Based on a paper published in [JAMA](https://jamanetwork.com/journals/jama/fullarticle/2608211) last year, the weight gain is increasing among US adults while there is no difference in the percentage of people that were trying to lose weight. The authors used the data from the National Health and Nutrition Examination Survey [NHANES](https://www.cdc.gov/nchs/nhanes/index.htm) from 1988 to 2014 and calculated the proportion of people tried to lose weight during the past 12 months.

Logistic regression analysis was used to estimate the odds of trying to lose weight for different periods from 1988 to 2014. I would like to reproduce these findings with R, given that NHANES data is publically available.

**Libraries and Datasets**

Load libraries

library(tidyverse)

library(RNHANES)

library(weights)

library(ggsci)

library(ggthemes)

The authors utilized the data from 1988 to 2014, but since the RNHANES package enables the data starting from 1999, I will include data from 1999 to 2014. Therefore, I cannot fully reproduce the results of the paper publish at JAMA.

Load the data using nhanes\_load\_data function from RNHANES package.

d99 = nhanes\_load\_data("DEMO", "1999-2000") %>%

select(SEQN, cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR) %>%

transmute(SEQN=SEQN, wave=cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR) %>%

left\_join(nhanes\_load\_data("BMX", "1999-2000"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI) %>%

left\_join(nhanes\_load\_data("WHQ", "1999-2000"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI, WHQ070)

d01 = nhanes\_load\_data("DEMO\_B", "2001-2002") %>%

select(SEQN, cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR) %>%

transmute(SEQN=SEQN, wave=cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR) %>%

left\_join(nhanes\_load\_data("BMX\_B", "2001-2002"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI) %>%

left\_join(nhanes\_load\_data("WHQ\_B", "2001-2002"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI, WHQ070)

d03 = nhanes\_load\_data("DEMO\_C", "2003-2004") %>%

select(SEQN, cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR) %>%

transmute(SEQN=SEQN, wave=cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR) %>%

left\_join(nhanes\_load\_data("BMX\_C", "2003-2004"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI) %>%

left\_join(nhanes\_load\_data("WHQ\_C", "2003-2004"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI, WHQ070)

d09 = nhanes\_load\_data("DEMO\_F", "2009-2010") %>%

select(SEQN, cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMIN2, WTINT2YR, WTMEC2YR) %>%

transmute(SEQN=SEQN, wave=cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC=INDFMIN2, WTINT2YR, WTMEC2YR) %>%

left\_join(nhanes\_load\_data("BMX\_F", "2009-2010"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI) %>%

left\_join(nhanes\_load\_data("WHQ\_F", "2009-2010"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI, WHQ070)

d11 = nhanes\_load\_data("DEMO\_G", "2011-2012") %>%

select(SEQN, cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMIN2, WTINT2YR, WTMEC2YR) %>%

transmute(SEQN=SEQN, wave=cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC=INDFMIN2, WTINT2YR, WTMEC2YR) %>%

left\_join(nhanes\_load\_data("BMX\_G", "2011-2012"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI) %>%

left\_join(nhanes\_load\_data("WHQ\_G", "2011-2012"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI, WHQ070)

d13 = nhanes\_load\_data("DEMO\_H", "2013-2014") %>%

select(SEQN, cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMIN2, WTINT2YR, WTMEC2YR) %>%

transmute(SEQN=SEQN, wave=cycle, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC=INDFMIN2, WTINT2YR, WTMEC2YR) %>%

left\_join(nhanes\_load\_data("BMX\_H", "2013-2014"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI) %>%

left\_join(nhanes\_load\_data("WHQ\_H", "2013-2014"), by="SEQN") %>%

select(SEQN, wave, RIAGENDR, RIDAGEYR, RIDRETH1, RIDEXPRG, INDFMINC, WTINT2YR, WTMEC2YR, BMXBMI, WHQ070)

The list of variables I will use in analysis:

* SEQN: id
* RIAGENDR: gender
* RIDAGEYR: age in years
* BMXBMI: body mass index
* INDFMINC: family income
* RIDRETH1: race/ethnicity
* RIDEXPRG: pregnancy
* WHQ070: intention to lose weight
* WTINT2YR: weighted variable for questionnaires
* WTMEC2YR: weighted variable for laboratory measurements

Merging the datasets, creating new variables, and excluding pregnant women.

dat = rbind(d99, d01, d03, d09, d11, d13) %>%

mutate(

race = recode\_factor(RIDRETH1,

`1` = "Mexian American",

`2` = "Hispanic",

`3` = "Non-Hispanic, White",

`4` = "Non-Hispanic, Black",

`5` = "Others"),

bmi = if\_else(BMXBMI >= 30, "Obese", "Overweight"),

pregnancy = if\_else(RIDEXPRG == 1, "Yes", "No", "No"),

tryweloss = if\_else(WHQ070 == 1, "Yes", "No")

) %>%

filter(BMXBMI >= 25, RIDAGEYR >=20, RIDAGEYR < 60, race != "Others", pregnancy == "No", WHQ070 %in% c(1,2))

I excluded pregnant women from the analysis as they are overweight or obese due to pregnancy and might have no intention to lose weight during pregnancy. The R code below calculates the weighted proportion of overweight and obesity as well as intention to lose weight for the total population.

with(dat, wpct(bmi, weight = WTMEC2YR))

*## Obese Overweight*

*## 0.4997831 0.5002169*

with(dat, wpct(tryweloss, weight = WTINT2YR))

*## No Yes*

*## 0.5235062 0.4764938*

The proportion of obese and overweight is 50%, as I excluded the normal weight individuals before. The intention to lose weight in the total population is 48%. Now, I will evaluate the “intention to lose weight” among overweight and obese people in different periods. (I will repeat the same code for each period studied. If you know how to short the code below please let me know. I don't know how to weight the proportion with the tidyverse package)

with(dat %>% filter(wave == "1999-2000", bmi == "Obese"),

wpct(tryweloss, weight = WTINT2YR))

*## No Yes*

*## 0.4555818 0.5444182*

with(dat %>% filter(wave == "1999-2000", bmi == "Overweight"),

wpct(tryweloss, weight = WTINT2YR))

*## No Yes*

*## 0.6000436 0.3999564*

There are 54% obese and 40% overweight people who are trying to lose weight in 1999-2000. Using the above, I calculated the proportions for other periods and made a table. I will focus only on those who lost weight to see how the trend goes over years in overweight and obese.

trend <- read.table(header=TRUE, text='

year bmi weightloss

1999-2000 Obese 54.44182

1999-2000 Overweight 39.99564

2001-2002 Obese 54.14932

2001-2002 Overweight 38.60543

2003-2004 Obese 53.6699

2003-2004 Overweight 42.97784

2009-2010 Obese 53.62087

2009-2010 Overweight 35.6233

2011-2012 Obese 57.5303

2011-2012 Overweight 41.99084

2013-2014 Obese 58.32048

2013-2014 Overweight 40.19653

')

trend %>%

ggplot(aes(year, weightloss, fill = bmi, group = bmi)) +

geom\_line(color = "black", size = 0.3) +

geom\_point(colour="black", pch=21, size = 3) +

theme(text = element\_text(family = "serif", size = 11), legend.position="bottom") +

scale\_fill\_jama() +

theme\_hc() +

labs(

title = "Trends of 'trying to lose weight' in the United States",

caption = "NHANES 1999-2014 survey\ndatascienceplus.com",

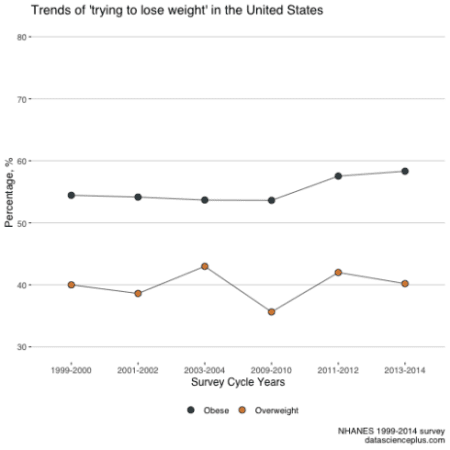
x = "Survey Cycle Years",

y = "Percentage, %",

fill = ""

) +

scale\_y\_continuous(limits=c(30,80), breaks=seq(30,80, by = 10))

Gives this plot:  
[](https://i2.wp.com/datascienceplus.com/wp-content/uploads/2018/07/unnamed-chunk-8-1-1.png?ssl=1)

There is a small increasing trend to lose weight in people with obesity in recent years.

**Logistic Regression**

A logistic regression model with *Poisson* distribution is used to evaluate the odds of intention to lose weight tryweloss in different time periods. In this analysis, several confounders were considered such as BMI status, gender, age, family income, and race.

mylogit <- glm(tryweloss2 ~ wave + bmi + RIAGENDR + RIDAGEYR + INDFMINC + race,

data = dat,

family = "poisson")

summary(mylogit)

*##*

*## Call:*

*## glm(formula = tryweloss2 ~ wave + bmi + RIAGENDR + RIDAGEYR +*

*## INDFMINC + race, family = "poisson", data = dat)*

*##*

*## Deviance Residuals:*

*## Min 1Q Median 3Q Max*

*## -1.2590 -0.9171 -0.7429 0.6490 1.1270*

*##*

*## Coefficients:*

*## Estimate Std. Error z value Pr(>|z|)*

*## (Intercept) -1.2545203 0.0886399 -14.153 < 2e-16 \*\*\**

*## wave2001-2002 0.0028610 0.0553019 0.052 0.95874*

*## wave2003-2004 0.0277646 0.0562594 0.494 0.62165*

*## wave2009-2010 -0.0434281 0.0519876 -0.835 0.40352*

*## wave2011-2012 0.0377786 0.0539726 0.700 0.48395*

*## wave2013-2014 0.0929779 0.0523326 1.777 0.07562 .*

*## bmiOverweight -0.3519162 0.0306745 -11.473 < 2e-16 \*\*\**

*## RIAGENDR 0.4184619 0.0303948 13.768 < 2e-16 \*\*\**

*## RIDAGEYR -0.0028538 0.0013459 -2.120 0.03398 \**

*## INDFMINC 0.0008816 0.0010161 0.868 0.38559*

*## raceHispanic 0.0847514 0.0584763 1.449 0.14725*

*## raceNon-Hispanic, White 0.1199298 0.0387668 3.094 0.00198 \*\**

*## raceNon-Hispanic, Black -0.0016056 0.0436180 -0.037 0.97064*

*## ---*

*## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1*

*##*

*## (Dispersion parameter for poisson family taken to be 1)*

*##*

*## Null deviance: 7330.3 on 10232 degrees of freedom*

*## Residual deviance: 6923.2 on 10220 degrees of freedom*

*## (88 observations deleted due to missingness)*

*## AIC: 16273*

*##*

*## Number of Fisher Scoring iterations: 5*

The results show that there are no significant odds (p>0.05) on the intention to lose weight when I compare different periods to 1999-2000. The small sample size could explain these findings for each survey.

However, let me get the exponential of these coefficients and confidence intervals.

exp(coef(mylogit))

*## (Intercept) wave2001-2002 wave2003-2004*

*## 0.2852126 1.0028651 1.0281537*

*## wave2009-2010 wave2011-2012 wave2013-2014*

*## 0.9575014 1.0385013 1.0974375*

*## bmiOverweight RIAGENDR RIDAGEYR*

*## 0.7033391 1.5196224 0.9971503*

*## INDFMINC raceHispanic raceNon-Hispanic, White*

*## 1.0008820 1.0884464 1.1274177*

*## raceNon-Hispanic, Black*

*## 0.9983957*

exp(confint(mylogit))

*## 2.5 % 97.5 %*

*## (Intercept) 0.2395745 0.3391149*

*## wave2001-2002 0.8999514 1.1178640*

*## wave2003-2004 0.9208583 1.1481368*

*## wave2009-2010 0.8650406 1.0606215*

*## wave2011-2012 0.9344651 1.1546960*

*## wave2013-2014 0.9907765 1.2164297*

*## bmiOverweight 0.6622089 0.7468264*

*## RIAGENDR 1.4318982 1.6130974*

*## RIDAGEYR 0.9945248 0.9997859*

*## INDFMINC 0.9988443 1.0028327*

*## raceHispanic 0.9697129 1.2196128*

*## raceNon-Hispanic, White 1.0452581 1.2168270*

*## raceNon-Hispanic, Black 0.9166538 1.0876073*

The odds ratio and 95% confidence intervals (95%CI) of losing weight in the period 2013-2014 is 1.10 (95% CI 0.99-1.22) compare to 1999-2000. The odds ratio (95%CI) of losing weight in overweight compared to obese individuals is significantly lower 0.70 (95% CI 0.66-0.75) indicating that obese individuals are trying more to lose weight than overweight (as we can expect).

My results were different from the paper published in JAMA. There are some differences from this analysis compared to the article. My reference is the period 1999-2000, whereas in the paper is 1988-1994. Moreover, I excluded pregnant women, when the paper keeps those in the analysis.