiovascular diseases are a relatively rare cause of death in the absence of major risk factors Almost 75% of the new cases of cardiovascular diseases occurring in developed countries in the 1970s and 1980s could be explained by inadequate diet and physical activity, expressed by high lipid levels, obesity, and increased blood pressure, associated with smoking. The study of these risk factors relates their presence and intensity to the development of the disease. These prospective studies were responsible for the development of risk scores, which allow for the estimation of the probability of developing a certain cardiovascular disease in a defined time interval. These risk scores, in addition to being useful in foretelling a cardiovascular event, aid in its prevention and in the reduction in its incidence in individuals and populations.

The present study aimed at disclosing the risk profile of coronary artery disease between two group which are control and treatment by estimating the risk of developing of coronary heart disease using statistical Analysis major factor which cause the death

The major risk factors:

The factor risk factor which contribute most are categorized in two ways.

1. **Traditional**
2. **Non-Traditional**



In this study we have selected following traditional method.

Age.

SysBp

Cholesterol

HDL

Diabetes

Smoking

Hypertension

Previous History

Data Source

For this study two group (Control and Treatment) of male and female taken. For these group we have the recorded data which will used to perform the analysis to investigate the risk score difference between control and treatment groups. We have a sample a size of 547 in control group which has male and female both between the age group of 29 to 91 and a sample size of 444 in treatment which also has male and female both between an age group of 33 to 84. From entire dataset we have selected the Traditional Factors (mentioned above) column to perform the statistical Analysis to build the risk difference between the Control and Treatment Group. Data has been manipulated using the R language and Categorical data like hypertension, diabetes and Previous.MI which has missing value, has been removed where in for continuous data like HDL, Cholesterol, Systolic BP has been replaced with mean value of that data. While Data manipulation and filtering outlier's presence was noticed in dataset however removing the outliers was resulting to very less sample size which is not good for Analysis as result will not that accurate if the sample would be small.

Methods

Statistical analysis was performed using Minitab statistical software, a significant difference between the groups is determined by analysis of variance (ANOVA) using smoking Status, and 2 sample t and paired t was used for assessment of mean difference between and within the groups. Statistically significant differences are marked with probability $p < 0.05$.

Comparison of the groups was performed on the basis age with a logistic linear regression analysis and the results were described as odds ratios with 95% confidence intervals (95% CIs).

The size of the samples of Control and Treatment group is sufficient to estimate the prevalence of risk factors to 5% significance level, with a confidence interval of 95%.

The parameters used for constructing the risk score were as follows: sex, age, smoking, diabetes, systolic blood pressure, cholesterol, and HDL.

As per the medical research conducted in past cholesterol and Sysbp are directly proportional to increase of risk getting a CHD (Coronary Heart Risk) wherein HDL is protective factor, so an average risk score is calculated for both group which is used a baseline character for Primary analysis.

AvgScore = Cholesterol+ Sysbp – HDL

Statistical Analysis:

For this study, we have divided the analysis in two categories considering our primary and secondary baseline endpoints.

1.Primary Endpoint Analysis 2. Secondary Endpoint Analysis

Primary Endpoint Analysis

1: Gender

First primary analysis is done on AvgScore calculated using the Cholesterol, Sysbp and HDL within Control and Treatment Group taking gender in account.

Firstly, we performed the subjective analysis using the scatterplots and box plots of control and treatment group males and females which was preceded with formal paired t test for both groups

Figure 1: Boxplots of Control and Treatment group based on gender comparing the mean difference.

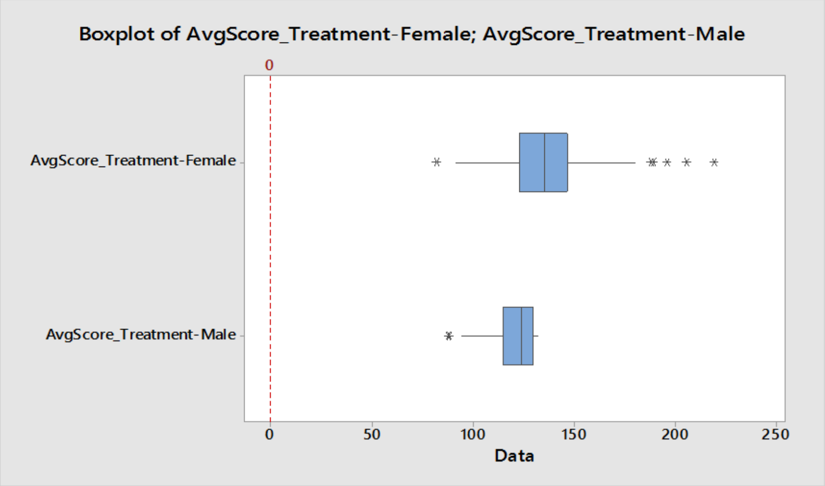


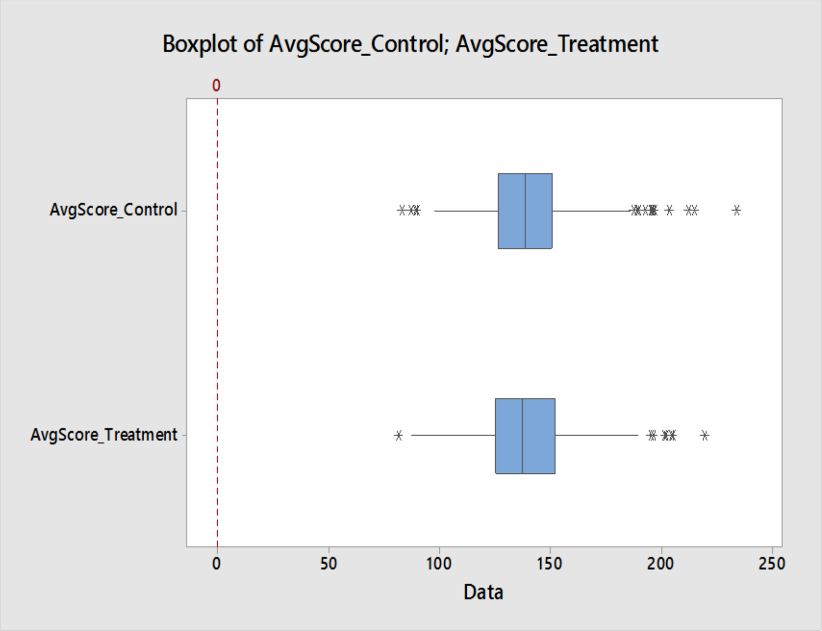
Table:

Descriptive Statistics: AvgScore_Control-Female; ... core_Control-Male

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
AvgScore_Control-Female	132	0	139.03	1.93	22.14	89.23	124.38	137.85	148.00
AvgScore_Control-Male	132	0	123.60	0.946	10.87	82.73	117.86	125.31	131.79

Variable	Maximum
AvgScore_Control-Female	233.83
AvgScore_Control-Male	137.81



Figures of boxplot shows that the boxplots overlap for both Control and Treatment groups which shows the signs of difference and the outliers are there with the possibility to make a diligent about the reliability of this value. Confirming the general idea that in the control group there is a difference between Early and After pregnancy because the boxplot far way for the zero. To confirm this, we must go for formal Pair t test within Control and Treatment group

Table 1: Paired t test of control male and female and Treatment male and female.

Estimation for Paired Difference	**Estimation for Paired Difference**
Mean StDev SE Mean 95% CI for μ_difference	Mean StDev SE Mean 95% CI for μ_difference
15.43 13.94 1.21 (13.03; 17.83)	16.28 14.97 1.32 (13.67; 18.90)
μ_difference: mean of (AvgScore_Control-Female - AvgScore_Control-Male) **Test**	*μ_difference: mean of (AvgScore_Treatment-Female - AvgScore_Treatment-Male)*

Null hypothesis H_0: μ_difference = 0 Alternative hypothesis H_1: μ_difference ≠ 0 T-Value P-Value 12.72 0.000	**Test** Null hypothesis H_0: μ_difference = 0 Alternative hypothesis H_1: μ_difference ≠ 0 T-Value P-Value 12.31 0.000

Our p value for control and treatment groups is .001 which is more less .05 which indicates strong evidence against the null hypothesis, so we reject the null hypothesis. That shows, Female of control and Treatment group has less change of getting the CHD. To investigate the difference of risk profile difference between the control and treatment

Group we chose for the subjective analysis.

Figure 2: Scatter plot of control and treatment group comparing the AvgScore.

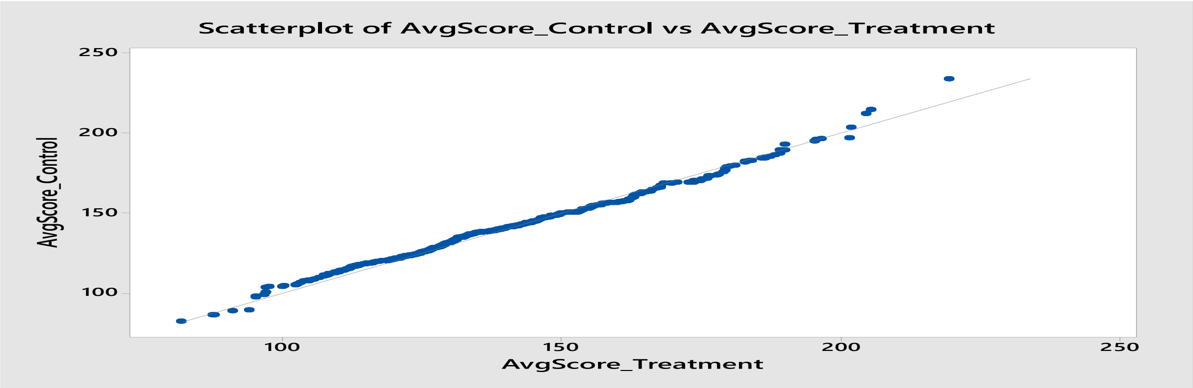


Figure shows that there is positive relation between control and treatment group which shows the evidence of no difference in risk profile between control and treatment. To make sure that our assumption is true we have carry out the formal analysis (2 sample t test).

Table 2:

```
         Two-Sample T-Test and CI: Control and Treatment
AvgScore

 μ₁: mean of AvgScore when Arm = Control

 μ₂: mean of AvgScore when Arm = Treatment

 Difference: μ₁ - μ₂
Equal variances are not assumed for this analysis.

Descriptive Statistics: AvgScore

 Arm          N   Mean  StDev  SE Mean
──────────────────────────────────────
 Control    429  140.2   21.0      1.0

 Treatment  429  139.4   22.2      1.1
Estimation for Difference

               95% CI for
 Differences   Difference
────────────────────────────
        0.89  (-2.01; 3.79)


Test

 Null hypothesis          H₀: μ₁ - μ₂ = 0

 Alternative hypothesis   H₁: μ₁ - μ₂ ≠ 0

 T-Value   DF  P-Value
──────────────────────
    0.60  853    0.549
```

Formal analysis shows that our 95% CI contain zero however the mean difference is positive which shows that risk profile between control and treatment does not differs. p-value (>0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.

2: Age, Gender Vs AvgScore.

Like gender age is also which makes a remarkable difference while investigate the risk profile of CHD. Correlation coefficient shows the positive equation so linear regression analysis is performed to find where the response variable will the manual calculated AvgScore dependent on the Age and Gender.

Table 3:

Analysis of Variance

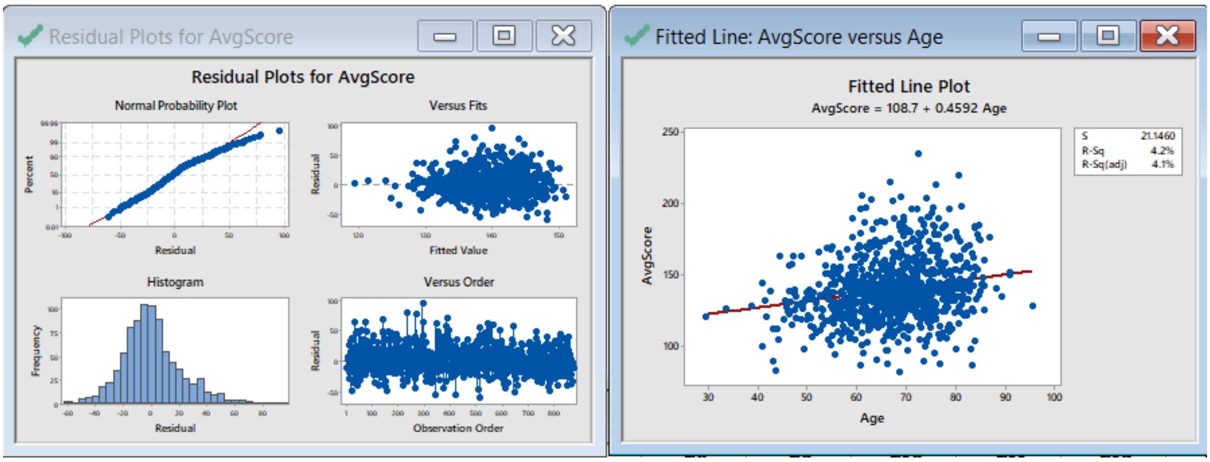
Source	DF	Adj SS	Adj MS	F-Value	**P-Value**
Regression	2	18630	9315.2	20.90	**0.000**
Age	1	17908	17908.3	40.18	**0.000**
Gender	1	1739	1739.2	3.90	**0.049**
Error	867	386389	445.7		
Lack-of-Fit	528	238328	451.4	1.03	**0.372**
Pure Error	339	148061	436.8		
Total	869	405019			

Table 4:

Coefficients

Term	**Coef**	SE Coef	T-Value	**P-Value**	VIF
Constant	**105.37**	5.35	19.69	**0.000**	
Age	**0.4758**	0.0751	6.34	**0.000**	1.01
Gender					
Male	**3.11**	1.57	1.98	**0.049**	1.01

Figure 3:



The histogram is roughly bell-shaped, so it is an indication that it is reasonable to assume that the residuals have a normal distribution. The pattern of the normal probability plot is straight, so this plot also provides evidence that it is reasonable to assume that the errors have a normal distribution.

In fits vs residual plot the variance is roughly the same all the way across and there are no worrisome patterns. There seems to be no difficulties with the model or data.

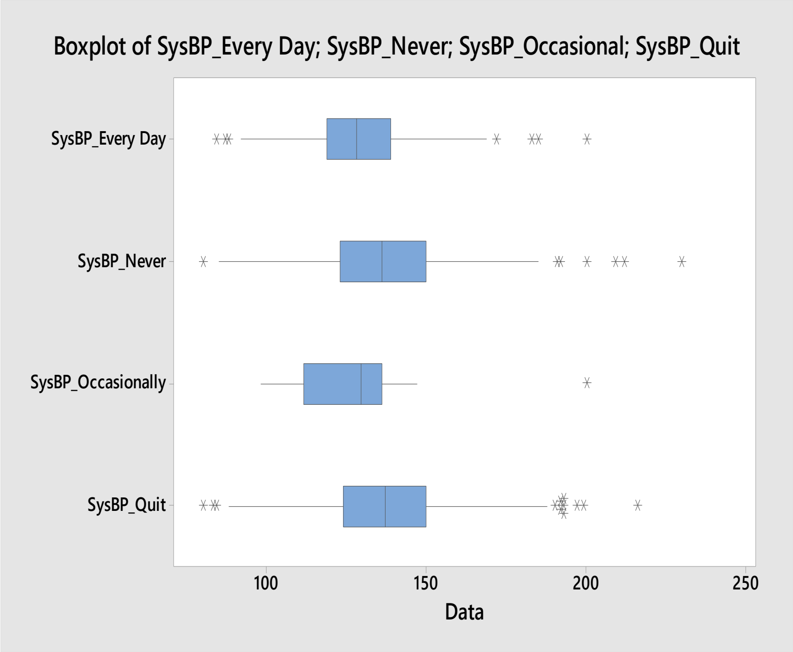
From order vs residual graph, it’s clear that the order of the data doesn’t give any information, so the sample is independent. There is no evidence of a bend in the fits vs residual plot, so we can assume that the model is linear.

And the graphs show that there is positive relation between the AvgScore and age which indicates that as age increase the risk profile of CHD will be strong in higher age group and female are comparatively has less risk with male with increasing age. Coefficient value of Gender and Age versus AvgScore is positive which shows a strong positive impact and provide evidence the risk profile increase as the age increase, while categorising the gender risk profile is high in Men comparing to women which can be confirm with the p value that is almost equal to .05.

3: Smoking.Status and SysBp

Smoking.Status is directly proportional to Sysbp which is a major factor to create the risk profile among the population, so comparison is done of smoking status with Sysbp through boxplots following the One Way Anova and Tukey Comparison.

Figure 4: Box plot **Figure 5: Tukey Comparison**



Descriptive Statistics: SysBP_Every Day; SysBP_Never; ... lly; SysBP_Quit

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
SysBP_Every Day	107	0	131.17	2.00	20.72	84.00	119.00	128.00	139.00	200.00

SysBP_Never	244	0	137.72	1.38	21.55	80.00	123.00	136.00	149.75	230.00
SysBP_Occasionally	22	0	128.09	4.52	21.18	98.00	111.50	129.50	136.00	200.00
SysBP_Quit	485	0	137.79	0.975	21.47	80.00	124.00	137.00	150.00	216.00

Boxplot for smoking status overlaps for all four group (Never, Quit, every day, Occasionally) and presence of outliers shows the weak proof for relying on this outcome however median line of all four pass through between the boxplots. But there is significant difference between the largest and smallest value of mean which supports to go for formal analysis.

Table 5:

```
        One-way ANOVA: SysBP versus Smoking.Status

 Method

 Null hypothesis          All means are equal
 Alternative hypothesis   Not all means are equal
 Significance level       α = 0.05
Equal variances were assumed for the analysis.

Factor Information

 Factor          Levels  Values
─────────────────────────────────────────────────────
 Smoking.Status       4  Every Day; Never; Occasionally; Quit

Analysis of Variance

 Source           DF   Adj SS   Adj MS  F-Value  P-Value
───────────────────────────────────────────────────────
 Smoking.Status    3     5736   1912.0     4.18    0.006
 Error           854   390861    457.7
```

Total	857	396597

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
21.3935	1.45%	1.10%	0.53%

Means

Smoking.Status	N	Mean	StDev	95% CI
Every Day	107	131.17	20.72	(127.11; 135.23)
Never	244	137.72	21.55	(135.04; 140.41)
Occasionally	22	128.09	21.18	(119.14; 137.04)
Quit	485	137.789	21.472	(135.882; 139.696)

Pooled StDev = 21.3935

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

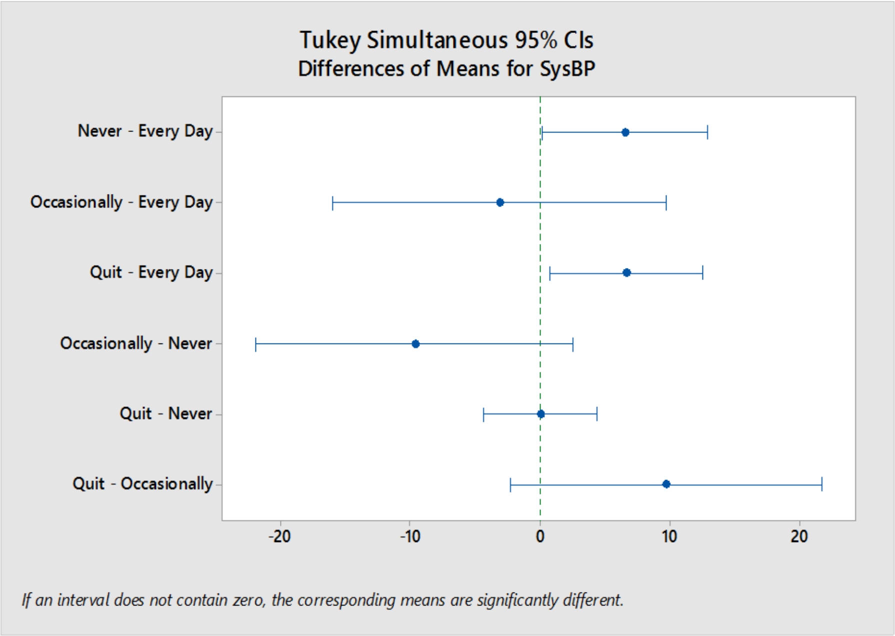
Smoking.Status	N	Mean	Grouping	
Quit	485	137.789	A	
Never	244	137.72	A	
Every Day	107	131.17		B
Occasionally	22	128.09	A	B

Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
Never - Every Day	6.56	2.48	(0.19; 12.92)	2.64	0.041
Occasionally - Every Day	-3.08	5.01	(-15.93; 9.78)	-0.61	0.928
Quit - Every Day	6.62	2.28	(0.76; 12.49)	2.90	0.020
Occasionally - Never	-9.63	4.76	(-21.86; 2.59)	-2.02	0.179
Quit – Never	0.06	1.68	(-4.25; 4.37)	0.04	1.000
Quit - Occasionally	9.70	4.66	(-2.27; 21.67)	2.08	0.160

Individual confidence level = 98.96%



If an interval does not contain zero, the corresponding means are significantly different.

Analysis of variance shows since the p value from the one-way ANOVA (p=0.006) is much smaller than 0.005, shows the strong evidence to reject the null hypothesis.

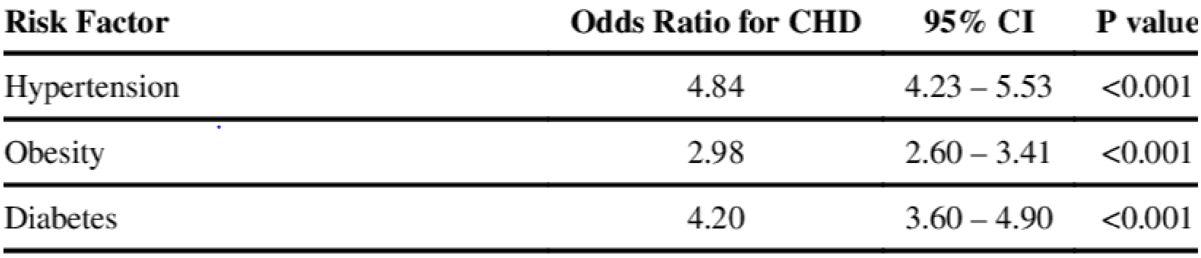
As we can conclude some difference among the Sysbp, it is required to identify that which Smoking status impacts are significantly different for risk score of CHD. This is achieved by using a multiple comparisons procedure for the difference in population mean of Sysbp between each pair of Smoking.Status.

Tukey comparison, shows that the group which quit the smoking shows the decreases in risk profile comparing to the everyday group and same is for Quit and everyday group as the p value is less than .05 and confidence interval contains the positive value.

Secondary Endpoint Analysis:

Our considered secondary endpoint for this study Previous.MI and Hypertension and Diabetes and according to the previous conducted research, these endpoint shows the relevant difference in negative directions in case of risk scoring of CHD among adult males and females which is the very most reason to include these in our study.

Table 7: - Odd ratio of FRS scores among adults.



Risk Factor	Odds Ratio for CHD	95% CI	P value
Hypertension	4.84	4.23 – 5.53	<0.001
Obesity	2.98	2.60 – 3.41	<0.001
Diabetes	4.20	3.60 – 4.90	<0.001

1: Previous.Mi

We carried out analysis first pie chart comparison between control and treatment which is followed by a formal analysis which took place as 2 proportion tests. To carry out the analysis research, the missing data is removed, and test is performed on filtered data for more accuracy in result.

Before carrying out the 2 proportion, tally cross tabulation chi square was conducted to evaluate the individual percentage.

Note: For pie chart and table for cross tabulation result please, follow the appendix.

Table- 8:

Test and CI for Two Proportions: Control & Treatment

Method

Event: Previous.MI_Control = Yes

p_1: proportion where Previous.MI_Control = Yes and Previous.MI_Treatment = No

p_2: proportion where Previous.MI_Control = Yes and Previous.MI_Treatment = Yes

Difference: p_1 - p_2

Descriptive Statistics: Previous.MI_Control

Previous.MI_Treatment	N	Event	Sample p
No	215	106	0.493023
Yes	214	109	0.509346

Estimation for Difference

Difference	95% CI for Difference
-0.0163225	(-0.110938; 0.078293)

CI based on normal approximation

Test

Null hypothesis	H_0: $p_1 - p_2 = 0$
Alternative hypothesis	H_1: $p_1 - p_2 \neq 0$

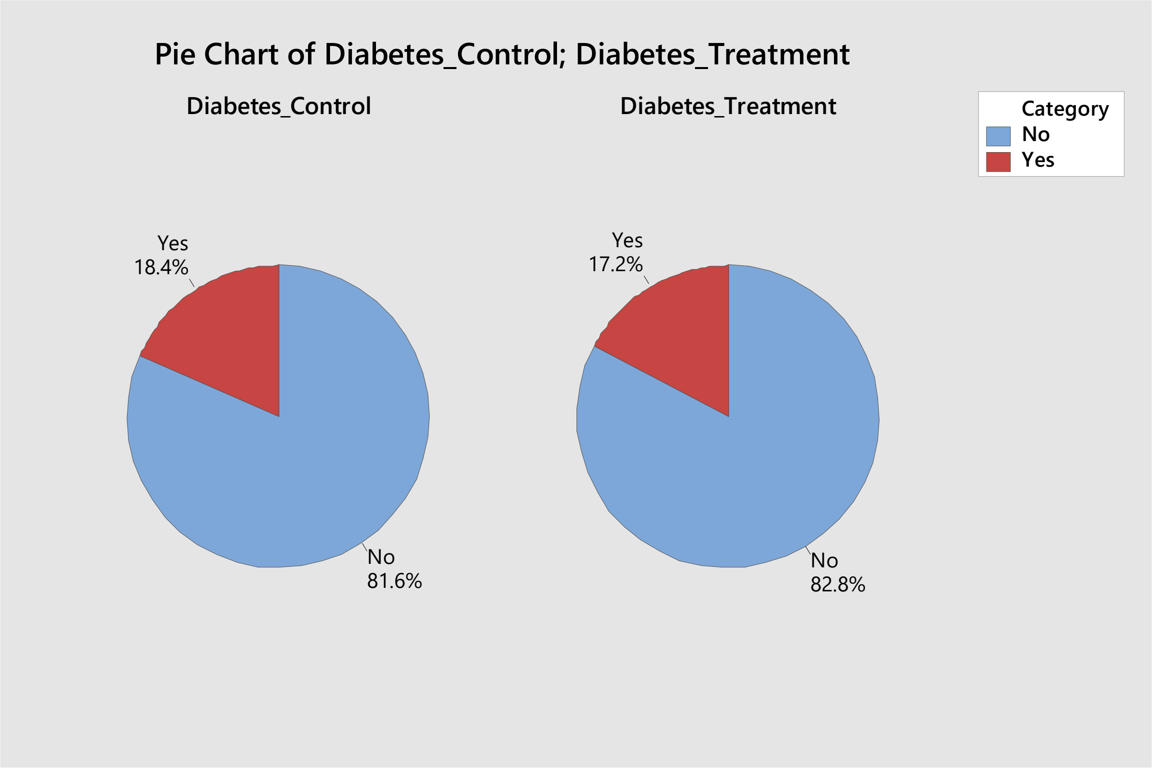
Method	Z-Value	***P-Value***
Normal approximation	-0.34	***0.735***
Fisher's exact		***0.772***

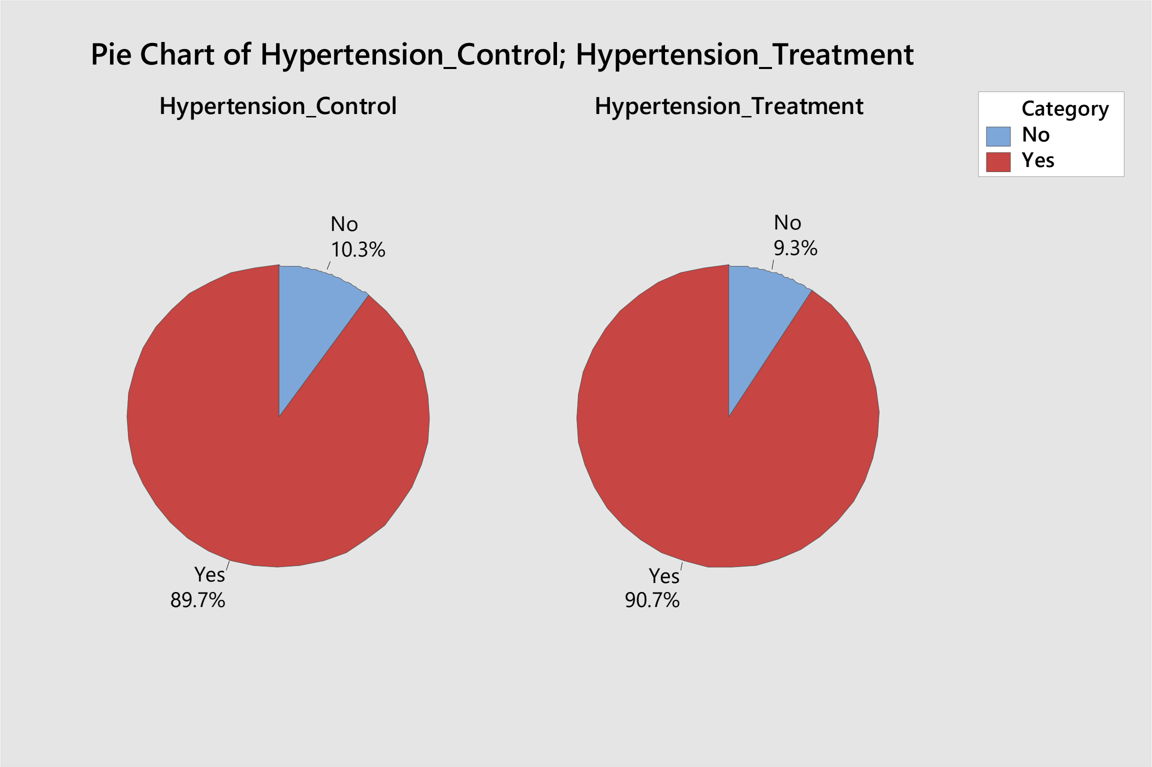
The 2 proportion result the weak evidence to reject the null hypothesis due to our 95% confidence interval between 7% to 11% and does contain Zero and our p>.05. Hence it indicates that there is not significant difference between the previous history of control and treatment.

2: Hypertension and Diabetes

Analysis was carried out on Hypertension and Diabetes. Subjective analysis of Pie chart and cross Chi square tabulation which was followed by a formal analysis of 2 proportion for both second baseline endpoint.

Figure: Pie Chart of control and treatment for Diabetes and Hypertension.





Note: Kindly follow appendix for the descriptive analysis.

Table 9:CI and Fisher exact p value evaluation for Hypertension

Test and CI for Two Proportions: ... _Treatment; Hypertension_Control

Method

Event: Hypertension_Treatment = Yes

p_1: proportion where Hypertension_Treatment = Yes and Hypertension_Control = No

p_2: proportion where Hypertension_Treatment = Yes and Hypertension_Control = Yes

Difference: p_1 - p_2

Descriptive Statistics: Hypertension_Treatment

Hypertension_Control	N	Event	Sample p
No	44	39	0.886364
Yes	385	350	0.909091

Estimation for Difference

Difference	95% CI for Difference
-0.0227273	(-0.120800; 0.075346)

CI based on normal approximation

Test

Null hypothesis	H_0: $p_1 - p_2 = 0$	
Alternative hypothesis	H_1: $p_1 - p_2 \neq 0$	

Method	Z-Value	P-Value
Normal approximation	-0.45	0.650
Fisher's exact		0.586

Table 10: CI and Fisher exact p value evaluation for Diabetes

Test and CI for Two Proportions: Diabetes_Control; ... betes_Treatment

Method

Event: Diabetes_Control = Yes

p_1: proportion where Diabetes_Control = Yes and Diabetes_Treatment = No

p_2: proportion where Diabetes_Control = Yes and Diabetes_Treatment = Yes

Difference: $p_1 - p_2$

Descriptive Statistics: Diabetes_Control

Diabetes_Treatment	N	Event	Sample p
No	355	68	0.191549
Yes	74	11	0.148649

Estimation for Difference

Difference	95% CI for Difference
0.0429006	(-0.047903; 0.133704)

CI based on normal approximation

Test

Null hypothesis	H_0: $p_1 - p_2 = 0$	
Alternative hypothesis	H_1: $p_1 - p_2 \neq 0$	

Method	Z-Value	P-Value
Normal approximation	0.93	0.354
Fisher's exact		0.509

Formal analysis for Hypertension and diabetes between control and treatment group has weak evidence against evidence null hypothesis due to fisher exact p value which is much bigger that .05 and 95% confidence intervals does contain zero. So, our analysis proves that Hypertension and Diabetes does not impact significantly the risk profile between the control and treatment. Primary and secondary analysis shows the difference respective the baseline point which should be closely examine while interpretation of results to ensure the accuracy of result for expected and fair results.**Result**

There was a statistically significant positive correlation between Avgscores and core parameters (age, gender) and the severity of CHD according to the $p < 0.05$. Also, there was no statistically significant difference between Hypertension, Diabetes and Previous. MI according to $p > 0.05$ in Control and Treatment arm. On the other hand, there was no statistically significant difference between Avgscore of Control and Treatment as the $p > 0.05$. Wherein paired t shows that there is significant difference between the Male and Female population of Control and Treatment Arm and female population has less risk score comparing to men's according to 95% CI and p value. Smoking was also an important factor in the difference between both groups as the population who quit and who smoke Every day has significant.

Conclusion

There is no significant between the treatment and control population among risk score however the Female of control group has 13% to 16% less change gets the CHD comparing to men wherein in treatment Females has 14 % to 19% less change to get the CHD comparing to men. Risk scores increases as the age increase in males of control and treatment group, including this population is quit or never smoking has less chances of developing the CHD.

APPENDIX A

DATA Manipulation Using R

Datasets used for the study were analysed using the R language. And while analysing it was observed that it contains the missing value for several fields.

Following are the steps which has been performed to filter the data before any statistical analysis.

Selection of column for analysis.

Replace NA with mean value for Continuous columns

Removing NA for categorical Column

Calculate Risk Score adding cholesterol, Sysbp and Subtract HDL.

Round Off them score to zero decimal

Now Filter AvgScore by Male and Female.

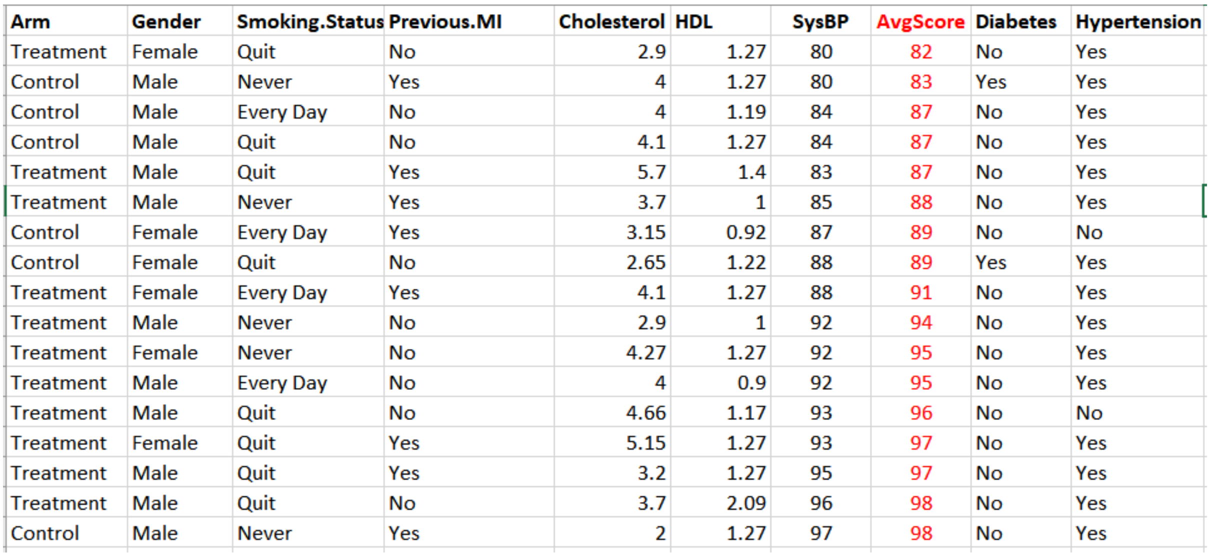
Refer to below mentioned code

```
#installing required package.

install.packages("dplyr")

library(dplyr)

#Reading the dataset

a<-read.csv("Sphere.csv")

#Printing summary to know mikssing data
```

```
summary(a)

#Selecting the desired data set to work on

a <-a %>%

  select(Arm,Gender,   Smoking.Status,        Previous.MI, Cholesterol,       HDL, SysBP,
           Diabetes,        Hypertension)

# Removing Missing value for cetegorical data

a<-a[!is.na(a$Smoking.Status), ]

a<-a[!is.na(a$Diabetes), ]

summary(a)

# Replacing na with mean for continous data



A$Cholesterol[which(is.na(A$Cholesterol))] <- mean(A$Cholesterol, na.rm = TRUE)

A$HDL[which(is.na(A$HDL))] <- mean(A$HDL, na.rm = TRUE)

A$SysBP[which(is.na(A$SysBP))] <- mean(A$SysBP, na.rm = TRUE)



#Box Plot to know the outlier for the continous data

boxplot(A$Cholesterol)

boxplot(A$HDL)

boxplot(A$SysBP)
```

```
#Calculating the Avgscore from continouos data.

a$Average<-a$SysBP + a$Cholesterol - a$HDL

print(a)
```

Filtered data for Analysis.



Arm	Gender	Smoking.Status	Previous.MI	Cholesterol	HDL	SysBP	AvgScore	Diabetes	Hypertension
Treatment	Female	Quit	No	2.9	1.27	80	82	No	Yes
Control	Male	Never	Yes	4	1.27	80	83	Yes	Yes
Control	Male	Every Day	No	4	1.19	84	87	No	Yes
Control	Male	Quit	No	4.1	1.27	84	87	No	Yes
Treatment	Male	Quit	Yes	5.7	1.4	83	87	No	Yes
Treatment	Male	Never	Yes	3.7	1	85	88	No	Yes
Control	Female	Every Day	Yes	3.15	0.92	87	89	No	No
Control	Female	Quit	No	2.65	1.22	88	89	Yes	Yes
Treatment	Female	Every Day	Yes	4.1	1.27	88	91	No	Yes
Treatment	Male	Never	No	2.9	1	92	94	No	Yes
Treatment	Female	Never	No	4.27	1.27	92	95	No	Yes
Treatment	Male	Every Day	No	4	0.9	92	95	No	Yes
Treatment	Male	Quit	No	4.66	1.17	93	96	No	No
Treatment	Female	Quit	Yes	5.15	1.27	93	97	No	Yes
Treatment	Male	Quit	Yes	3.2	1.27	95	97	No	Yes
Treatment	Male	Quit	No	3.7	2.09	96	98	No	Yes
Control	Male	Never	Yes	2	1.27	97	98	No	Yes

APPENDIX B

This appendix is used to support the statistical analysis which was performed on the sphere data to investigate the difference between two arms control and treatment.

Figure 1: Pie chart of Control and Treatment for Previous.MI

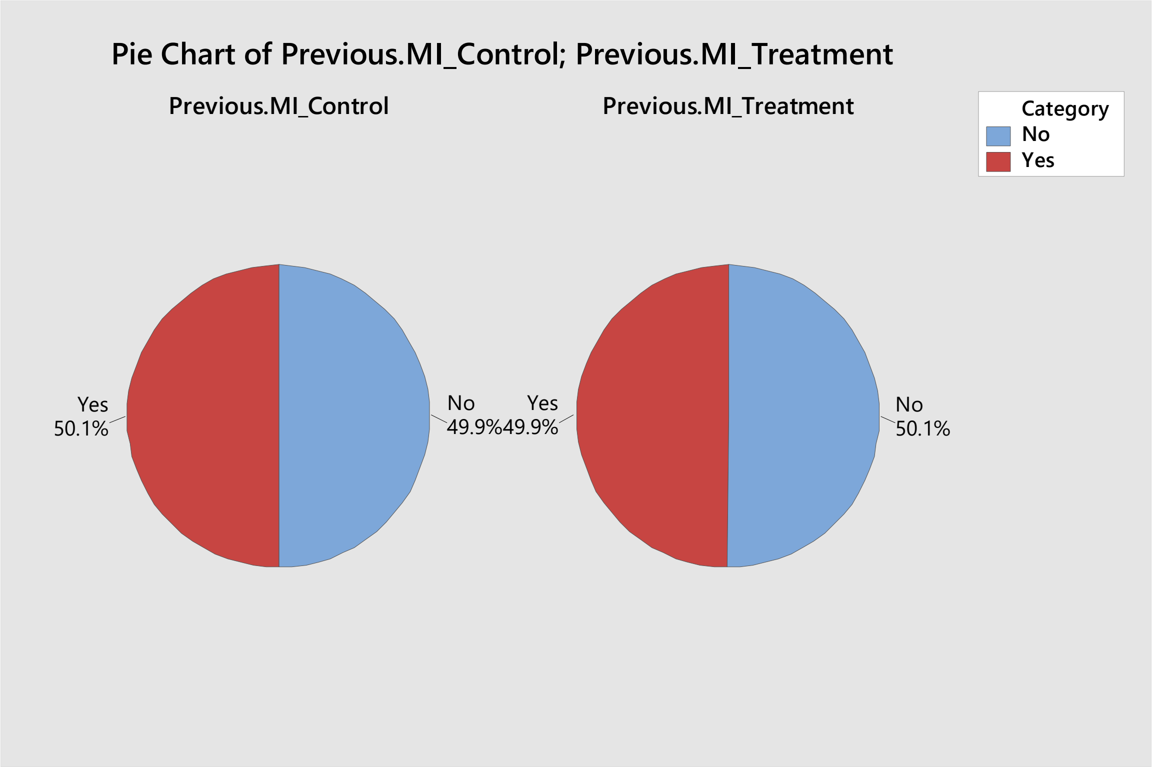


Table 1:

Tabulated Statistics: Previous.MI_Control; Previous.MI_Treatment
Rows: Previous.MI_Control   Columns: Previous.MI_Treatment

	No	Yes	All
No	109	105	214
	50.93	49.07	100.00
Yes	106	109	215
	49.30	50.70	100.00
All	215	214	429

50.12 49.88 100.00

Figure2: Scatterplot of AvgScore for Female and Male within Control group

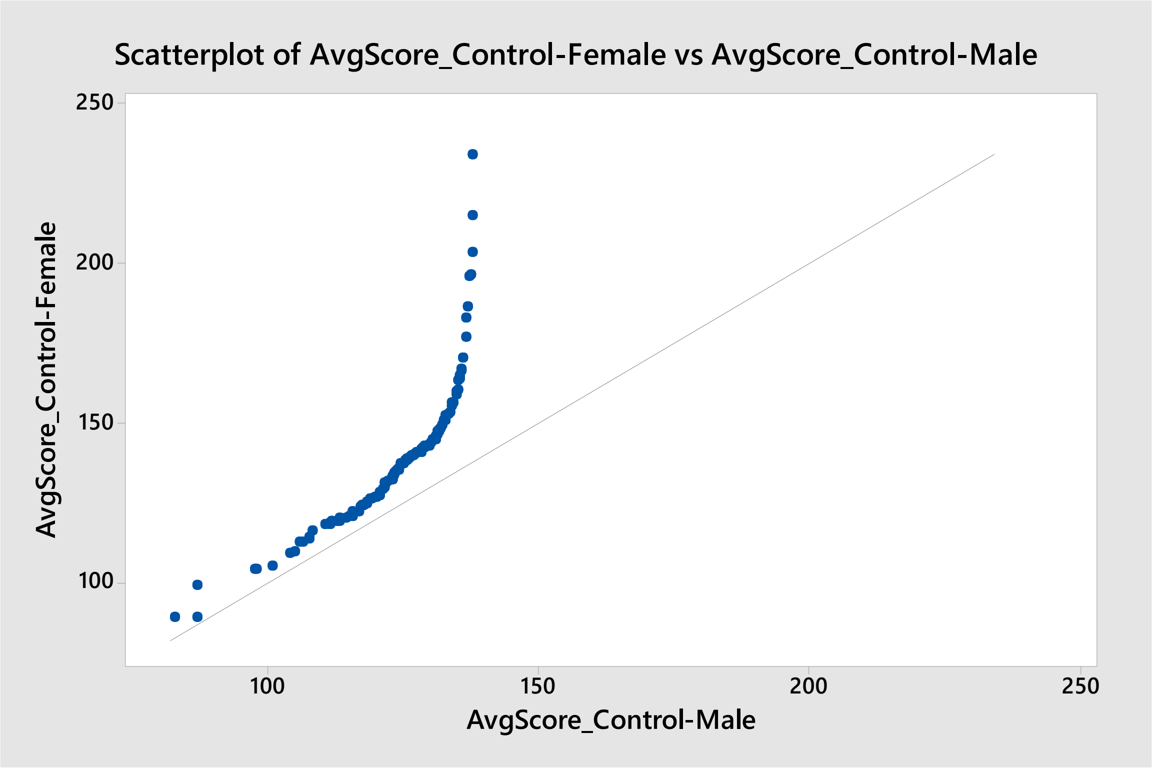


Figure2: Scatterplot of AvgScore for Female and Male within Treatment group

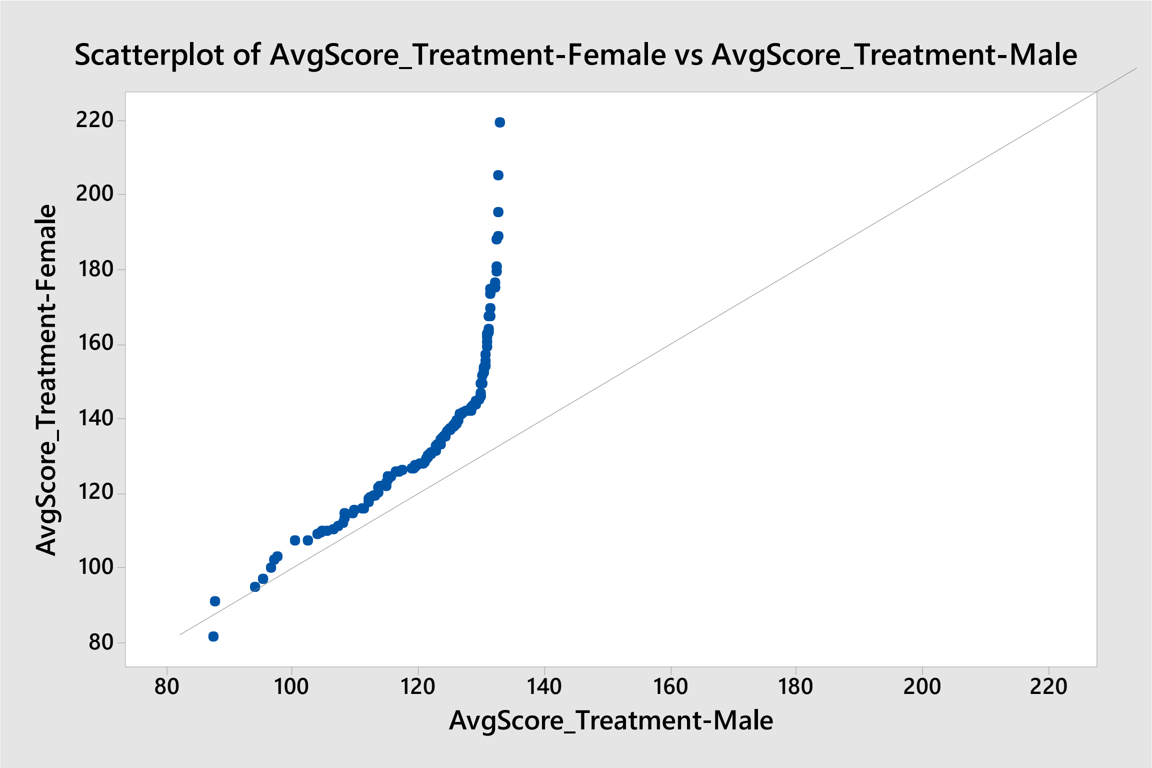


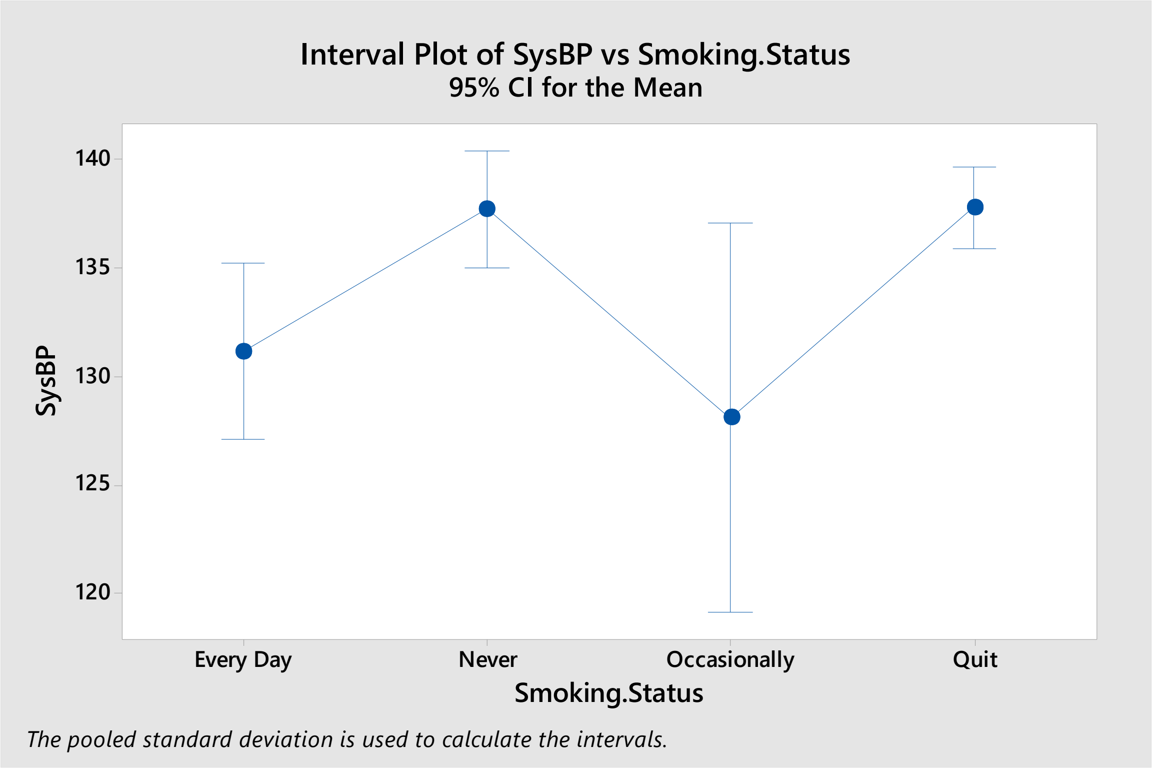
Table 2:

Descriptive Statistics: AvgScore_Control; AvgScore_Treatment

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
AvgScore_Control	429	0	140.25	1.02	21.04	82.73	126.27	138.83	150.71	233.83
AvgScore_Treatment	429	0	139.36	1.07	22.22	81.63	125.55	137.20	152.12	219.40

Figure 3: Interval plot for Anova of Smoking History.



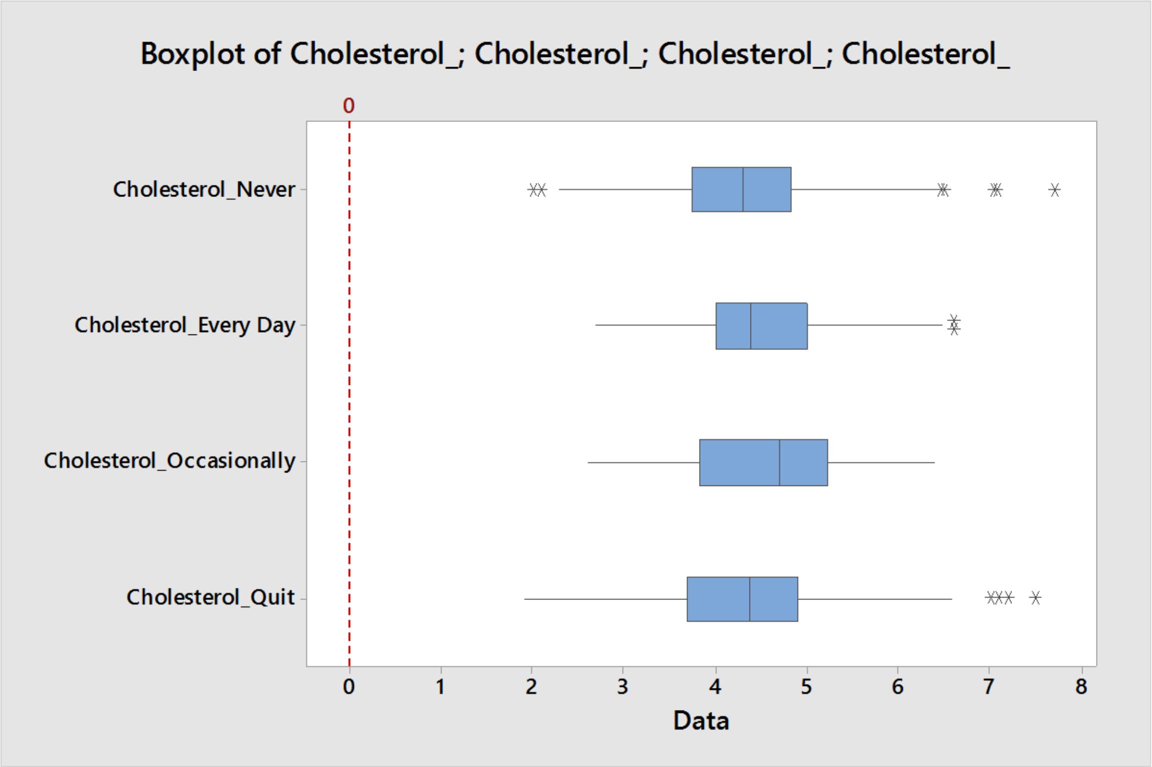
Interval Plot of SysBP vs Smoking.Status

95% CI for the Mean

The pooled standard deviation is used to calculate the intervals.

ONEWAY ANOVA BETWEEN SMOKING.STATUS AND CHOLESTROL.

Figure 4 : Box Plot



Descriptive Statistics: Cholesterol_Every Day; ... nally; Cholesterol_Quit
Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Cholesterol_Every Day	107	0	4.5233	0.0812	0.8396	2.7000	4.0000	4.4000	5.0000
Cholesterol_Never	244	0	4.3611	0.0584	0.9125	2.0000	3.7475	4.3000	4.8225
Cholesterol_Occasionally	22	0	4.565	0.209	0.980	2.600	3.838	4.700	5.225
Cholesterol_Quit	485	0	4.3443	0.0409	0.9001	1.9000	3.7000	4.3700	4.9000

Variable	Maximum
Cholesterol_Every Day	6.6000
Cholesterol_Never	7.7000
Cholesterol_Occasionally	6.400
Cholesterol_Quit	7.5000

One-way ANOVA: Cholesterol versus Smoking.Status
Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	α = 0.05

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Smoking.Status	4	Every Day; Never; Occasionally; Quit

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Smoking.Status	3	3.644	1.2147	1.50	0.212
Error	854	689.370	0.8072		
Total	857	693.014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.898457	0.53%	0.18%	0.00%

Means

Smoking.Status	N	Mean	StDev	95% CI
Every Day	107	4.5233	0.8396	(4.3528; 4.6937)
Never	244	4.3611	0.9125	(4.2483; 4.4740)
Occasionally	22	4.565	0.980	(4.189; 4.941)
Quit	485	4.3443	0.9001	(4.2642; 4.4243)

Pooled StDev = 0.898457

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Smoking.Status	N	Mean	Grouping
Occasionally	22	4.565	A
Every Day	107	4.5233	A
Never	244	4.3611	A
Quit	485	4.3443	A

Means that do not share a letter are significantly different.

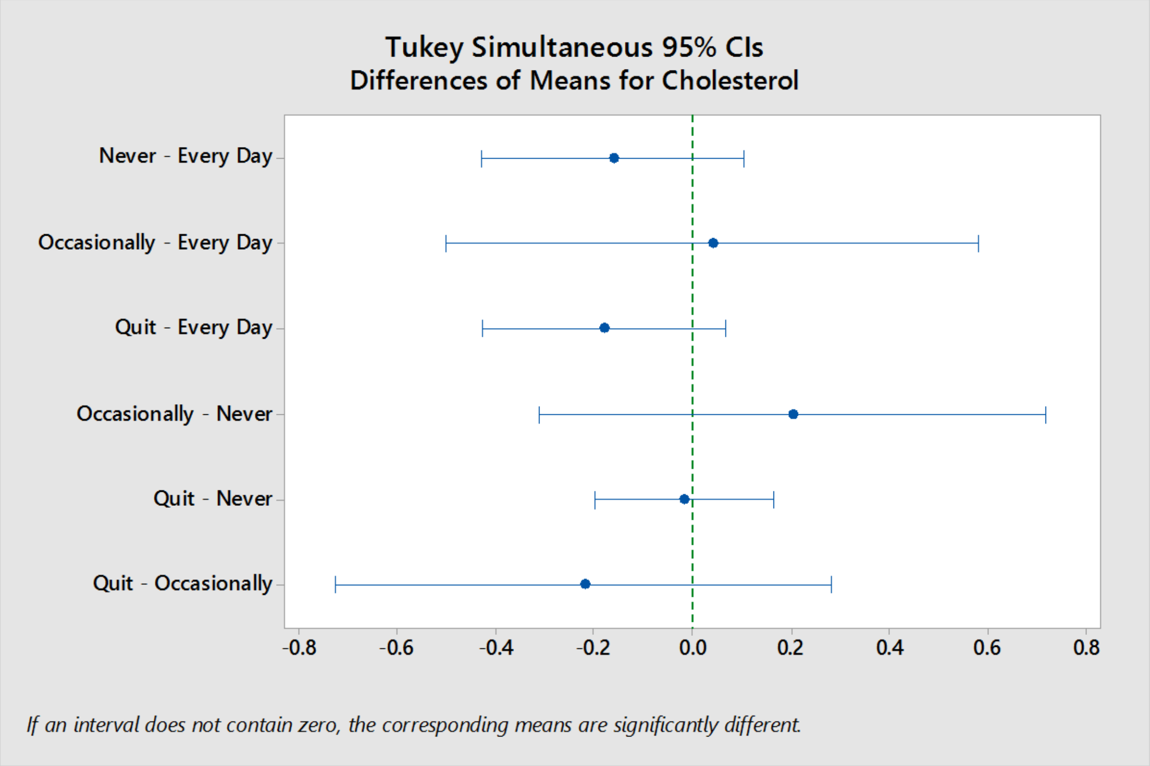
Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
Never - Every Day	-0.162	0.104	(-0.430; 0.105)	-1.56	0.404
Occasionally - Every Day	0.041	0.210	(-0.499; 0.581)	0.20	0.997
Quit - Every Day	-0.1790	0.0960	(-0.4253; 0.0673)	-1.87	0.243

Occasionally - Never	0.203	0.200	(-0.310; 0.717)	1.02	0.740
Quit - Never	-0.0169	0.0705	(-0.1979; 0.1641)	-0.24	0.995
Quit - Occasionally	-0.220	0.196	(-0.723; 0.282)	-1.12	0.674

Individual confidence level = 98.96%

Tukey Simultaneous 95% Cis


Tukey Simultaneous 95% CIs
Differences of Means for Cholesterol
Never - Every Day
Occasionally - Every Day
Quit - Every Day
Occasionally - Never
Quit - Never
Quit - Occasionally
If an interval does not contain zero, the corresponding means are significantly different.