Project Report: Depression and Obesity

Josh Christensen and Anna Leman

4/14/2021

Introduction

Obesity and depression are two diseases that are becoming more and more frequent in our world at large. These diseases generally have a negative effect on the quality of life of those experiencing them and have been researched and studied extensively. They are both treatable but many still struggle with their harmful effects.

The question we are considering is if these two problems are linked and if depression has a causal relationship with obesity. According to Bryne, O'Brien-Simpson, Mitchell, and Allen, depression has an association with obesity but the exact relationship between them is not totally known (2015). They found that depression amongst adolescents could potentially be a risk for obesity (Bryne et al., 2015). Another study done investigating depression and obesity among Mexican people, found that there was an association between the two for women, but not for men (Zavala et al., 2017). We hope to look at this relationship as well, to find out if depression has a causal effect on obesity in American adults.

Data

Our data was collected by the CDC's National Center for Health Statistics from their National Health and Nutrition Examination Survey from the years 2017-2018. This data can be found at the following link: National Health and Nutrition Examination Survey.

We searched for variables that we were interested in and found the data file where the variable resides. This data has information covering the demographics of the participants, the questionnaire data and the examination data. We chose to only look at participants with all their data present. Each data set was downloaded into SAS as a SAS export transport file which was then converted from a SAS library to a CSV in SAS. From here we imported the CSV files into R, where we pulled out the variables we were interested in and merged them into a data set by respondent sequence number (respondent ID).

Each row is an American individual who has taken this NHANES survey. Our problem requires information on depression, BMI, education level, diet, income, exercise and race. These variables are required because we want to assess the impact of depression on obesity (elevated BMI), and we want to control for diet, income, and exercise to see if depression truly has an effect on obesity. Below is the causal graph where we were able to identify D (income), A (diet) and B (exercise) as causes of C (depression) and G (obesity). We then applied the disjunctive cause criterion by controlling for these variables in our causal analysis. In all the causal graph includes the following variables: A - Diet , B - Exercise, C - Depression, D- Income, E - Education, F - Race, G - Obesity.

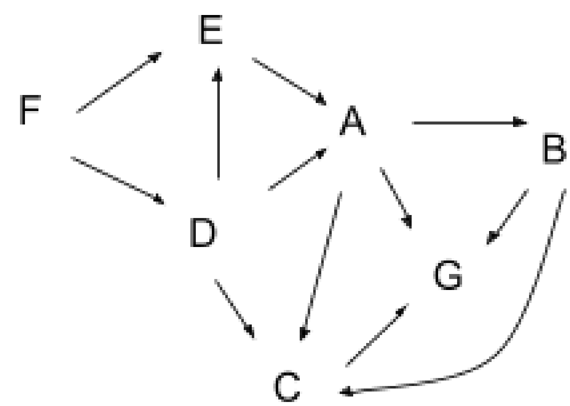

Figure 1: Directed Acyclical Causal Graph

Variables

Our unit of measure for our response variable is Body Mass Index (BMI) and anyone who classifies as obese on this scale (a score of 30 or higher) was counted as an obese individual. We have several explanatory variables. The first is diet. Each participant was asked "In general, how healthy is {your/his/her} overall diet?" The respondent answered one of the following: 1 = Excellent, 2 = Very good, 3 = Good, 4 = Fair, 5 = Poor, 7 = Refused, and 9 = Don't know.

The next explanatory variable is depression, which is what we are considering our treatment. For depression we have collected 9 questions on the participants current mental state. If the participant answered 1 or more times to the following question we will assume they are depressed: "Over the last 2 weeks, how often have you been bothered by the following problems: Thoughts that you would be better off dead or of hurting yourself in some way?" If participants answered that any of the following had bothered them 3 or more times over the last two weeks we also considered them depressed. The symptoms were as follows: little interest or pleasure in doing things, feeling down, depressed, or hopeless, trouble falling or staying asleep, or sleeping too much, feeling tired or having little energy, poor appetite or overeating, feeling bad about yourself - or that you are a failure or have let yourself or your family down, trouble concentrating on things, such as reading the newspaper or watching TV, or moving or speaking so slowly that other people could have noticed or the opposite - being so fidgety or restless that you have been moving around a lot more than usual.

We also used education level. The education level of the participants was gathered via this question: What is the highest grade or level of school {you have/SP has} completed or the highest degree {you have/s/he has} received? The answers were as follows: 1 = less than 9th grade, 2 = 9th - 12th grade without diploma, 3 = High School graduate or GED recipient , 4 = Some college or AA degree, 5 = college graduate or above, 7 = Refused , or 9 = Don't Know.

Income was reported as a categorical variable with varying ranges for each category.We removed those who had responded as above $20000, below $20000, didn't know, and refused to respond. After the removal we took the average of each range so that we had a quantitative variable instead, with the exception of the top bracket which was $100000 and above. This upper bracket was about 20% of our data, so we took the average income for the upper class of Americans (which is roughly 20% of the population) and used this value instead. This left us with the following incomes to represent each category: $2500, $7500, $12500, $17500, $22500, $30000, $40000, $50000, $60000, $70000, $87500, $188000.

Exercise was measured with the following question - {Are you/Is s/he} now doing any of the following: increasing {your/his/her} physical activity or exercise? If the response is a 1 then the answer is yes, if the response is a 2 then the answer is no, 7 is a refusal, 9 is a "don't know".

Race was tracked with differing codes. Respondents who self-identified as "Mexican American" were coded as 1 regardless of their other race-ethnicity identities. Otherwise, self-identified "Hispanic" ethnicity would result in code "2, Other Hispanic". All other non-Hispanic participants were then categorized based on their self-reported races: non-Hispanic white 3, non-Hispanic black 4, non-Hispanic Asian 6, and other non-Hispanic races including non-Hispanic multiracial 7.

In table one the zero column represents nondepressed individuals while the one column represents depressed individuals. Healthy diet was the largest standardized mean difference we observed. Upon closer examination we can see that the entire distribution for depressed people is flipped towards worse diets. The education level also had a large standardized mean difference, with depressed individuals having attained higher education levels on average than nondepressed individuals.

Table 1: Demographic and Health Characteristics

	Overall	0	1	SMD
n	4198	3338	860	
Obesity = 1 (%)	1765 (42.0)	1342 (40.2)	423 (49.2)	0.181
Race (%)				0.242
1	531 (12.6)	423 (12.7)	108 (12.6)	
2	361 (8.6)	279 (8.4)	82 (9.5)	
3	1570 (37.4)	1203 (36.0)	367 (42.7)	
4	951 (22.7)	780 (23.4)	171 (19.9)	
6	562 (13.4)	489 (14.6)	73 (8.5)	
7	223 (5.3)	164 (4.9)	59 (6.9)	
Income (mean (SD))	69099.33 (62899.85)	73635.86 (64154.52)	51491.28 (54326.29)	0.373
Education.Level (%)				0.386
1	285 (6.8)	201 (6.0)	84 (9.8)	
2	452 (10.8)	320 (9.6)	132 (15.3)	
3	1026 (24.4)	801 (24.0)	225 (26.2)	
4	1398 (33.3)	1101 (33.0)	297 (34.5)	
5	1034 (24.6)	915 (27.4)	119 (13.8)	
7	1 (0.0)	0 (0.0)	1 (0.1)	
9	2 (0.0)	0 (0.0)	2 (0.2)	
Increasing.Exercise (%)				0.092
1	2502 (59.6)	2019 (60.5)	483 (56.2)	
2	1694 (40.4)	1318 (39.5)	376 (43.7)	
9	2 (0.0)	1 (0.0)	1 (0.1)	
Healthy.Diet (%)				0.413
1	292 (7.0)	252 (7.5)	40 (4.7)	
2	853 (20.3)	726 (21.7)	127 (14.8)	
3	1638 (39.0)	1349 (40.4)	289 (33.6)	
4	1116 (26.6)	839 (25.1)	277 (32.2)	
5	299 (7.1)	172 (5.2)	127 (14.8)	

Methods

To answer our question on whether depression causes obesity we ran a causal matched pairs analysis. In order to fit this model we needed to make sure our assumptions were met. The assumptions are ignorability, SUTVA, positivity and consistency. We checked the ignorability assumption which states that given pre-treatment covariates X, treatment assignment is independent of the potential outcomes. This assumption is met by adjusting our sample through matching, based on the propensity scores, in such a way that obesity and depression are independent of each other. The Stable Unit Treatment Value Assumption (SUTVA) states that treatment assignment of one unit does not affect the outcome of another unit. We are assuming that each individual in our sample is independent of each other. We are also treating depression as general depressive behavior, not focusing on one type of depression, so this assumption is met. Positivity states that treatment assignment was not deterministic. The positivity assumption is met through subsetting our sample by propensity scores so that there are no outliers. There are treated and untreated individuals for every propensity score that we are including. Consistency states that the setting has not changed in between the treatment occurring and us observing the outcome so that the potential outcome that individual on that treatment matches our observed outcome. This assumption is reasonable in our case since our data comes from a cross-sectional study and both the treatment and the response are observed simultaneously.

To analyze this data we found the propensity scores for each individual, using the variables income, diet, and exercise to assess their impact on depression. We graphed the propensity scores of those with depression against those without depression. After looking at this graph we found that there were two individuals with depression that needed to be dropped based on their propensity scores being larger than that of the highest nondepressed propensity score. After these two individuals were dropped the propensity scores for the depressed and nondepressed aligned nicely.

After the propensity scores had been subsetted we matched each depressed individual with 3 nondepressed individuals based on how closely related their propensity scores were, with a caliper of 1.5. We used this matched data to create a new table 1 which compared the depressed individuals with the nondepressed individuals. Based on the matched data we were then able to find the odds ratio and complete a McNemar's test for our outcome analysis.

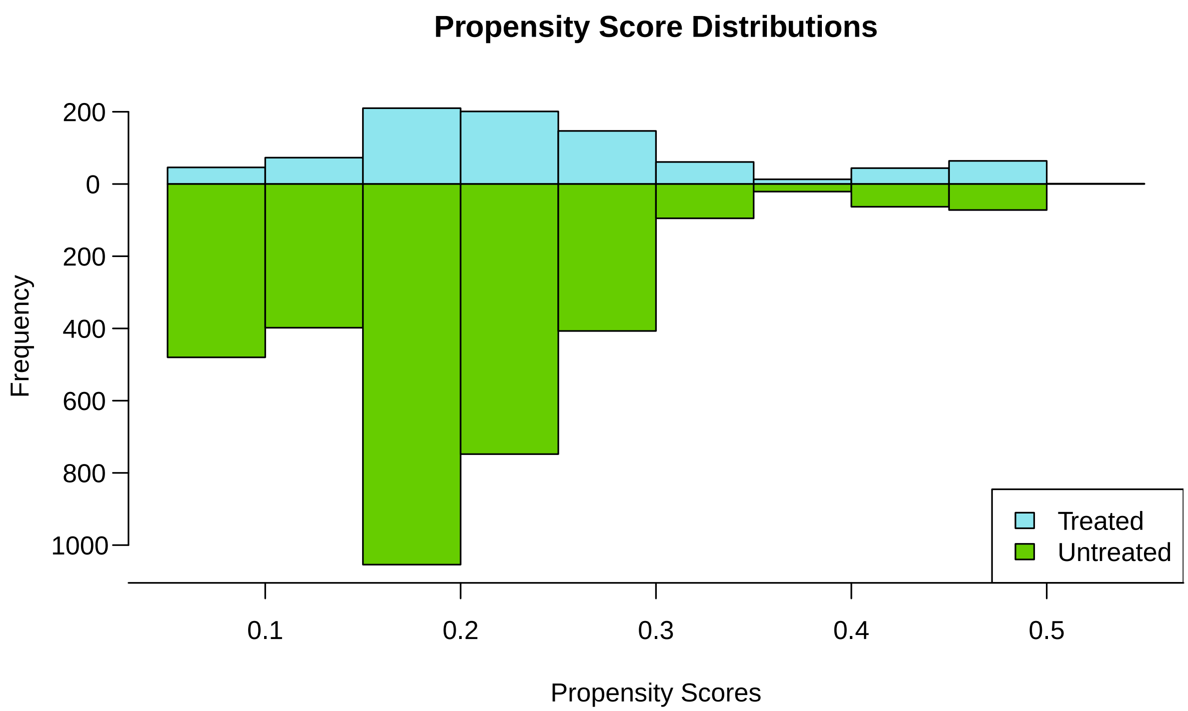

Figure 2: Propensity Score Distributions

Table 2: Matched Demographics and Health Characteristics

	0	1	SMD
n	2448	816	
Income (mean (SD))	56371.32 (53945.32)	53085.17 (55192.66)	0.060
Increasing.Exercise (%)			0.029
1	1414 (57.8)	470 (57.6)	
2	1033 (42.2)	346 (42.4)	
9	1 (0.0)	0 (0.0)	
Healthy.Diet (%)			0.142
1	104 (4.2)	40 (4.9)	
2	418 (17.1)	127 (15.6)	
3	956 (39.1)	287 (35.2)	
4	798 (32.6)	277 (33.9)	
5	172 (7.0)	85 (10.4)	

Results

Our odds ratio for our outcome analysis was 1.239 (95% confidence interval: 1.057, 1.452) which indicates that the odds of being obese are 1.239 times higher if you are depressed. McNemar's test indicated that the difference in proportions of depressed individuals that are obese and nondepressed individuals that are obese are significantly different with a p-value of approximately zero.

In order to validate our results we performed a sensitivity analysis in which we reproduced our analysis under slightly different assumptions. First we reproduced the analysis by using all of the available covariates to calculate propensity scores, rather than only those that satisfied the disjunctive cause criterion. This analysis produced an odds ratio of 1.25 and a McNemar's test p-value of approximately zero. Both of these results are consistent with our original analysis. Next we varied the caliper size and number matches. Finally we recreated the analysis using different estimated incomes for the top income bracket. The resulting odds ratios and McNemar's p-values for these various combinations are listed in the table below. We can see from the table that regardless of how we modify our assumptions and methods we reach the same conclusion that depression does relate to a statistically significant increase in likelihood of obesity.

Table 3: Sensitivity Analysis

Caliper	Matches	Income	Odds	Pvals
0.1	3	$188,000	1.257	0.000
0.5	3	$188,000	1.201	0.000
1.0	3	$188,000	1.209	0.000
1.5	3	$188,000	1.294	0.000
2.0	3	$188,000	1.260	0.000
1.5	1	$188,000	1.252	0.032
1.5	2	$188,000	1.221	0.000
1.5	3	$125,000	1.275	0.000
1.5	3	$150,000	1.264	0.000
1.5	3	$200,000	1.231	0.000

Conclusion

Our hypothesis for our research was that depression would have a causal relationship with obesity. After completing our analysis, our results support this hypothesis. Our odds ratio was above 1 and our McNemar's test was significant. These two things led us to believe that our hypothesis was correct. Obesity and depression are huge problems across the world and they are only getting bigger. Understanding how they play into one another could be helpful in treating these problems. Our analysis generalizes the relationship between obesity and depression to adults, and confirms the findings of Bryne, O'Brien-Simpson, Mitchell, and Allen (2015). It also supports the research done previously on Mexican populations as stated above and generalizes to Americans (Zavala et al., 2017).

In general depression can lead to a lack of interest in activity, and a lack of energy. Both of these things can lead to a decline in the quality of diet and physical exercise. These things are two of the largest factors that play into obesity. Because of this it makes sense that depression would have a causal relationship with obesity. This analysis, while interesting, could be taken further to more fully understand the impact of depression on obesity. In furthering the research we would recommend a cohort study to see if people develop depression and whether or not after that they became obese. We would also recommend studying more covariates. There could be more research done to better identify which covariates should be included in a causal analysis. These are a few recommendations for furthering the research on this topic.

Our data was from a cross-sectional study which is inherently restrictive, and we had to find creative ways to define obesity, depression and income. The information gathered from this survey was extensive and there were more variables we could have considered, but we chose what felt most applicable to obesity and depression. Through the use of this data we were able to identify a causal relationship between obesity and depression which could be helpful for future learning and research.

References

Byrne, M., O'Brien-Simpson, N., Mitchell, S., Allen, N., Byrne, M. L., O'Brien-Simpson, N. M., Mitchell, S. A., & Allen, N. B. (2015, Adolescent-Onset Depression: Are Obesity and Inflammation Developmental Mechanisms or Outcomes? Child Psychiatry & Human Development,46, 839-850. https://search.lib.byu.edu/byu/record/edsbyu.eft.110340160

Zavala, G. A., Kolovos, S., Chiarotto, A., Bosmans, J. E., Campos-Ponce, M., Rosado, J. L., & Garcia, O. P. (2018, Association between obesity and depressive symptoms in Mexican population. Social Psychiatry & Psychiatric Epidemiology, 53, 639-646. https://search.lib.byu.edu/byu/record/edsbyu.pbh.129685441

Appendix

```
knitr::opts_chunk$set(echo = FALSE, fig.dim = c(8,5), tidy.opts=list(width.cutoff=65), tidy=TRUE, fig.al
```

```
/* SAS CODE FOR CREATING CSVS */
libname in xport "C:\Users\joshu\OneDrive\Documents\Winter 2021\Biostatistics\Project\DPQ_J.XPT";
proc copy inlib=in outlib=work;
run;

libname in xport "C:\Users\joshu\OneDrive\Documents\Winter 2021\Biostatistics\Project\BMX_J.XPT";
proc copy inlib=in outlib=work;
run;

libname in xport "C:\Users\joshu\OneDrive\Documents\Winter 2021\Biostatistics\Project\DEMO_J.XPT";
proc copy inlib=in outlib=work;
run;

libname in xport "C:\Users\joshu\OneDrive\Documents\Winter 2021\Biostatistics\Project\DIQ_J.XPT";
proc copy inlib=in outlib=work;
run;

libname in xport "C:\Users\joshu\OneDrive\Documents\Winter 2021\Biostatistics\Project\DBQ_J.XPT";
proc copy inlib=in outlib=work;
run;

libname in xport "C:\Users\joshu\OneDrive\Documents\Winter 2021\Biostatistics\Project\MCQ_J.XPT";
proc copy inlib=in outlib=work;
run;

proc export data=Dpq_j outfile='C:/Users/joshu/OneDrive/Documents/Winter 2021/Biostatistics/Project/dpq_
run;

proc export data=Bmx_j outfile='C:/Users/joshu/OneDrive/Documents/Winter 2021/Biostatistics/Project/bmx_
run;
```

```
proc export data=Demo_j outfile='C:/Users/joshu/OneDrive/Documents/Winter 2021/Biostatistics/Project/dem
run;


proc export data=Diq_j outfile='C:/Users/joshu/OneDrive/Documents/Winter 2021/Biostatistics/Project/diq_
run;


proc export data=Dbq_j outfile='C:/Users/joshu/OneDrive/Documents/Winter 2021/Biostatistics/Project/dbq_
run;


proc export data=Mcq_j outfile='C:/Users/joshu/OneDrive/Documents/Winter 2021/Biostatistics/Project/mcq_
run;
# Packages
library(Matching)
library(tableone)
library(formattable)
library(knitr)
# Set working directory relative to project
setwd("Project")

# Initialize data frames
bmx <- read.csv("bmx_j.csv", header = TRUE)
dbq <- read.csv("dbq_j.csv", header = TRUE)
demo <- read.csv("demo_j.csv", header = TRUE)
diq <- read.csv("diq_j.csv", header = TRUE)
dpq <- read.csv("dpq_j.csv", header = TRUE)
mcq <- read.csv("mcq_j.csv", header = TRUE)

# Initialize merged data frame
data1 <- dpq[,1:10]
data2 <- merge(data1,demo[,c("SEQN","RIDRETH3","INDFMIN2","DMDEDUC2")], by = "SEQN")
data3 <- merge(data2,bmx[,c("SEQN","BMXBMI")], by = "SEQN")
data4 <- merge(data3,mcq[,c("SEQN","MCQ371B")], by = "SEQN")
```

```r
data5 <- merge(data4,dbq[,c("SEQN","DBQ700")], by = "SEQN")

# Omit NAs
data <- na.omit(data5)

# Rename columns
colnames(data)[11:16] <- c("Race","Income","Education.Level","BMI","Increasing.Exercise","Healthy.Diet")

# initialize data$depression
data$Depression <- NA

# Create data$obesity
data$Obesity <- as.numeric(data$BMI > 30)

# Create depression variable
for(i in 1:nrow(data)){
  if(data$DPQ090[i] > 0 | data$DPQ010[i] > 2 | data$DPQ020[i] > 2 | data$DPQ030[i] > 2 | data$DPQ040[i]
    data$Depression[i] <- 1
  }else{
    data$Depression[i] <- 0
  }
}

# Create logical depression for use in matching
data$Logical.Depression <- as.logical(data$Depression)

# Make variables factors
data$Race <- factor(data$Race)
data$Education.Level <- factor(data$Education.Level)
data$Increasing.Exercise <- factor(data$Increasing.Exercise)
data$Healthy.Diet <- factor(data$Healthy.Diet)
data$Depression <- factor(data$Depression)
```

```r
data$Obesity <- factor(data$Obesity)

# Remove unused columns
data <- data[-c(2:10,14)]

# Remove above 20,000 and below 20,000
data <- data[!(data$Income == 13 | data$Income == 12 | data$Income == 77 | data$Income == 99),]
rownames(data) <- 1:nrow(data)

# Fix the income column

# Create vector of the averages for the income ranges included in the dataset
newincome <- numeric(nrow(data))

# Create if else statements to replace categorical labels with averages
for(i in 1:nrow(data)){
  if(data$Income[i] == 1){
    newincome[i] <- 2500
  }
  if(data$Income[i] == 2){
    newincome[i] <- 7500
  }
  if(data$Income[i] == 3){
    newincome[i] <- 12500
  }
  if(data$Income[i] == 4){
    newincome[i] <- 17500
  }
  if(data$Income[i] == 5){
    newincome[i] <- 22500
  }
  if(data$Income[i] == 6){
```

```r
    newincome[i] <- 30000
  }
  if(data$Income[i] == 7){
    newincome[i] <- 40000
  }
  if(data$Income[i] == 8){
    newincome[i] <- 50000
  }
  if(data$Income[i] == 9){
    newincome[i] <- 60000
  }
  if(data$Income[i] == 10){
    newincome[i] <- 70000
  }
  if(data$Income[i] == 14){
    newincome[i] <- 87500
  }
  if(data$Income[i] == 15){
    newincome[i] <- 188000
  }
}

# Replace the income column with new income
data$Income <- newincome

# Create initial table one
tab1 <- CreateTableOne(vars = c("Obesity","Race","Income","Education.Level","Increasing.Exercise","Heal
# Export to kable
kableone(print(tab1,smd = TRUE), caption = "Demographic and Health Characteristics")
# Fit a propensity score model using logistic regression
psmodel <- glm(Depression ~ Income + Increasing.Exercise + Healthy.Diet,
               family = binomial,
```

```r
              data = data)

# Store the fitted values for each observation
pscore <- psmodel$fitted.values

# Check overlap to evaluate positivity assumption
h1 <- hist(pscore[data$Depression == 1], plot = FALSE, breaks = 10)
h2 <- hist(pscore[data$Depression == 0], plot = FALSE, breaks = 10)
h2$counts <- -h2$counts
hmax <- max(h1$counts)
hmin <- min(h2$counts)
x <- c(h1$breaks, h2$breaks)
xmax <- max(x)
xmin <- min(x)
plot(h1, ylim = c(hmin, hmax), col = "cadetblue2", xlim = c(xmin, xmax),
     main = "Propensity Score Distributions",
     xlab = "Propensity Scores", axes = FALSE)
lines(h2, col = "chartreuse3")
axis(1, at = seq(0, 1, by = .1))
axis(2, at = seq(-1000, 200, by = 200), labels = c(1000,800,600,400,200,0,200), las = 2, hadj = .8)
legend("bottomright", legend = c("Treated","Untreated"),
       fill = c("cadetblue2","chartreuse3"))
# Find max untreated
max.untreated <- max(pscore[data$Depression == 0])

# Find which treated individuals have higher propensity scores
rm.ind <- which(pscore[data$Depression == 1] > max.untreated)

# Find min treated
min.treated <- min(pscore[data$Depression == 1])

# Find which untreated individuals have lower propensity scores
```

```r
#which(pscore[data$Depression == 0] < min.treated)

# There are no untreated individuals with lower propensity scores than the lowest treated individual

# Remove the individuals identified above
data <- data[-rm.ind,]
pscore <- pscore[-rm.ind]
# Create logit function and then match on propensity scores
logit <- function(p){log(p)-log(1-p)}
psmatch <- Match(Tr = data$Logical.Depression, M = 3, X = logit(pscore), replace = FALSE, caliper = 1.5)
matched <- data[c(psmatch$index.treated[seq(1,length(psmatch$index.treated), by = 3)],psmatch$index.cont


# Create a matched table one
tab2 <- CreateTableOne(vars = c("Income", "Increasing.Exercise", "Healthy.Diet"), strata = "Depression"

# Export to kable
kableone(print(tab2, smd = TRUE), caption = "Matched Demographics and Health Characteristics")
# Response variable separated by treatment
obesity_trt <- as.numeric(matched$Obesity[matched$Logical.Depression==TRUE])-1
obesity_con <- as.numeric(matched$Obesity[matched$Logical.Depression==FALSE])-1

# Calculate numbers for the odds ratio
a <- sum(obesity_trt)
b <- length(obesity_trt)-sum(obesity_trt)
c <- sum(obesity_con)
d <- length(obesity_con)-sum(obesity_con)

# Proportions of obese in depressed and not depressed
p_trt <- a/length(obesity_trt)
p_con <- c/length(obesity_con)
```

```r
# Calculate odds ratio
odds <- (a*d)/(b*c)

# Calcualte SE for log odds ratio
se <- sqrt(1/a+1/b+1/c+1/d)

# CI
ci <- c("Lower" = exp(log(odds) - 1.96*se), "Upper" = exp(log(odds) + 1.96*se))

# Create outcome table
response.test <- mcnemar.test(table(matched[,c("Depression","Obesity")]))
# Sensitivity analysis

# Write function to rerun outcome analysis, but vary a few parameters
sensitivity <- function(caliper, matches, propformula, topincome){
  data <- na.omit(data5)
  colnames(data)[11:16] <- c("Race","Income","Education.Level","BMI","Increasing.Exercise","Healthy.Die
  data$Depression <- NA
  data$Obesity <- as.numeric(data$BMI > 30)
  for(i in 1:nrow(data)){
    if(data$DPQ090[i] > 0 | data$DPQ010[i] > 2 | data$DPQ020[i] > 2 | data$DPQ030[i] > 2 | data$DPQ040[
      data$Depression[i] <- 1
    }else{
      data$Depression[i] <- 0
    }
  }
  data$Logical.Depression <- as.logical(data$Depression)
  data$Race <- factor(data$Race)
  data$Education.Level <- factor(data$Education.Level)
  data$Increasing.Exercise <- factor(data$Increasing.Exercise)
  data$Healthy.Diet <- factor(data$Healthy.Diet)
  data$Depression <- factor(data$Depression)
```

```
data$Obesity <- factor(data$Obesity)
data <- data[-c(2:10,14)]
data <- data[!(data$Income == 13 | data$Income == 12 | data$Income == 77 | data$Income == 99),]
rownames(data) <- 1:nrow(data)
newincome <- numeric(nrow(data))
for(i in 1:nrow(data)){
  if(data$Income[i] == 1){
    newincome[i] <- 2500
  }
  if(data$Income[i] == 2){
    newincome[i] <- 7500
  }
  if(data$Income[i] == 3){
    newincome[i] <- 12500
  }
  if(data$Income[i] == 4){
    newincome[i] <- 17500
  }
  if(data$Income[i] == 5){
    newincome[i] <- 22500
  }
  if(data$Income[i] == 6){
    newincome[i] <- 30000
  }
  if(data$Income[i] == 7){
    newincome[i] <- 40000
  }
  if(data$Income[i] == 8){
    newincome[i] <- 50000
  }
  if(data$Income[i] == 9){
    newincome[i] <- 60000
```

```
    }
    if(data$Income[i] == 10){
      newincome[i] <- 70000
    }
    if(data$Income[i] == 14){
      newincome[i] <- 87500
    }
    if(data$Income[i] == 15){
      newincome[i] <- topincome
    }
  }
  data$Income <- newincome
  psmodel <- glm(propformula,
                 family = binomial,
                 data = data)
  pscore <- psmodel$fitted.values
  psmatch <- Match(Tr = data$Logical.Depression, M = matches, X = logit(pscore), replace = FALSE, calipe
  matched <- data[c(psmatch$index.treated[seq(1,length(psmatch$index.treated), by = matches)],psmatch$in
  obesity_trt <- as.numeric(matched$Obesity[matched$Logical.Depression==TRUE])-1
  obesity_con <- as.numeric(matched$Obesity[matched$Logical.Depression==FALSE])-1
  a <- sum(obesity_trt)
  b <- length(obesity_trt)-sum(obesity_trt)
  c <- sum(obesity_con)
  d <- length(obesity_con)-sum(obesity_con)
  p_trt <- a/length(obesity_trt)
  p_con <- c/length(obesity_con)
  odds <- (a*d)/(b*c)
  response.test <- mcnemar.test(table(matched[,c("Depression","Obesity")]))

  return(list("odds" = odds, "pval" = response.test$p.value))
}
# Use the function to check a few different variations on our original analysis
```

```r
# Analysis with all possible covariates included in the propensity scores
sens1 <- sensitivity(caliper = 1.5, matches = 3, propformula = Depression ~ Income + Increasing.Exercise

# Analysis with different calipers
sens2 <- sensitivity(caliper = .1, matches = 3, propformula = Depression ~ Income + Increasing.Exercise
sens3 <- sensitivity(caliper = .5, matches = 3, propformula = Depression ~ Income + Increasing.Exercise
sens4 <- sensitivity(caliper = 1, matches = 3, propformula = Depression ~ Income + Increasing.Exercise +
sens5 <- sensitivity(caliper = 1.5, matches = 3, propformula = Depression ~ Income + Increasing.Exercise
sens6 <- sensitivity(caliper = 2, matches = 3, propformula = Depression ~ Income + Increasing.Exercise +

# Analysis with different matches
sens7 <- sensitivity(caliper = 1.5, matches = 1, propformula = Depression ~ Income + Increasing.Exercise
sens8 <- sensitivity(caliper = 1.5, matches = 2, propformula = Depression ~ Income + Increasing.Exercise

# Analysis with different top incomes
sens9 <- sensitivity(caliper = 1.5, matches = 3, propformula = Depression ~ Income + Increasing.Exercise
sens10 <- sensitivity(caliper = 1.5, matches = 3, propformula = Depression ~ Income + Increasing.Exerci
sens11 <- sensitivity(caliper = 1.5, matches = 3, propformula = Depression ~ Income + Increasing.Exerci
# Table of manipulations and results

# Create data frame of combinations
sensitivity.df <- data.frame(Caliper = c(.1,.5,1,1.5,2,1.5,1.5,1.5,1.5,1.5), Matches = c(3,3,3,3,3,1,2,3

# Reformat currency
sensitivity.df$Income <- currency(sensitivity.df$Income,digits = 0)

# Export to kable
kable(sensitivity.df, digits = 3, caption = "Sensitivity Analysis")
```