Regular Marijuana Usage and its Influence on a Sedentary Lifestyle

By: Ankur Bambhrolia

Background/Motivation for the Study

The main proponent of marijuana is tetrahydrocannabinol (THC) which interacts with the cannaboid receptors of the brain. This pathway is responsible for regulating "learning and memory, emotional processing, sleep, temperature control, pain control, inflammatory and immune responses, and eating" (Grinspoon 2021). THC acts upon these receptors which is how it gives users the "munchies", relieves pain, and promotes habit formation.

The main premise for the legalization of marijuana is that the drug is not inherently addictive, and that marijuana has medical applications unlike legalized drugs such alcohol and cigarettes. However, due to the popularity of the positive effects, proponents seem to diminish the negative aspects of marijuana, and propose that the usage of marijuana does not significantly alter one's quality of life.

Quality of life is a catch all term, and it is easier to break it down into three main components: physical health, psychological health, and social health. Previous studies have shown that there is a negative correlation between marijuana usage and the user's quality of life.

Physically, recreational use of marijuana has led to questions about the amount of inhaled THC, and the purity of the THC smoked. This in turn leads to "coughing, increased sputum or phlegm, wheezing, shortness of breath and inflammation of lung tissue" in smokers, which tends to go away once drug usage stops (Neff et al. 2021). While marijuana use is not directly linked to increased heart disease in the same way nicotine is, there is some evidence that links cannabis use to the development of heart disease (Sydney 2002). Furthermore, marijuana also causes "increased heart rates" and "intense nausea and vomiting" (NDIA 2019). Lastly, marijuana increases appetite, but also serves as a protective factor against obesity (Brook et al. 2003).

Psychologically and socially, marijuana has a myriad of negative effects. TCH alters brain chemistry and brain development. Symptoms include changes in mood, hallucinations, impaired memory, and difficulty with critical thinking skills (NDIA 2019). As aforementioned, marijuana promotes habit development, making it harder for user to make significant changes to their lives. Additionally, as users build up a tolerance of THC, users begin to smoke more, worsening the physical and psychological effects. Due to the brain impairments from THC, studies have indicated that those who smoke are less likely to be less educated, have a lower income and are more likely to have a workplace accident than their non-smoking counterparts (NDIA 2023).

Research Question and Hypothesis

I would like to see if regular marijuana users live a more sedentary lifestyle compared to non-regular marijuana users using BMI, heart rate and physical activity as indicators of an active lifestyle and general health. I anticipate that regular marijuana users have a higher BMI, pulse and are less physically active than their non-regular smoking counterparts. In short, there are 3 hypothesis to test:

1. The association of Regular Marijuana Usage and Physical Activity
    a. H_0 : There is no association between Regular Marijuana Usage and Physical Activity
    b. H_A : There is an association between Regular Marijuana Usage and Physical Activity
2. Regular Marijuana BMI mean = Non-Regular Marijuana BMI mean
    a. $H_0 : \mu_{RegularBMI} - \mu_{NonRegularBMI} = 0$
        i. Null Hypothesis: There is no difference in the mean BMI of regular marijuana users and non-regular marijuana users.
    b. $H_A : \mu_{RegularBMI} - \mu_{NonRegularBMI} \neq 0$
        i. Alternative Hypothesis: There is a difference in the mean BMI of regular marijuana users and non-regular marijuana users.
3. Regular Marijuana Pulse mean = Non-Regular Marijuana Pulse mean
    a. $H_0 : \mu_{RegularPulse} - \mu_{NonRegularPulse} = 0$
        i. Null Hypothesis: There is no difference in the mean pulse (Heart Rate) of regular marijuana users and non-regular marijuana users.
    b. $H_A : \mu_{RegularPulse} - \mu_{NonRegularPulse} \neq 0$
        i. Alternative Hypothesis: There is a difference in the mean BMI of regular marijuana users and non-regular marijuana users.

Data Description and Exploratory Data Analysis

The data analyzed was taken by the National American Nutrition Examination Surveys (NHANES) from 2009-2012. 76 variables were measured across 10,000 patients.

Below are the list of variables of interest in answering the research question posed above:

1. *NHANERMJ*:
    a. The NHANES table filtered to only include those who answered "Yes" or "No" to regularly smoking Marijuana which is the "predictor" variable. "*Yes*" means that a participant has been/is smoking marijuana regularly. "*No*"

means that a participant has not smoked marijuana regularly. *Regularly* is defined as smoking for a minimum once a month over the course of a year.

b. This is the population of interest. All "NA" values were removed from the table and a result, 4941 participants remain of the original 10,000. The age range for this population is 18-59 years of age.

c. This table will be filtered further into "YRMJ" and "NRMJ". "YRMJ" only includes data in which people responded "yes" to regularly smoking marijuana. "NRMJ" only includes data in which people responded "No" to regularly smoking marijuana.

2. *BMI*: Body Mass Index (kg/m^2). A general indicator of health.

3. *Pulse*: Heart Rate measured in beats per minutes (BPM).

4. *PhyActive*: Participant engages in moderate or vigorous-intensity sports, fitness or recreational activities (Yes or No)*
   The age range for the participants in this filtered sample is 18-59 years of age.

Physical Activity

```
library(NHANES)
library(dplyr)

library(colorspace)
library(tidyr)


#Filtering Data by Regular Marijuana Users
NHANESRMJ <- NHANES %>% drop_na(RegularMarij)
NHANESRMJ$RegularMarij <- factor( NHANESRMJ$RegularMarij, levels=c('Yes', 'No
'))
NHANESRMJ$PhysActive <- factor(NHANESRMJ$PhysActive,levels=c('Yes', 'No'))
```

First we remove any "NA" values for the "RegMarij" column leaving us with a sample of 4941 people, which will serve as our sample size moving forward.

```
#Contigency Table of people who are physically active grouped by regular mari
juana usage

physical_active_table <- addmargins(table(NHANESRMJ$RegularMarij,NHANESRMJ$Ph
ysActive,
                  dnn = c("Regular Marijuana Use","Physically Active")))
```

```
physical_active_table
```

```
                     Physically Active
Regular Marijuana Use  Yes   No  Sum
                  Yes  723  643 1366
                  No  2128 1447 3575
                  Sum 2851 2090 4941
```

```
#Proportion of people who are physically active and not physically active res
pectively:

2851/4941 #Physically Active
```

```
[1] 0.5770087
```

```
2090/4941 #Not Physically Active
```

```
[1] 0.4229913
```

```
#Proportion of people who smoke marijuana regularly and do not smoke marijuan
a regularly repectively:

1366/4941 #Regular Smokers
```

```
[1] 0.2764623
```

```
3575/4941 #Not Regular Smokers
```

```
[1] 0.7235377
```

```
#Proportion of people who regularly smoke and are physically active:

723/1366
```

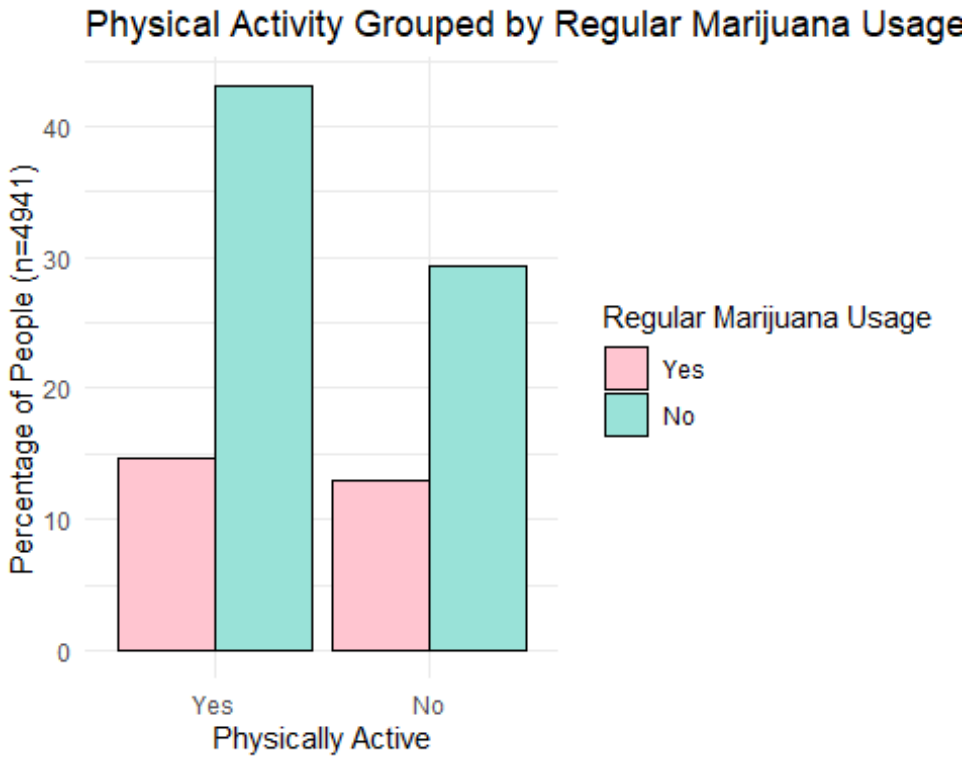
```
[1] 0.5292826
```

```
#Proportion of people who regularly smoke and are not physically active:
643/1366
```

```
[1] 0.4707174
```

```
#Releative Frequency of Physically active people grouped together by regular
smoking status (n=4941).
NHANESRMJ %>%
  ggplot(aes(x=PhysActive, y=after_stat(count/sum(count))*100, fill=as.factor
(RegularMarij)))+
  geom_bar(color = "black", position="dodge")+
  theme_minimal()+
  labs(
    title = "Physical Activity Grouped by Regular Marijuana Usage",
    fill = "Regular Marijuana Usage",
    x = "Physically Active",
```

```
    y = "Percentage of People (n=4941)")+
  scale_fill_discrete_qualitative(palette = "Pastel 1")
```



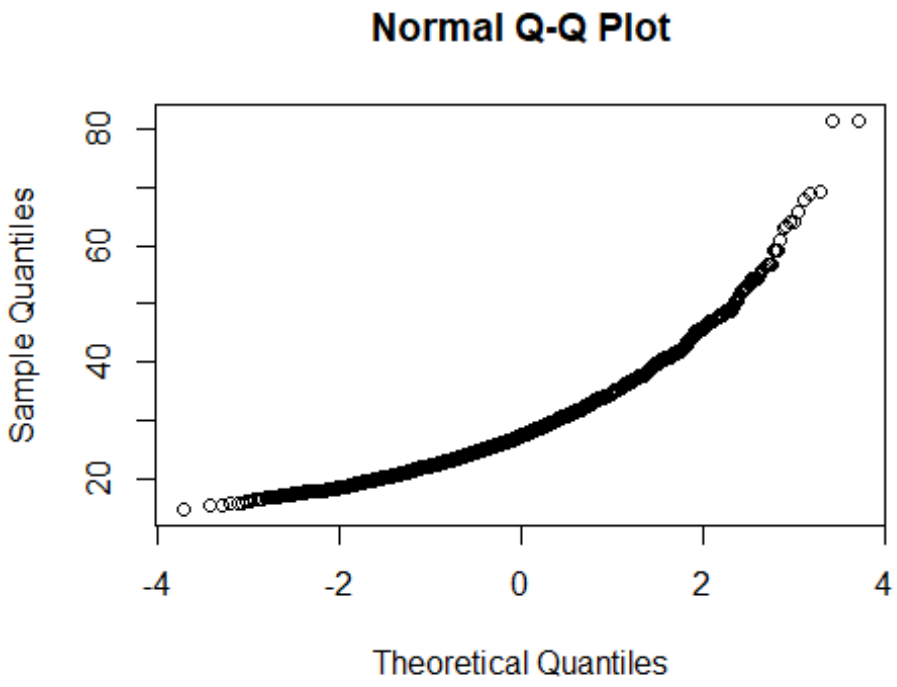
Looking at the sample and dividing them those who are physically active and those that are not, we see that about 58% of people are physically active and 42% of people are not physically active. From the sample, about 28% of people smoke marijuana regularly whereas 72% do not smoke marijuana regularly. This is represented by the colors of the bar graph, which means roughly a quarter of the population smokes marijuana regularly. 47% of people who smoke regularly are not physically active where as 53% of people who do smoke regularly are physically active. This 6% difference is visualized in the pink bar in the "yes" group is slightly higher than the pink bar in the "no" group.

This suggests that people who smoke marijuana are more physically active than their non-regular smoking counterparts, which is the opposite of what I anticipated. Due to marijuana's influence on developing habits and how marijuana makes the most mundane tasks entertaining to the smoker, I thought that regular marijuana smokers would be less inclined to be physically active.

The difference in size between the red and blue bars in the physically active group is greater in the difference in size than the in the non-physically active group. This supports that there is an association between the two variables because the bars would have a relatively similar height difference in each group if there was no association between the variables.

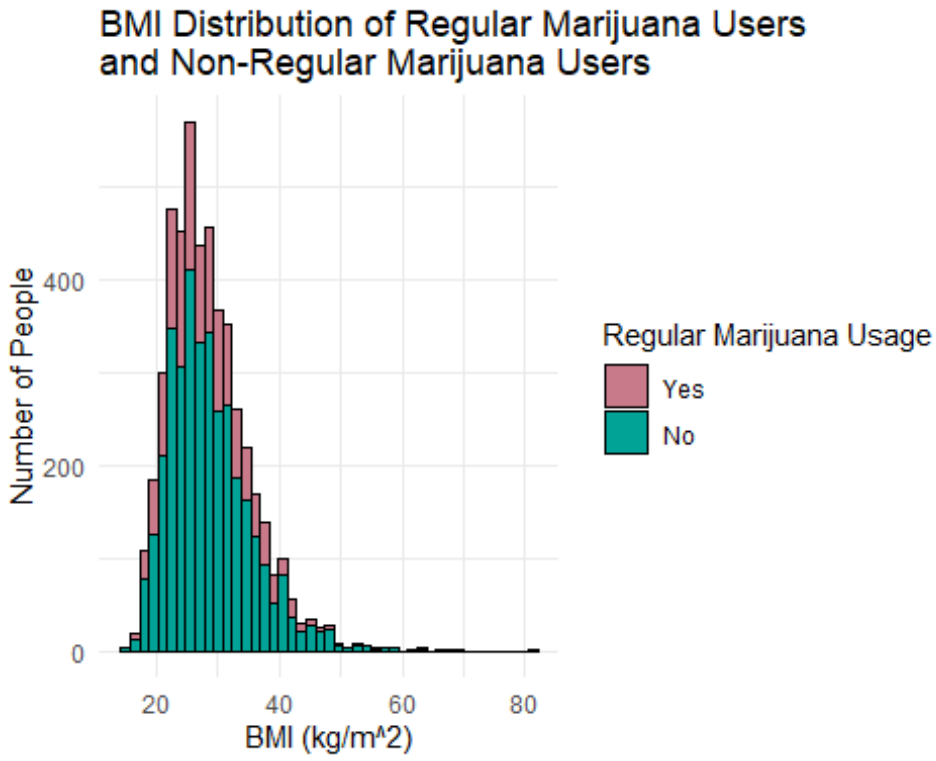
BMI

```
#Normality Test for BMI
qqnorm(NHANESRMJ$BMI)
```



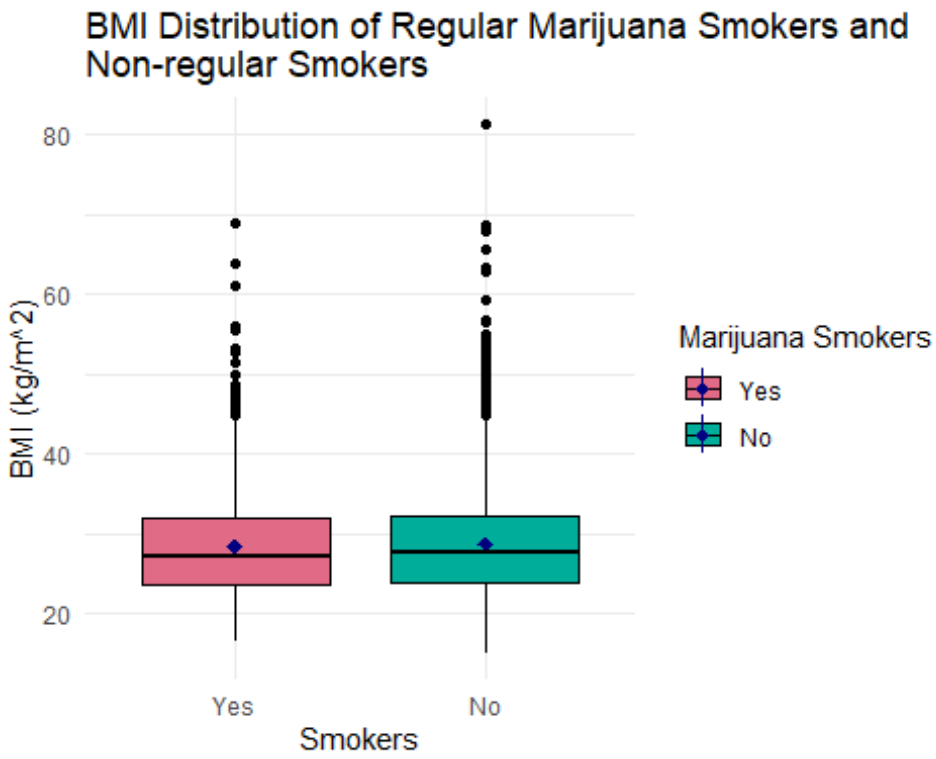

```
#BMI histogram split by regular smokers and non-regular smokers
NHANESRMJ %>%
  ggplot(aes(x=BMI, fill= RegularMarij))+
  geom_histogram(bins=45,color="black")+
  theme_minimal()+
  scale_fill_discrete_qualitative(palette = "Dark 2")+
  labs(
    title= "BMI Distribution of Regular Marijuana Users\nand Non-Regular Mari
juana Users",
    x= "BMI (kg/m^2)",
    y= "Number of People",
    fill= "Regular Marijuana Usage")

Warning: Removed 26 rows containing non-finite values (`stat_bin()`).
```




```
#Boxplot of BMI grouped by Smokers and non-smokers

NHANESRMJ %>%
  ggplot(aes(x=RegularMarij, y=BMI, fill= as.factor(RegularMarij)))+
  geom_boxplot(color="black")+
  theme_minimal()+
  labs(
    title= "BMI Distribution of Regular Marijuana Smokers and\nNon-regular Sm
okers",
    x= "Smokers",
    y= "BMI (kg/m^2)",
    fill = "Marijuana Smokers")+
  theme_minimal()+
  scale_fill_discrete_qualitative(palette = "Dark 3")+
  stat_summary(fun.y="mean",color="navy", shape=18, size=.5)
```




```
#Notable BMI Values

#26 People dropped for missing BMI Values
summary(NHANESRMJ$BMI, na.rm=TRUE)

#Summary Stats for BMI Values
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  15.02   23.81   27.46   28.66   32.20   81.25      26
```

```
BMI_Stats <- NHANESRMJ %>% drop_na(BMI)
BMI_Stats <- BMI_Stats %>%
  group_by(RegularMarij) %>%
  summarize(
    Mean_BMI = mean(BMI),
    Median_BMI = median(BMI),
    Min_BMI = min(BMI),
    Max_BMI = max(BMI),
    IQR_BMI = IQR(BMI),
    SD_BMI = sd(BMI),
    People = n())
BMI_Stats
```

```
# Summary Stats for BMI
  RegularMarij Mean_BMI Median_BMI Min_BMI Max_BMI IQR_BMI SD_BMI People
  <fct>           <dbl>      <dbl>   <dbl>   <dbl>   <dbl>  <dbl>  <int>
1 Yes              28.3       27.1    16.4    69      8.4    6.66   1355
2 No               28.8       27.6    15.0    81.2    8.27   6.89   3560
```

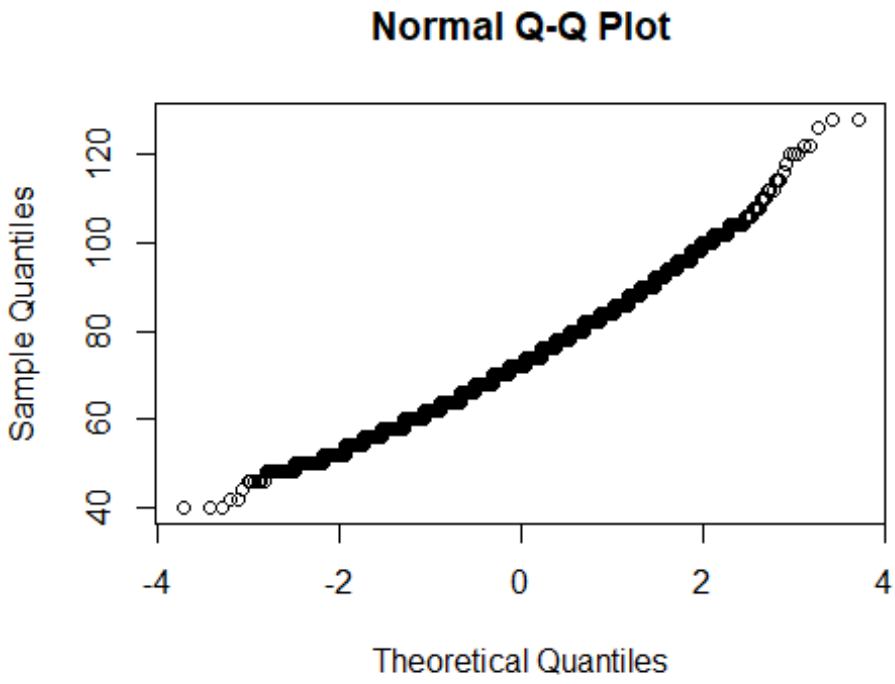
26 people out of the 4941 were dropped due to missing BMI data bring the sample size down to 4915. The BMI histogram shows the distribution of BMI split by regular marijuana smokers and non-regular marijuana smokers.

The BMI QQ plot and histogram indicates that the BMI values are not normally distributed. There is a slight right skew, indicating that the best test for this data is a Wilcox-Ranking test. In general, most of the BMI values are between 20 kg/m^2 and 40 kg/m^2. The box-plot compares the BMI distribution between regular marijuana smokers and non-regular smokers. The non-regular smokers have a larger BMI range, more upper outliers and a slightly higher mean and median than their regular smoking counterparts. The standard deviation in the non-regular smoking group is slightly greater than the regular smoking group. However, the IQR in the non-regular smoking group is slightly smaller than the regular smoking group. The maximum BMI value of the non-regular smoking group (81.2 kg/m^2) is much larger than the largest BMI for the regular smoking group (69.0 kg/m^2). The table above summarizes important statistics. Most values are very close to each other in each group except for the number of people in each group and the maximum BMI values.

From a macro perspective, this suggests that there is no difference in BMI in regular marijuana smokers and non-regular marijuana smokers. I had expected people who smoke regularly to have a higher BMI partially due to how marijuana induces the user to become hungry and would eat more than their non-regular smoking counterparts. Additionally I also anticipated that regular marijuana smokers would be less inclined to be physically active than non-regular marijuana smokers, resulting in a higher BMI.

Pulse

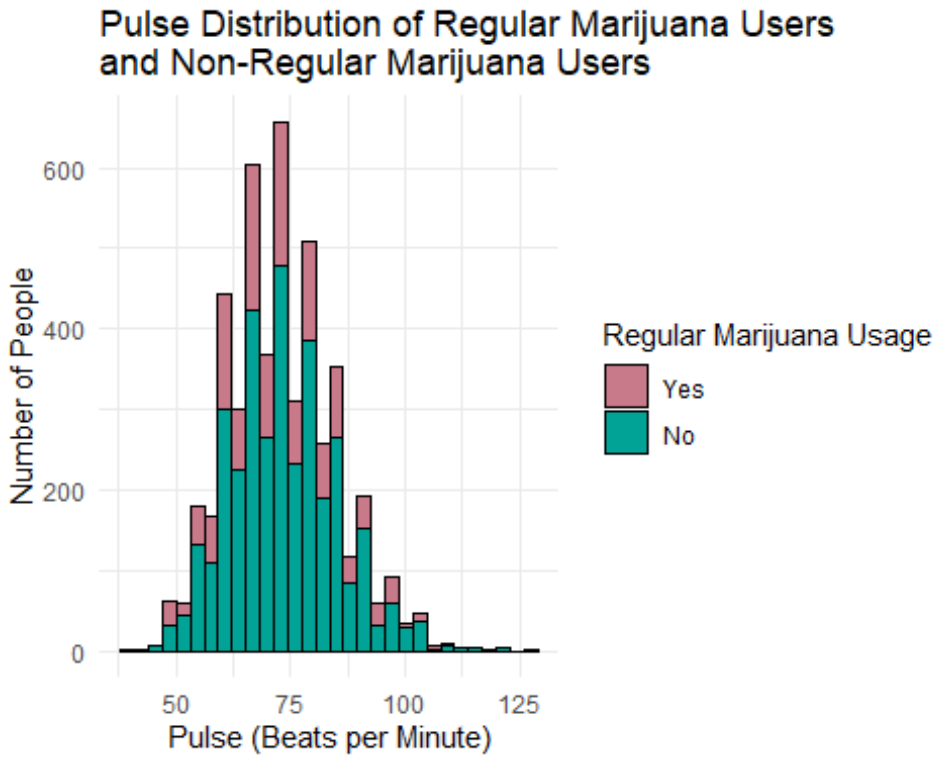
```
#Nomrality Test for Pulse
qqnorm(NHANESRMJ$Pulse)
```




```
#Histogram for Pulse (n= 4863, NA=78) grouped by Smoking Status

NHANESRMJ %>%
  ggplot(aes(x=Pulse, fill= RegularMarij))+
  geom_histogram(bins=30,color="black")+
  theme_minimal()+
  scale_fill_discrete_qualitative(palette = "Dark 2")+
  labs(
    title= "Pulse Distribution of Regular Marijuana Users\nand Non-Regular Ma
rijuana Users",
    x= "Pulse (Beats per Minute)",
    y= "Number of People",
    fill= "Regular Marijuana Usage")
```

```
Warning: Removed 78 rows containing non-finite values (`stat_bin()`).
```



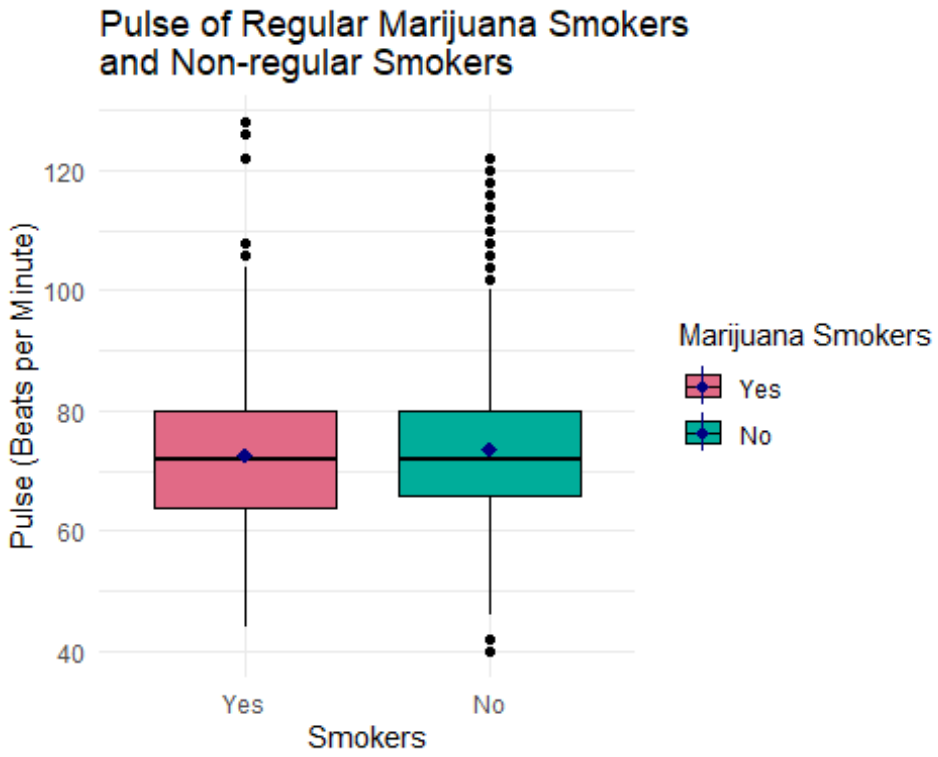

```
#Boxplot of Pulse Distribution grouped by Smoking Status

NHANESRMJ %>%
  ggplot(aes(x=RegularMarij, y=Pulse, fill= as.factor(RegularMarij)))+
  geom_boxplot(color="black")+
  theme_minimal()+
  labs(
    title= "Pulse of Regular Marijuana Smokers\nand Non-regular Smokers",
    x= "Smokers",
    y= "Pulse (Beats per Minute)",
    fill = "Marijuana Smokers")+
  theme_minimal()+
  scale_fill_discrete_qualitative(palette = "Dark 3")+
  stat_summary(fun.y="mean",color="navy", shape=18, size=.5)
```

```
Warning: Removed 78 rows containing non-finite values (`stat_boxplot()`).

Warning: Removed 78 rows containing non-finite values (`stat_summary()`).

Warning: Removed 2 rows containing missing values (`geom_segment()`).
```




```
#Pulse Summary Stats

summary(NHANESRMJ$Pulse)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  40.00   64.00   72.00   73.27   80.00  128.00      78

Pulse_Stats <- NHANESRMJ %>% drop_na(Pulse)
Pulse_Stats <- Pulse_Stats %>%
  group_by(RegularMarij) %>%
  summarize(
    Mean_Pulse = mean(Pulse),
    Median_Pulse = median(Pulse),
    Min_Pulse = min(Pulse),
    Max_Pulse = max(Pulse),
    IQR_Pulse = IQR(Pulse),
    SD_Pulse = sd(Pulse),
    People = n())
Pulse_Stats

# Summary Stats for Pulse
  RegularMarij Mean_Pulse Median_Pulse Min_Pulse Max_Pulse IQR_Pulse SD_Pulse
  <fct>             <dbl>        <dbl>     <int>     <int>     <dbl>    <dbl>
1 Yes                72.6           72        44       128        16     12.0
2 No                 73.5           72        40       122        14     11.6
# ℹ 1 more variable: People <int>
```

Of the original 4941 people, 78 were removed due to not having measured a pulse value, creating a sample size of 4863 people.

The pulse histogram and QQ Plot demonstrate a normal distribution, indicating that a two-sample t-test would be appropriate to compare the pulse means of the Regular smoking group and the non-regular smoking group. The upper and lower quartile of the pulse histogram is 80 to 64 beats per minute (BPM) with a max of 128 BPM and a minimum of 40 BPM. The mean and median pulse values are very close together (73.27 and 72.00 BPM respectively) which is typical of a normal distribution.

The pulse boxplot distributes the pulses by regular marijuana smokers and non-regular marijuana smokers. The median values of both groups are exactly the same and the mean values are very close to each other. The non-regular marijuana smoking group has both upper and lower outliers whereas the regular marijuana smoking group only has upper outliers, and they are more extreme than the outliers of the non-regular marijuana smoking group. Additionally, the IQR and standard deviation of the non-regular smoking group is slightly smaller than the regular smoking group. While the range of measured pulses are very similar (84 and 82 BPM for the regular smoking group and the non-regular smoking group respectively), the minimum and maximum pulse of the regular smoking group are both higher than the minimum and maximum pulse of the non-regular smoking group.

Similar to the BMI data, there does not seem to be much difference between the heart rates of regular smokers and non-regular smokers. This supports the null hypothesis; however, the people who regularly smoke have a slightly higher heart rate which is what I initially suspected.

Analysis

Physical Activity Analysis

Here we look at the association between people who regularly smoke marijuana and people who are or are not physically active. The hypothesis are as follows:

- The association of Regular Marijuana Usage and Physical Activity
    - H_0 : There is no association between Regular Marijuana Usage and Physical Activity
    - H_A : There is an association between Regular Marijuana Usage and Physical Activity

Since we trying to determine if there is an association between two categorical variables, it is best to use a chi-squared test so long as the expected values are all greater than 5.

```
#Contingency Table for Physically Active Regular Marijuana Users
PhysAct_table <-table(NHANESRMJ$RegularMarij,NHANESRMJ$PhysActive,dnn =
c("Regular Marijuana Use", "Physically Active"))
PhysAct_table_total <- addmargins(PhysAct_table)
PhysAct_table_total
```

```
                     Physically Active
Regular Marijuana Use  Yes   No  Sum
                  Yes  723  643 1366
                  No  2128 1447 3575
                  Sum 2851 2090 4941
```

```
#Chi-Squared
test <- chisq.test(PhysAct_table)
test
```

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  PhysAct_table
X-squared = 17.35, df = 1, p-value = 3.109e-05
```

```
round(test$expected) #all greater than 5
```

```
                     Physically Active
Regular Marijuana Use  Yes   No
                  Yes  788  578
                  No  2063 1512
```

```
test$stdres
```

```
                     Physically Active
Regular Marijuana Use       Yes        No
                  Yes -4.197532  4.197532
                  No   4.197532 -4.197532
```

```
#Code for table formations
expected <- cbind(c(788,2063), c(578,1512))

rownames(expected) <- c("Regular Marijuana Use (Yes)","Regular Marijuana Use
(No)")
colnames(expected) <-c("Physically Active (Yes)", "Physically Active (No)")
observed <- cbind(c(723,2128), c(643,1447))
rownames(observed) <- c("Regular Marijuana Use (Yes)","Regular Marijuana Use
(No)")
colnames(observed) <-c("Physically Active (Yes)", "Physically Active (No)")


library(knitr)
knitr::kable(expected, caption = "Expected Observations")
```

Expected Observations

	Physically Active (Yes)	Physically Active (No)
Regular Marijuana Use (Yes)	788	578
Regular Marijuana Use (No)	2063	1512

```
knitr::kable(expected, caption = "Sample Observations")
```

Sample Observations

	Physically Active (Yes)	Physically Active (No)
Regular Marijuana Use (Yes)	723	643
Regular Marijuana Use (No)	2128	1447

Based on the chi-squared test, there is significant association ($p= 0 < 0.05 = \alpha$) between the regularly smoking marijuana and physical activity. This lines up with the visual in that the ratios between the regular marijuana usage bars are not the same for the physical activity groups. The gap in bars is larger in the physically active group than the gap in bars from the non-physically active group.

The chi-squared results indicate that there were more people than expected from those non-active regular smokers and active non-smokers. We expected to see 788 physically active regular marijuana smokers but in reality we saw 723 people. Additionally, we expected to see 2063 physically active non-regular smokers but we observed 2128 physically active non-regular smokers. This is further complimented by the fact we expected to see 578 not physically active regular marijuana smokers, yet the actual number of people observed is higher at 643 people.

BMI Analysis

Next we will look at the relationship between the BMI of people who do and do not regularly smoke marijuana. Since the BMI distribution has a right-skew and many outliers, we will perform a two-sample t-test along with the Wilcoxon Rank Sum Test. The hypothesis for BMI is as follows:

- Regular Marijuana BMI mean = Non-Regular Marijuana BMI mean
    - $H_0 : \mu_{RegularBMI} - \mu_{NonRegularBMI} = 0$
        - Null Hypothesis: There is no difference in the mean BMI of regular marijuana users and non-regular marijuana users.
    - $H_A : \mu_{RegularBMI} - \mu_{NonRegularBMI} \neq 0$

- Alternative Hypothesis: There is a difference in the mean BMI of regular marijuana users and non-regular marijuana users.

```
#Two Sample T-Test for Mean
t.test(BMI~RegularMarij, data = NHANESRMJ)


    Welch Two Sample t-test

data:  BMI by RegularMarij
t = -2.037, df = 2523, p-value = 0.04176
alternative hypothesis: true difference in means between group Yes and group
No is not equal to 0
95 percent confidence interval:
 -0.85825721 -0.01633102
sample estimates:
mean in group Yes  mean in group No
         28.34294          28.78024

#Wilcoxon Test for Median
wilcox.test(BMI~RegularMarij, conf.int=TRUE, data = NHANESRMJ)


    Wilcoxon rank sum test with continuity correction

data:  BMI by RegularMarij
W = 2324583, p-value = 0.04951
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -7.599529e-01 -7.556621e-05
sample estimates:
difference in location
            -0.3899595

#Table for Results
```

The t-test provides estimates of the true mean of BMI values of regular smokers and non-regular smokers. Meanwhile, the Wilcoxon test provides estimations of the true median BMI values of the of regular and non-regular smokers.

According to the t-test, there is a significant difference in the true BMI means of regular smokers and non-regular smokers ($p=0.04 < 0.05 = \alpha$). Based on this we, can reject the null hypothesis and conclude that there is a significant difference in the true mean BMI between those that smoke marijuana regularly and those that do not. We can be 95% confident that the true difference in BMI between those who smoke marijuana regularly and do not smoke marijuana regularly is negative (-0.858 BMI to -0.016 BMI), indicating that those who do not regularly smoke marijuana have a higher BMI than those who do smoke marijuana.

Furthermore, the Wilcoxon test also states that we reject the null hypothesis ($p = 0.049 < 0.05 = \alpha$). Since this is based off the median values, we can conclude that there is

not a significant difference in the median BMI values of those who regularly smoke marijuana and those who do not. Since the BMI distribution contains many outliers and is right-skewed, the Wilcoxon test is worth testing as it not as sensitive to outliers. From this test, we can either fail to reject the null hypothesis or reject the null hypothesis depending on the number of significant figures one wishes to go out to. I chose to go with 3 significant figures as a result of comparing it fairly with the critical value, and to remain consistent with the rest of the p-values in this paper.

Finally, the 95% confidence interval states that the minimum difference between the means is -7.557×10^{-7}. This value, while extremely close to 0, is not actually 0. Therefore we can be 95% confident there is a difference in the true BMI medians between those who regularly smoke marijuana and those who do not regularly smoke marijuana is between (-0.760 BMI to 0 BMI). While zero is included in this range, it is the extreme end of the confidence interval. As a result, the Wilcoxon suggests that we should technically reject the null hypothesis; however, for practical purposes, we should not reject the null hypothesis. As a result, it is-at best-inconclusive.

Between the two tests, we should reject the null hypothesis and conclude that there is a significant difference in the true mean and true median of those who regularly smoke marijuana and those who do not. In both cases, the confidence intervals are negative which suggests that who do not smoke regularly have a higher BMI (at least in terms of the mean) than the BMI of those who smoke regularly. In terms of a difference in the medians, the results are inconclusive.

Pulse Analysis

Finally, we will look at the heart rate (Pulse Variable) and determine if there is a difference between the heart rate of those who regularly smoke marijuana and those who do not. Since the *Pulse* variable follows a normal distribution, the two-sample t-test would be the best for this statistical analysis. The hypothesis is as follows:

- Regular Marijuana Pulse mean = Non-Regular Marijuana Pulse mean
  - $H_0 : \mu_{RegularPulse} - \mu_{NonRegularPulse} = 0$
    - Null Hypothesis: There is no difference in the mean pulse (Heart Rate) of regular marijuana users and non-regular marijuana users.
  - $H_A : \mu_{RegularPulse} - \mu_{NonRegularPulse} \neq 0$
    - Alternative Hypothesis: There is a difference in the mean BMI of regular marijuana users and non-regular marijuana users.

```
#Two Sample T-Test for Mean
t.test(Pulse~RegularMarij, data = NHANESRMJ)


    Welch Two Sample t-test

data:  Pulse by RegularMarij
t = -2.4843, df = 2353.5, p-value = 0.01305
```

```
alternative hypothesis: true difference in means between group Yes and group
No is not equal to 0
95 percent confidence interval:
 -1.6894225 -0.1988796
sample estimates:
mean in group Yes  mean in group No
         72.58824          73.53239
```

Since p = 0.01 < 0.05 = α, we reject the null hypothesis and conclude that there is a significant difference between the true average *Pulse.* Additionally, we can be 95% confident that the true difference in the average heart rate is between -1.689 to -0.199 beats per minute (BPM), meaning that those who do not smoke marijuana have a higher heart rate than those who do smoke marijuana.

Conclusions

```
conclusion <- cbind(c(3.109e-5,0.0417,0.0495,0.0130),
                    c("NA",-0.858,-0.760,-1.689),
                    c("NA", -0.016,0,-0.199), c("reject","reject","reject","r
eject"))
rownames(conclusion)=c("Physically Active (Chi-Squared)","BMI(mean)","BMI(med
ian)","Pulse(mean)")
colnames(conclusion) <- c("p-values","95% Confidence Lower Bound","95% Confid
ence Upper Bound","Null Hypothesis")

conclusion
```

```
                              p-values    95% Confidence Lower Bound
Physically Active (Chi-Squared) "3.109e-05" "NA"
BMI(mean)                     "0.0417"    "-0.858"
BMI(median)                   "0.0495"    "-0.76"
Pulse(mean)                   "0.013"     "-1.689"
                              95% Confidence Upper Bound Null Hypothesis
Physically Active (Chi-Squared) "NA"                     "reject"
BMI(mean)                     "-0.016"                   "reject"
BMI(median)                   "0"                        "reject"
Pulse(mean)                   "-0.199"                   "reject"
```

```
knitr::kable(conclusion, caption = "Stastical Analysis Summary")
```

Statical Analysis Summary (See Data Description for Variable Units)

	p-values	95% Confidence Lower Bound	95% Confidence Upper Bound	Null Hypothesis
Physically Active (Chi-Squared)	3.109e-05	NA	NA	reject
BMI(mean)	0.0417	-0.858	-0.016	reject
BMI(median)	0.0495	-0.76	0	reject
Pulse(mean)	0.013	-1.689	-0.199	reject

Regular smokers are less active, have a lower heart and a lower BMI on average then their non-regular smoking counterparts. The statistical summary table above summarizes of the p-values and the 95% Confidence Intervals for each test performed.

The most notable finding is that there is an association between regularly smoking marijuana and being physically active. The chi-squared test demonstrated that there is a *lower* number of regular smokers who are physically active, and a *higher* number of regular smokers who are not physically active than expected. Additionally the number of non-regular smokers who physically work out is higher than expected. This is further reinforced by the bar graph. The gaps in the bars when groups by physically activity are not identical, which indicates an underlying association between regular marijuana use and physical activity. Because of this, *I would tell someone who regularly smokes marijuana that is not good for their general health as they are more prone to living a sedentary lifestyle.*

The next step in this research would be to determine if people started smoking marijuana *before* or *after* becoming physically active. Marijuana aids in habit formation, and thus it likely that the time at which people became regular smokers influences this data. Is a person who is physically active likely to stay physically active upon regular marijuana usage? Is a person who already is not physically active likely to become physically active if they start using marijuana? Does marijuana reinforce existing physical activity behavior (or lack thereof)? These questions are not answered by this study. However, this study does conclude that there is scientific and statistical difference in the physical activity of those who do and do not regularly smoke marijuana.

Although there is a statistically significant difference between true means of the BMI and heart rate of regular smokers and non-regular smokers, that difference is tiny. Per the 95% Confidence Interval for the BMI mean, we can be 95% confident that non-regular smokers have a higher BMI than regular smokers by about -0.858 to -0.016 kg/m^2. A less than 1 BMI difference is not scientifically meaningful. For practical purposes, there is not a difference in BMIs between the two group even through there is a statistical significance.

Furthermore, this is the same case when looking at the heart rate differences between regular marijuana smokers and the non-regular smokers. The 95% confidence interval for the differences in heart rate is -1.689 to -0.198 BPM. A difference in about 1 BPM is not a very meaningful from a biological perspective, and could be due to a variety of factors. However, there is a statistical significance between the two groups that rejects the null hypothesis. Typically, those who are more sedentary have a higher heart rate; however, a difference of 1 BPM is not enough to biologically conclude that regular marijuana use is healthy or unhealthy, as a healthy heart rate fluctuates within a population and even within an individual. Regardless, non-regular marijuana smokers have a higher mean BPM than regular marijuana smokers in this study implying that regular smokers are healthier than their non-regular counterparts.

Lastly, the BMI had a right-skew distribution which led me to use the Wilcoxon-Rank test to compare the difference in the mediums between the BMI of regular smokers and non-regular smokers. The p-value of this test is 0.049 and the confidence interval on the higher end is -7.55×10^{-5} which I rounded to zero in the table above. I concluded that

there was a statistical significant difference between the two groups. However, I believe that conclusion should best be left to the reader. If one chooses to round 0.049 to 0.05, then one could fail to reject the null hypothesis, and this would be consistent with the 95% Confidence Interval as the upper limit (-7.55×10^{-5}) can reasonably be rounded to zero. In this case, a difference of zero would fall within the confidence interval which would be consistent with the p-value results. I chose to reject the null hypothesis due in an attempt to keep the results as accurate as possible. Nonetheless, it is important to point out where subjective decisions have been made in the process of the statistical analysis. Regardless, the difference between the true median BMI values, even if we take the lower bound of the confidence interval (-0.760), is not biologically significant. Comparing the median values is more accurate and reflective of the BMI distribution between the two groups due to the sheer number of outliers-highlighted by the boxplot. This also supports the conclusion we got from comparing the difference in the mean BMI values. Neither one are biologically significant.

From the study, we can conclude that regular marijuana smokers are live a more sedentary life than their non-regular smoking peers. There is a significant association between being physically active and smoking marijuana in which those who regularly smoke marijuana are not as physically active as those who do not regularly smoke marijuana. Paradoxically, there does not seem to be much of a biological difference between regular smokers and non-regular smokers based on BMI and heart rate estimates. Ironically, non-regular smokers have a slightly higher BMI and heart rate on average. There are many factors that may account for this, including the use of marijuana itself. Marijuana has been shown to be a protective factor against obesity. Additionally, marijuana is also linked to a *higher* heart rates, but the study did not specify if the heart rate leads to a higher rate while experiencing a marijuana high or if an elevated heart rate is long term effect of regularly smoking marijuana. *While regular marijuana smokers may be more sedentary than their non-regular smokers, the difference in their health seem to be minimal based on the estimates of this study*. From here, the next parameters to measure would be the frequency of physical activity, as well as measuring other health and social variables associated with a sedentary lifestyle such as stroke frequency, blood pressure, and the amount of time spent indoors vs outdoors.

Works Cited

Brook, Judith S et al. “Physical factors, personal characteristics, and substance use: associations with obesity.” *Substance abuse* vol. 34,3 (2013): 273-6. doi:10.1080/08897077.2013.770425

Grinspoon,Peter. “The endocannabinoid system: Essential and Mysterious”*Harvard Health Publishing*. 11 August 2021. https://www.health.harvard.edu/blog/the-endocannabinoid-system-essential-and-mysterious-202108112569

Hasan, Khaled M. “Cannabis Unveiled: An Exploration of Marijuana's History, Active Compounds, Effects, Benefits, and Risks on Human Health.” *Substance abuse: research and treatment* vol. 17 11782218231182553. 20 Jun. 2023, doi:10.1177/11782218231182553

NIDA. "Cannabis (Marijuana) DrugFacts." *National Institute on Drug Abuse*, 24 Dec. 2019, https://nida.nih.gov/publications/drugfacts/cannabis-marijuana Accessed 9 Sep. 2023.

NIDA. "How does marijuana use affect school, work, and social life?." *National Institute on Drug Abuse*, 17 Apr. 2023, https://nida.nih.gov/publications/research-reports/marijuana/how-does-marijuana-use-affect-school-work-social-life Accessed 9 Sep. 2023.

Neff, Sarah,et al.“Inhaled Marijuana and the Lungs.” American Thoracic Society, 2021. https://www.thoracic.org/patients/patient-resources/resources/marijuana.pdf. Accessed 09 September 2023.

Sidney, S. (2002), “Cardiovascular Consequences of Marijuana Use”. *The Journal of Clinical Pharmacology*, 42: 64S-70S. https://doi.org/10.1002/j.1552-4604.2002.tb06005.x