The Effects of Heavy Drinking, Smoking, Anxiety and Depression on Stroke Odds: A Cross-Sectional Study

Robert Wardrup

Elizabeth Murphy

Natalie Requinez

Joshua Averitt

Though numerous studies have investigated the relationships between lifestyle risk factors and stroke outcome, few have done so while controlling as age and race/ethnicity range. Utilizing quasi-binomial logistic regression, we show that the odds of having had a stroke increase with being a daily smoker, days of anxiety, and having a depression diagnosis.

Keywords: Stroke, Alcohol, Smoking, Anxiety, Depression

Introduction

Strokes nearly affect 800,000 people in the U.S every year and is the fifth leading cause of death in the U.S (American Stroke Association, 2017a). Despite increased efforts on education and general public knowledge about stroke and other cardiovascular risk diseases, the prevalence rates have shown little improvement within the past ten years (Mozaffarian et al., 2015). Previous research has shown that a variety of controllable risk factors that are strongly associated with stroke outcome. Although behaviors such as smoking and poor diet have been well-known risk factors, other factors such as stress, lack of sleep, depression, and anxiety still have unclear associations [(Risk Factors Stroke Association, (American Stroke Association, 2017b).

Stress, lack of sleep, and mental disorders such as depression and anxiety are issues prominent throughout the United States. Anxiety is the most common mental illness and approximately 18% of the U.S population is affected while depression affects nearly 7% of adults (Anxiety and Depression Association of America, 2017). Those who are diagnosed with anxiety also have a 50%

chance of being diagnosed with depression and vice versa (Anxiety and Depression Association of America, 2017). Furthermore, approximately, 30% of Americans suffer from lack of sleep (Centers for Disease Control and Prevention, 2015). Although average stress levels have decreased since the recession, the average stress scores for American adults is still at moderate levels and is increased with lower household incomes (American Psychological Association, 2015). Each of these factors has a considerable significance to public health as issues on their own. If someone is suffering from multiple risk factors, the risk of stroke could be significant and possibly result in death.

The goal of this study is to test the if there is an association between the different risk factors, particularly with stress, lack of sleep, depression, and anxiety, with the outcome stroke. This will be compared to previous trends in the same survey data to see if results are consistent and hopefully contribute additional knowledge. This will be determined using a chi squared analysis and logistics regressions collected from the BRFSS National Survey. The BRFSS National Survey collects data from 50 states, D.C, and the even throughout the territories collecting telephone survey data. It is in partnership with the CDC and has been useful for providing large quantities of health related data (Centers for Disease Control and Prevention, 2014). The results of this study will be compared to previous trends in the same survey data to see if results are consistent and hopefully contribute additional knowledge to help prevent stroke and improve quality of life.

Methods

Josh's Text

Survey Weighting

The BRFSS survey data were weighted using the raking method, which is a two-part methodology to help insure unbiased results by accounting for noncoverage and nonresponse bias and forcing the total number of cases to equal the population estimate of each state in the United States (Centers for Disease Control and Prevention, 2007). Raking works by repeatedly adjusting weight across a set of selected variables until the weights converge and the survey population totals are equal to the census population totals for each selected variable (Fricker and Andersen, 1993).

Dependent and Independent Variables

The dependent variable, stroke outcome, was recoded to a binary variable for use in a logistic regression model. Depression diagnosis and heavy drinker were also recoded into a binary variable. For each variable used in the analysis, *don't know / unsure* and *refused* responses were dropped. Table 1 lists the variables included in the model.

Table 1: Dependent and Independent Variables Used in Logistic Regression Model

Variable Name	Variable Type
Stroke Diagnosis	Dichotomous (DV)
Heavy Alcohol Drinker	Dichotomous
Depression Diagnosis	Dichotomous
Sex	Dichotomous
Days Anxious	Continuous
Daily Sleep Hours	Continuous
Smoker Status	Categorical
Race/Ethnicity	Categorical
Age Group	Categorical

A subject was designated as a heavy drinker if they were either an adult male who reported consuming more than 14 drinks per week or an adult female who reported consuming more than 7 drinks per week. This value was calculated after asking the subjects, "during the past 30 days, on the days when you drank, about how many drinks did you drink on average?" and "during the past 30 days, how many days per week or month did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage or liquor?" (**?**).

Logistic Regression Model

The first model used in calculating the logistic regression utilized the standard binomial function to predict the log odds of binary outcome k. The binomial formula, is given as

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where k represents a stroke outcome, n represents the sample size, and p is the probability that a stroke will occur. [**Explain overdispersion**] Before being checked for overdispersion by calculating ϕ, given by the formula

$$\phi = \frac{1}{(n-p-1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \hat{y}_i.$$

A threshold of $\phi > 1$, was used to determine if the data were overdispersed, possibly leading to unstable estimates. To account for the overdispersion, ϕ was included as a model parameter, giving the formula

$$P(k) = \binom{n}{k} p(p + k\phi)^{k-1} (1 - p - k\phi)^{n-k}.$$

[**Explain the quasi-binomial family more**]

Results

The weighted number of strokes in the United States in 2016 was 8,020,080, representing approximately 2.5% of the total 2016 United States population. Out of these, 3.20% were female and 3.13% were male.

Table 2: Proportion of Stroke Outcomes in Males and Females

Diagnosis	Male	Female
No Stroke	96.87%	96.80%
Stroke	3.13%	3.20%

A chi-squared test of independence revealed that there there is not statistically significant evidence that stroke outcome and sex is independent from one another at an alpha level of 0.05, χ^2 (3, N=486,303) = 1.73, p=0.48.

The proportion of subjects who have suffered from a stroke by smoker status is shown in Table 3, and visually in Figure 2. The percentage of subjects who have had a stroke is highest for the subjects who smoke daily, followed by subjects who are former smokers, followed by subjects who smoke some days. A possible explanation for the lower percentage of stroke in subjects who smoke some days than subjects who are former smokers is that stroke victims may simply have ceased smoking after having a stroke.

Table 3: Proportion of Stroke Outcomes and Smoker Status

Smoker Status	No Stroke	Stroke
Never Smoked	97.8%	2.2%
Former Smoker	95.4%	4.6%
Smokes Some Days	96.2%	3.8%
Smokes Daily	95.0%	5.0%

The proportions shown in Table 3 are depicted as a mosaic plot in Figure 2. The mosaic plot is a graphical depiction the proportions within each table cell, shaded by the difference from the expected observation. Blue depicts a higher than expected number of observations for that cell, and red depicts a lower than expected number of observations for that cell. As shown in Figure 2, there were a greater than expected number of non-smokers that have never had a stroke, and a greater than expected number of smokers and former smokers who have had a stroke.

A chi-squared test of independence revealed that there is statistically significant evidence that stroke outcome and smoking status were not independent from one another, χ^2 (3, N=486,303) = 2352.5, $p < 0.001$.

The mean hours of sleep for those who had a stroke was 6.99 hours and 6.98 for those who did not have a stroke.

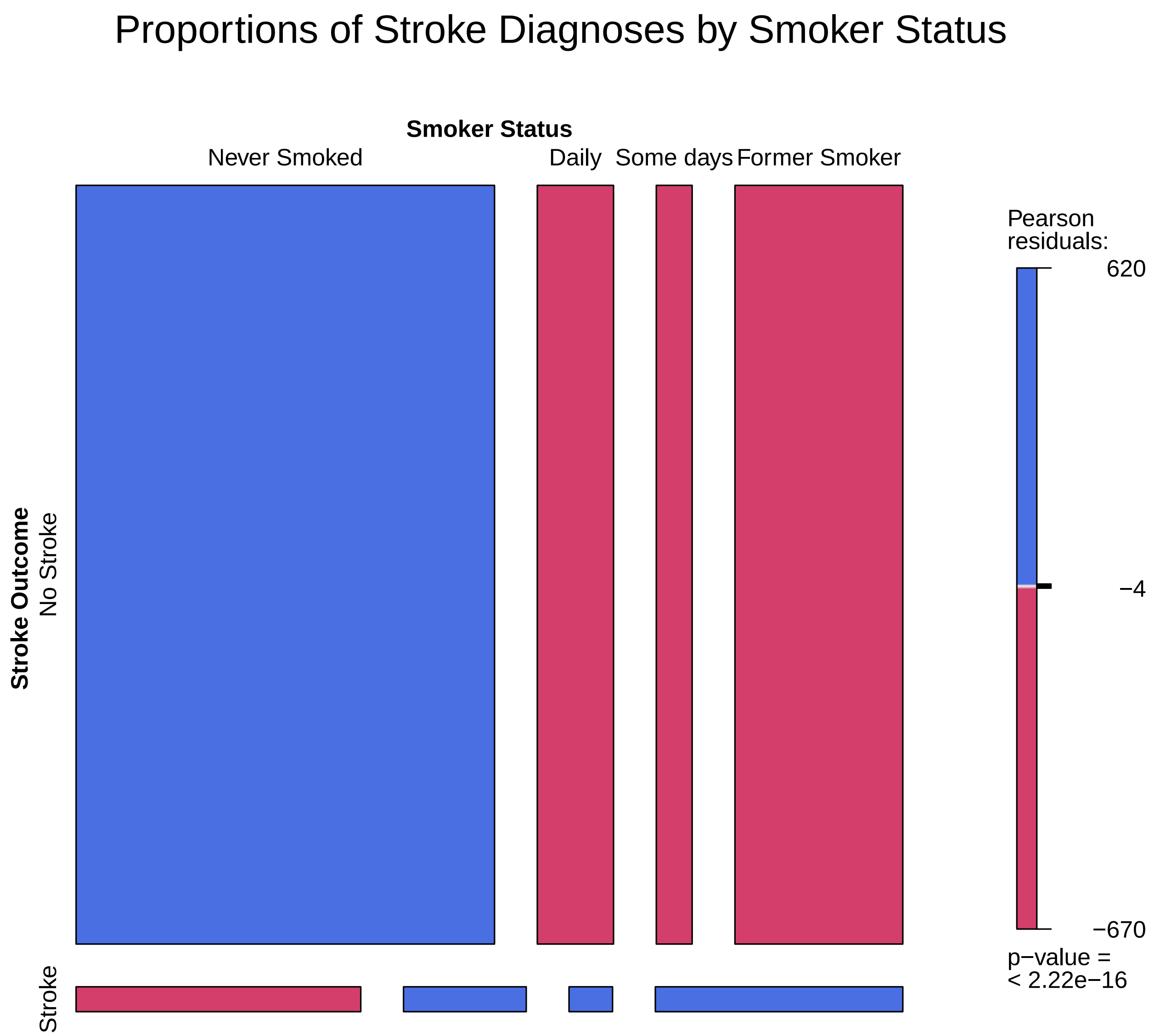

Figure 1: Proportion of Stroke Outcomes and Smoker Status

Table 4: Mean Values of Selected Risk Factors

Factor	Outcome	Mean
Days Anxious	No Stroke	4.79
	Stroke	8.05
Nightly Sleep Hours	No Stroke	6.98
	Stroke	6.99
Number Drinks Weekly	No Stroke	79.07
	Stroke	40.31

Logistic Regression Model

The binomial logistic regression model resulted in a ϕ value of 2.52, indicating that there is overdispersion in the estimates and the quasi-binomial function should be used instead.

The results of the quasi-binomial logistic regression (Table 4) indicate that, controlling for age, race/ethnicity, and sex, Several of the tested risk factors were significantly correlated with stroke outcome. Being a daily smoker, Monthly days with anxiety, and diagnosed depression were all positively associated with greater odds of stroke.

Daily cigarette smokers had 1.68 times the odds of having had a stroke than those who have never smoked cigarettes. For each additional day heavily affected by anxiety,the odds of having had a stroke were 1.03 times higher, and for those who have been diagnosed with depression the odds of stroke were 2.12 times higher. Of the control variables, age of 25 years old and above were the only factors significantly associated with stroke outcome, with an odds ratio that increased with age.

Variable	OR (95% CI)	P-Value

Table 5: Logistic Regression Model Estimating Effects of Risk Factors & Demographic Variables on Stroke Outcome Odds.

Variable	OR (95% CI)	P-Value
Risk Factors		
Drinking		
Non-Heavy Drinker	1 (Baseline Factor)	
Heavy Drinker	0.989 (0.41 - 2.38)	0.98
Smoking		
Non-Smoker	1 (Baseline Factor)	
Former Smoker	1.24 (0.91 - 1.70)	0.18
Smoker (Some days)	1.65 (0.95 - 2.88)	0.08
Smoker (Daily)	1.68 (1.12 - 2.53)	0.01
Mental Health		
Monthly Days Anxious	1.03 (1.01 - 1.04)	<0.01
No Depression Diagnosis	1 (Baseline Factor)	
Depression Diagnosis	2.12 (1.48 - 3.04)	<0.01
Daily Sleep Hours	0.91 (0.82 - 1.00)	0.06
Sex		
Male	1 (Baseline Factor)	
Female	0.77 (0.58 - 1.02)	0.07
Race/Ethnicity		
Other Race / Non-Hispanic	1 (Baseline Factor)	
Hispanic	1.42 (0.43 - 4.70)	0.57
Black Only / Non-Hispanic	2.41 (0.88 - 6.62)	0.09
White Only / Non-Hispanic	1.87 (0.73 - 4.83)	0.20

Variable	OR (95% CI)	P-Value
Multiracial / Non-Hispanic	2.92 (0.90 - 9.50)	0.08
Age Group		
Age 18 to 24	1 (Baseline Factor)	
Age 25 to 34	3.92 (0.80 - 19.66)	0.10
Age 35 to 44	9.33 (2.00 - 43.51)	<0.01
Age 45 to 54	12.35 (2.79 - 54.71)	<0.01
Age 55 to 64	20.93 (4.80 - 91.34)	<0.01
Age 65 or Older	44.87 (10.36 - 194.43)	<0.01

The effect of each significant risk factor, stratified by depression diagnosis, on the probability of having had a stroke is shown in Figure 2. The probability of a person having a stroke increases with the number of days a person is significantly affected by anxiety. This probability is multiplied if a person has been diagnosed with depression. For non-smokers with no depression diagnosis who did not signifantly suffer from anxiety, the probability of having had a stroke approaches zero. If the same subject was diagnosed with depression, the probability of having had a stroke nears 5%.

The probability of having had a stroke is multiplied in daily smokers, beginning at approximately 5% for daily smokers who have not been diagnosed with depression nor have suffered from significant anxiety. With a depression diagnosis, the probability originates at approximately 7% for subjects with zero days of anxiety and increases to 15% for subjects who experienced thirty days of signiticant anxiety. This relationship is similar for subjects who currently smoke some days.

Conclusion

Our study assessed the association between heavy drinking, anxiety, depression on stroke odds. We used the 2016 Behavioral Risk Factor Surveillance System dataset from the CDC. There was a significant difference between being a daily smoking and odds of a stroke as the p-value was .01. Also, there was a significant difference for anxiety, and depression for the odds of a stroke, p-value

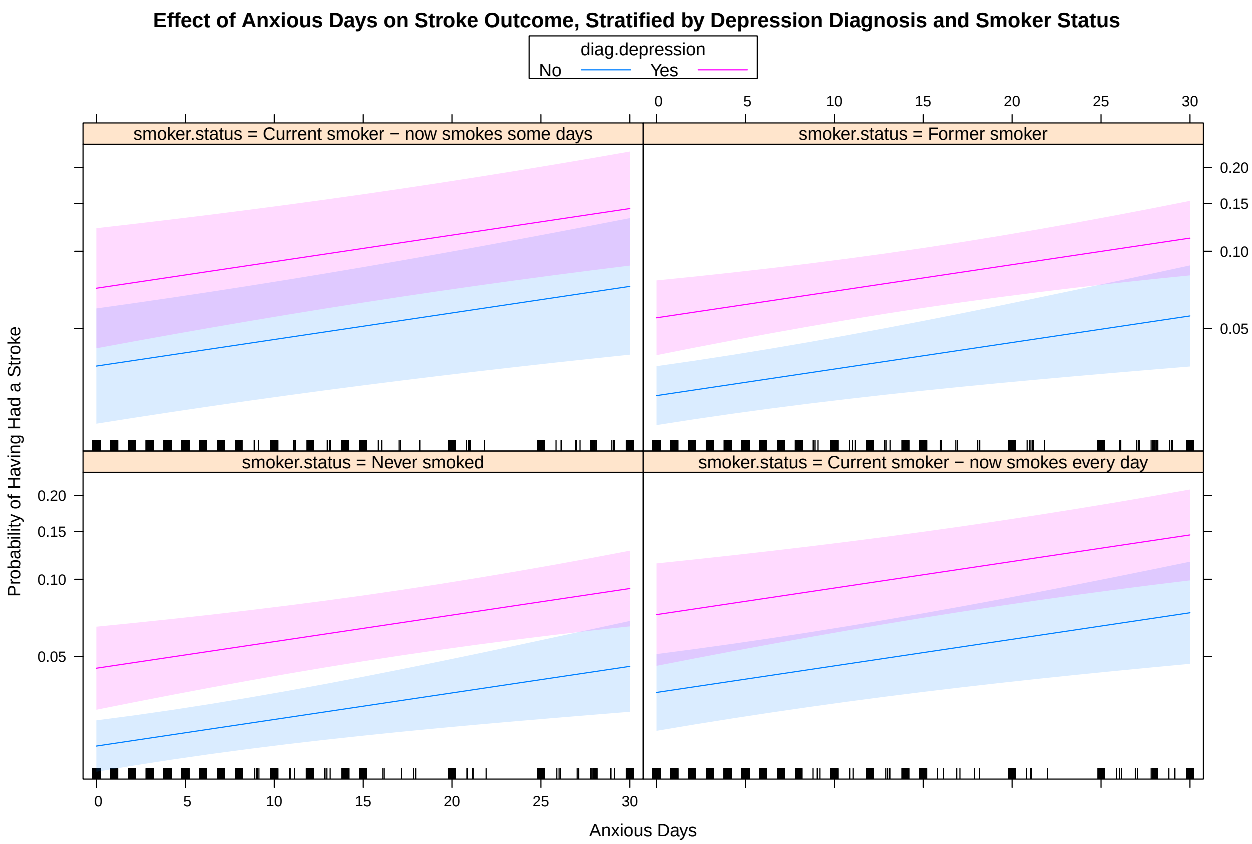

Figure 2: Effects of Anxious Days on Stroke Outcome, Stratified by Depression Diagnosis and Smoker Status

was less than .01 for both variables. We found know no significant different between race or the amount of sleep hours someone receives., and whether or not an individual is a heavy drinker.

Our study has a specific limitation due to its study design. The study was a cross-sectional one, causality and risk cannot be inferred from our data. Each variable tested affected one another, so heavy drinking influenced the variables anxiety and depression. And anxiety can affect heavy drinking and depression. The influence of variables on each other demonstrates that causality cannot be inferred in a cross-sectional study.

In conclusion, variables depression, anxiety, and being a daily smoker all were significantly different for stroke odds. This conclusion means that an individual who is either a daily smoker, diagnosed with depression or anxiety all have an increased change of having a stroke. The 2016 BRFSS dataset provided a large population data set to help monitor variables that affect stroke odds. Future studies can look at the association can look further at the effect of mental health diagnosis on stroke odds, as well as how smoking can increase an individual's odds of a stroke.

References

American Psychological Association (2015). Paying with our health. *Stress in America*, pages 1–23.

American Stroke Association (2017a). Impact of Stroke (Stroke statistics).

American Stroke Association (2017b). Understanding Stroke Risk.

Anxiety and Depression Association of America (2017). Facts & Statistics | Anxiety and Depression Association of America, ADAA.

Centers for Disease Control and Prevention (2007). Behavioral Risk Factor Surveillance System Survey (BRFSS) data. 2009(9/8/2008).

Centers for Disease Control and Prevention (2014). CDC - About BRFSS.

Centers for Disease Control and Prevention (2015). Insufficient Sleep Is a Public Health Problem | Features | CDC.

Fricker, R. D. and Andersen, L. (1993). Raking: An important and often overlooked survey tool. Technical report.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., De Ferranti, S., Després, J. P., Fullerton, H. J., Howard, V. J., Huffman, M. D., Judd, S. E., Kissela, B. M., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Liu, S., Mackey, R. H., Matchar, D. B., McGuire, D. K., Mohler, E. R., Moy, C. S., Muntner, P., Mussolino, M. E., Nasir, K., Neumar, R. W., Nichol, G., Palaniappan, L., Pandey, D. K., Reeves, M. J., Rodriguez, C. J., Sorlie, P. D., Stein, J., Towfighi, A., Turan, T. N., Virani, S. S., Willey, J. Z., Woo, D., Yeh, R. W., and Turner, M. B. (2015). *Heart disease and stroke statistics-2015 update : A report from the American Heart Association*, volume 131.

Code

Import Data

```
# This script downloads the required data from the internet
# and saves it to ./Data/Raw

source("./Functions/detachAll.R") # Detach all packages to help with plyr/dplyr issues
detachAllPackages()
library(foreign)
library(openxlsx)
library(fiftystater)

# Download the 2016 BRFSS Data from CDC ----
OUTFILE = "./Data/Raw/LLCP2016.xpt "
URL <- "https://www.cdc.gov/brfss/annual_data/2016/files/LLCP2016XPT.zip"
if (!file.exists(OUTFILE))
{
  # Only download if file doesn't already exist
  ZIPFILE = "./Data/Raw/LLCP2016XPT.zip"
  download.file(URL, ZIPFILE, method = "auto")
  unzip(zipfile = ZIPFILE, overwrite = TRUE) # Extract data
  file.remove(ZIPFILE) # Delete zipfile
```

```r
  rm(ZIPFILE) # Clean up cruft
  brfss <- read.xport(OUTFILE) #Read in the data
  rm(OUTFILE)
  rm(URL)
  file.remove(ZIPFILE)
} else
{
  # If file exists, read it into memory
  brfss <- read.xport(OUTFILE)
  rm(OUTFILE)
  rm(URL)
}

# Get CSV file of BRFSS state designations ----
CSVPATH <- "./Data/Raw/brfss_state.csv"
download.file("https://raw.githubusercontent.com/arilamstein/complexsurvey/master/brfss_state.csv
    ",
  CSVPATH)
state.codes <- read.csv(CSVPATH)
rm(CSVPATH)

# Get the census data ----
CENSUS_PATH <- "./Data/Raw/population.xlsx"
download.file("https://www2.census.gov/programs-surveys/popest/tables/2010-2016/state/totals/
    nst-est2016-01.xlsx",
  CENSUS_PATH)
census <- readWorkbook(CENSUS_PATH, sheet = 1, startRow = 1,
  colNames = TRUE, rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
  skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
  namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)

census <- census[-c(seq(1, 8)), ] # Remove census rows 1-8
census <- census[-c(seq(53, 57)), ] # Remove census rows 53-57
census <- census[, c(1, 10)] # Only keep state name column and 2016 population column
colnames(census) <- c("STATE", "Pop.2016") # Replace column names
census$STATE <- substring(census$STATE, 2) # Remove leading '.' character from state names

# Cleanup
rm(CENSUS_PATH)
gc()
```

Parse Data

```r
# This script parses the data downloaded in 01_Import.R,
# creates the survey design object, and saves it to
# ./Data/transformed.rds for use in step 4.

library(Hmisc)
library(classInt)
library(dplyr)
library(survey)
library(srvyr)
```

```
# a single-PSU stratum makes no contribution to the variance
# (for multistage sampling it makes no contribution at that
# level of sampling).
options(survey.lonely.psu = "certainty")

# Add state names to DF
state.codes$X_STATE <- state.codes$VALUE
state.codes <- state.codes[!(state.codes$VALUE == ".D" | state.codes$VALUE ==
  ".R"), ] # These vals don't exist in BRFSS
brfss.survey1 <- merge(brfss, state.codes, by = "X_STATE")
brfss.survey1$STATE <- as.factor(brfss.survey1$STATE)

# Garbage cleanup
rm(brfss)
rm(state.codes)
gc()

# Remove unused columns
brfss.survey1 <- brfss.survey1[, -which(names(brfss.survey1) %in%
  c("X_STATE", "SEQNO", "VALUE"))]

# Recode the BRFSS variables according to the codebook ----
brfss.survey1 <- brfss.survey1 %>% mutate(race.eth = car::recode(X_RACEGR3,
  "1='White␣only,␣Non-Hispanic';␣2='Black␣only,␣Non-Hispanic';␣3='Other␣race␣only,␣Non-
      Hispanic';␣4='Multiracial,␣Non-Hispanic';␣5='Hispanic';␣9=NA"),
  race.eth = as.factor(race.eth), age.grp = car::recode(X_AGE_G,
    "1='Age␣18␣to␣24';␣2='Age␣25␣to␣34';␣3='Age␣35␣to␣44';␣4='Age␣45␣to␣54';␣5='Age␣55␣to␣
        64';␣6='Age␣65␣or␣older'"),
  age.grp = as.factor(age.grp), sex = car::recode(SEX, "1='Male';␣2='Female';␣9=NA"),
  sex = as.factor(sex), daily.sleep.hrs = car::recode(SLEPTIM1,
    "77=NA;␣99␣=NA"), daily.sleep.hrs <- as.integer(daily.sleep.hrs),
  diag.depression = car::recode(ADDEPEV2, "1='Yes';␣2='No';␣7=NA;␣9=NA"),
  diag.depression = factor(diag.depression, labels = c("No␣Depression␣Diagnosis",
    "Depression␣Diagnosis")), days.anxious = car::recode(QLSTRES2,
    "88='0';␣77=NA;␣99=NA"), days.anxious <- as.integer(days.anxious),
  diag.stroke = car::recode(CVDSTRK3, "1=1;␣2=0;␣7=NA;␣9=NA"),
  diag.stroke = factor(diag.stroke, labels = c("No␣Stroke",
    "Stroke")), alc.heavy.drinker = car::recode(X_RFDRHV5,
    "1='No';␣2='Yes';␣9=NA"), alc.heavy.drinker = as.factor(alc.heavy.drinker),
  smoker.status = car::recode(X_SMOKER3, "1␣=␣'Daily';␣2='Some␣days';␣3='Former␣Smoker';␣
      4='Never␣Smoked';␣9=NA"),
  smoker.status = as.factor(smoker.status))

# Only keep the variables we need
brfss.survey1 <- brfss.survey1[, c("X_LLCPWT", "X_STSTR", "X_PSU",
  "STATE", "race.eth", "age.grp", "sex", "daily.sleep.hrs",
  "diag.depression", "diag.stroke", "days.anxious", "alc.heavy.drinker",
  "smoker.status")]

# Flush memory
gc()

# Create the survey design ----
print("Creating␣Survey␣Design␣Object.␣This␣will␣take␣some␣time...")
```

```r
brfss.survey.design <- svydesign(nest = T, ids = ~X_PSU, strata = ~X_STSTR,
  weights = ~X_LLCPWT, data = brfss.survey1)

# Don't need these now. Save RAM.
rm(brfss.survey1)

# Save the transformed data ----
saveRDS(brfss.survey.design, "Data/transformed.rds")
```

Analyse Data

```r
# This script analyzes the data parsed out in 03_Transform.R
# and saves it out to ./Data/results.rds for use in the
# RMarkdown script.

library(ggplot2)
library(ggeffects)
library(srvyr)
library(survey)
library(RColorBrewer)
library(fiftystater)
library(plyr)
library(dplyr)
library(Hmisc)
library(data.table)
library(stargazer)
library(scales)
library(effects)
library(vcd)
source("./Functions/pub_graphs.R") # Nice graphs

plot_odds<-function(x, title = NULL){
  tmp<-data.frame(cbind(exp(coef(x)), exp(confint(x))))
  odds<-tmp[-1,]
  names(odds)<-c('OR', 'lower', 'upper')
  odds$vars<-row.names(odds)
  ticks<-c(seq(.1, 1, by =.1), seq(0, 10, by =1), seq(10, 100, by =10))

  g <- ggplot(odds, aes(y= OR, x = reorder(vars, OR))) +
    geom_point() +
    geom_errorbar(aes(ymin=lower, ymax=upper), width=.2) +
    scale_y_log10(breaks=ticks, labels = ticks) +
    geom_hline(yintercept = 1, linetype=2) +
    coord_flip() +
    labs(title = title, x = 'Variables', y = 'OR') +
    theme_bw()

  return(g)
}

# a single-PSU stratum makes no contribution to the variance
# (for multistage sampling it makes no contribution at that
# level of sampling).
```

```
options(survey.lonely.psu = "certainty")

# Disable scientific notation
options(scipen = 999)

# Initialize an empty list to store the results ----
results <- list()

# Load the transformed data ----
print("Reading␣transformed.rds␣into␣memory...")
brfss.survey.design <- readRDS("Data/transformed.rds")

# Specific transformations ---- Relevel some of the factors
# so that the regression output is easier to interpret.
print(("Setting␣logistic␣regression␣baseline␣levels..."))
brfss.survey.design$variables$sex <- relevel(brfss.survey.design$variables$sex,
  ref = "Male")
brfss.survey.design$variables$diag.depression <- relevel(brfss.survey.design$variables$diag.depression,
  ref = "No␣Depression␣Diagnosis")
brfss.survey.design$variables$smoker.status <- relevel(brfss.survey.design$variables$smoker.status,
  ref = "Never␣Smoked")
brfss.survey.design$variables$race.eth <- relevel(brfss.survey.design$variables$race.eth,
  ref = "Other␣race␣only,␣Non-Hispanic")
brfss.survey.design$variables$age.grp <- relevel(brfss.survey.design$variables$age.grp,
  ref = "Age␣18␣to␣24")
brfss.survey.design$variables$alc.heavy.drinker <- relevel(brfss.survey.design$variables$alc.heavy.
    drinker,
  ref = "No")

# Models ---- Logistic regression on stroke diagnosis status
# ----
print("Running␣binomial␣logistic␣regression...")
results$stroke.binom <- svyglm(diag.stroke ~ alc.heavy.drinker +
  smoker.status + days.anxious + diag.depression + daily.sleep.hrs +
  sex + age.grp + race.eth, brfss.survey.design, family = "binomial")

print("Running␣quasibinomial␣logistic␣regression...")
results$stroke.qbinom <- svyglm(diag.stroke ~ alc.heavy.drinker +
  smoker.status + days.anxious + diag.depression + daily.sleep.hrs +
  sex + age.grp + race.eth, brfss.survey.design, family = "quasibinomial")

results$stroke.nandySuggestions <- svyglm(diag.stroke ~ alc.heavy.drinker + smoker.status + days.
    anxious + diag.depression + daily.sleep.hrs +
                                    sex*daily.sleep.hrs + sex*days.anxious + age.grp + race.eth, brfss.
                                        survey.design, family = "quasibinomial")

# Calculate OR and 95% CI for estimate
print("Calculating␣OR␣&␣95%␣CI...")
results$stroke.qb.ORCI <- exp(cbind(OR = coef(results$stroke.qbinom),
  confint(results$stroke.qbinom)))
results$stroke.b.ORCI <- exp(cbind(OR = coef(results$stroke.binom),
  confint(results$stroke.binom)))
results$stroke.nandy.ORCI <- exp(cbind(OR = coef(results$stroke.nandySuggestions),
                                    confint(results$stroke.nandySuggestions)))
```

```
# Write report to file in ./Output/Reports/logit.html
OR.vector <- exp(results$stroke.qbinom$coef)
CI.vector <- exp(confint(results$stroke.qbinom))
p.values <- summary(results$stroke.qbinom)$coefficients[, 4]
stargazer(results$stroke.qbinom, coef = list(OR.vector), ci = T,
  ci.custom = list(CI.vector), p = list(p.values), single.row = T,
  type = "html", title = "Logistic Regression Model Estimating Effects of Risk Factors &
      Demographic Variables on Stroke Outcome Odds",
  out = "./Output/Reports/logit.html", omit = c("Constant"))

# Tables ----

# Summary statistics
print("Calculating mean values for sleep, days anxious and total strokes.")
results$mean.sleep.by.outcome <- svyby(~daily.sleep.hrs, ~diag.stroke,
  brfss.survey.design, svyciprop, vartype = "ci", method = "mean",
  na.rm = T)
results$mean.days.anxious.by.outcome <- svyby(~days.anxious,
  ~diag.stroke, brfss.survey.design, svyciprop, vartype = "ci",
  method = "mean", na.rm = T)
results$total.strokes <- svytotal(~diag.stroke, design = brfss.survey.design,
  na.rm = T)

# Two-way tables & Chi Sq. Tests----
print("Creating two-way tables...")
results$stroke.sex <- na.omit(svytable(~diag.stroke + sex, design = brfss.survey.design))
colnames(results$stroke.sex) <- c("Male", "Female")
rownames(results$stroke.sex) <- c("No stroke diagnosis", "Stroke diagnosis")
results$stroke.sex.p <- prop.table(results$stroke.sex, 2)
results$chsq.by.sex <- svychisq(~diag.stroke + sex, design = brfss.survey.design)

results$stroke.heavy.drinker <- na.omit(svytable(~diag.stroke +
  alc.heavy.drinker, design = brfss.survey.design))
colnames(results$stroke.heavy.drinker) <- c("Non-heavy Drinker",
  "Heavy Drinker")
rownames(results$stroke.heavy.drinker) <- c("No stroke diagnosis",
  "Stroke diagnosis")
results$stroke.heavy.drinker.p <- prop.table(results$stroke.heavy.drinker,
  2)

results$stroke.depression <- na.omit(svytable(~diag.stroke +
  diag.depression, design = brfss.survey.design))
colnames(results$stroke.depression) <- c("No Depression Diagnosis",
  "Depression Diagnosis")
rownames(results$stroke.depression) <- c("No Stroke Diagnosis",
  "Stroke Diagnosis")
results$stroke.depression.p <- prop.table(results$stroke.depression,
  1)
results$chsq.by.depression <- svychisq(~diag.stroke + diag.depression,
  design = brfss.survey.design)

results$stroke.smoker <- na.omit(svytable(~diag.stroke + smoker.status,
  design = brfss.survey.design))
```

```
results$stroke.smoker.p <- prop.table(results$stroke.smoker,
  1)
results$chsq.by.smoker <- svychisq(~diag.stroke + smoker.status,
  design = brfss.survey.design, statistic = "Chisq")

# Plots ----
plot_odds(results$stroke.qbinom)

print("Creating␣boxplots...")
results$box.sleep <- svyboxplot(daily.sleep.hrs ~ diag.stroke,
  design = brfss.survey.design, col = brewer.pal(2, "Set3"),
  all.outliers = F, xlab = "Outcome", ylab = "Hours␣of␣Sleep",
  main = "Hours␣of␣Sleep␣and␣Stroke␣Outcome")

results$box.anxious <- svyboxplot(days.anxious ~ diag.stroke,
  design = brfss.survey.design, col = brewer.pal(2, "Set3"),
  all.outliers = F, xlab = "Outcome", ylab = "Days␣of␣Anxiety",
  main = "Days␣Affected␣by␣Anxiety␣and␣Stroke␣Outcome")

# Effects
print("Creating␣effects␣plots...")
results$effects.age.income.sex <- Effect(focal.predictors = c("income",
  "age.grp", "sex"), mod = results$stroke.qbinom)

results$effects.age.sex <- Effect(focal.predictors = c("age.grp",
  "sex", "diag.depression"), mod = results$stroke.qbinom)

results$effects.anxiety.depression <- Effect(focal.predictors = c("days.anxious",
  "sex", "diag.depression"), mod = results$stroke.qbinom)

results$effects.sleep.depression <- Effect(focal.predictors = c("daily.sleep.hrs",
                                                                "smoker.status", "diag.depression"
                                                                    ), mod = results$stroke.
                                                                    qbinom)

results$effects.anxiety.depression.smoking <- Effect(focal.predictors = c("days.anxious",
  "smoker.status", "diag.depression"), mod = results$stroke.qbinom)

# Age Group & Stroke Probability Graph
print("Creating␣graphs...")
results$agegrp.odds.bar <- ggplot(ggpredict(results$stroke.qbinom,
  "age.grp"), aes(factor(x, labels = c("18␣to␣24", "**␣25␣to␣34",
  "**␣35␣to␣44", "**␣45␣to␣54", "**␣55␣to␣64", "**␣65+")),
  predicted)) + geom_point() + geom_errorbar(aes(min = conf.low,
  max = conf.high)) + coord_flip() + labs(x = "Age␣Group",
  y = "Stroke␣Probability", caption = "**␣-␣Significant␣at␣=0.05") +
  theme_Publication() + scale_fill_Publication() + scale_colour_Publication()

# Depression diagnosis status Probability Graph
results$depression.odds.bar <- ggplot(ggpredict(results$stroke.qbinom,
  "diag.depression"), aes(factor(x, labels = c("No␣Depression␣Diagnosis",
  "Depression␣Diagnosis")), predicted)) + geom_point() + geom_errorbar(aes(min = conf.low,
  max = conf.high)) + coord_flip() + ggtitle("Depression␣&␣Stroke␣Probability␣Estimate") +
```

```r
  labs(x = "Depression Diagnosis", y = "Stroke Probability",
    caption = "** - Significant at =0.05") + theme_Publication() +
  scale_fill_Publication() + scale_colour_Publication()

# Income Probability Graph
results$income.prob.graph <- ggplot(ggpredict(results$stroke.qbinom,
  terms = c("income", "diag.depression")), aes(factor(x, labels = c("Income < $10,000",
  "Income < $15,000", "Income < $20,000", "Income < $25,000",
  "Income < $35,000", "Income < $50,000", "Income < $75,000",
  "Income > $75,000")), predicted, colour = group)) + geom_point() +
  geom_errorbar(aes(min = conf.low, max = conf.high)) + coord_flip() +
  ggtitle("Depression & Stroke Probability Estimate") + labs(x = "Depression Diagnosis",
  y = "Stroke Probability", caption = "** - Significant at =0.05") +
  theme_Publication() + facet_wrap(~group) + scale_fill_Publication() +
  scale_colour_Publication()

# Education level probability graph
results$sleep.pred <- ggpredict(results$stroke.qbinom, terms = c("daily.sleep.hrs",
  "diag.depression"))
results$sleep.pred$group <- car::recode(results$sleep.pred$group,
  "'No'='No Depression Diagnosis'; 'Yes'='Depression Diagnosis';")

results$sleep.hrs.scatter <- ggplot(results$sleep.pred, aes(x,
  predicted, colour = group)) + geom_point(show.legend = F) +
  geom_line(show.legend = F) + geom_ribbon(aes(ymin = conf.low,
  ymax = conf.high, linetype = NA), alpha = 0.15, show.legend = F) +
  facet_wrap(~group) + ggtitle("Average Hours of Sleep & Stroke Probability Estimate") +
  labs(x = "Hours of Sleep", y = "Stroke Probability", caption = "") +
  theme_Publication() + scale_colour_Publication()

# Days anxious probability graph
results$days.anxious.pred <- ggpredict(results$stroke.qbinom,
  terms = c("days.anxious", "diag.depression"))
results$days.anxious.pred$group <- car::recode(results$days.anxious.pred$group,
  "'No'='No Depression Diagnosis'; 'Yes'='Depression Diagnosis';")

results$days.anxious.scatter <- ggplot(results$days.anxious.pred,
  aes(x, predicted, colour = group)) + geom_point(show.legend = F) +
  geom_line(show.legend = F) + geom_ribbon(aes(ymin = conf.low,
  ymax = conf.high, linetype = NA), alpha = 0.15, show.legend = F) +
  facet_wrap(~group) + ggtitle("Days Anxious & Stroke Probability Estimate") +
  labs(x = "Days Anxious in Past Month", y = "Stroke Probability",
    caption = "") + theme_Publication() + scale_colour_Publication()

# Income group bargraph
results$total.income.groups <- svytotal(~income, design = brfss.survey.design,
  na.rm = T)
results$total.income.groups <- melt(results$total.income.groups)
colnames(results$total.income.groups) <- c("Total", "SE")
results$total.income.groups <- setDT(results$total.income.groups,
  keep.rownames = TRUE)[]

results$bargraph_income <- ggplot(results$total.income.groups,
  aes(rn, Total)) + geom_bar(stat = "identity") + theme_Publication() +
```

```
  scale_colour_Publication() + scale_y_continuous(label = comma) +
  scale_x_discrete(labels = c("<$10,000", "$10,000␣-␣$15,000",
    "$15,000␣-␣$20,000", "$20,000␣-␣$25,000", "$25,000␣-␣$35,000",
    "$35,000␣-␣$50,000", "$50,000␣-␣$75,000", ">$75,000")) +
  ggtitle("2016␣United␣States␣Household␣Income␣Ranges", subtitle = "Weighted␣BRFSS␣Data") +
  labs(x = "Income␣Range")

# Age group bargraph
results$total.age.groups <- svytotal(~age.grp, design = brfss.survey.design,
  na.rm = T)

results$total.age.groups <- melt(results$total.age.groups)
colnames(results$total.age.groups) <- c("Total", "SE")
results$total.age.groups <- setDT(results$total.age.groups, keep.rownames = TRUE)[]

results$bargraph_age.grp <- ggplot(results$total.age.groups,
  aes(rn, Total)) + geom_bar(stat = "identity") + theme_Publication() +
  scale_colour_Publication() + scale_y_continuous(label = comma) +
  scale_x_discrete(labels = c("18␣to␣24␣Years␣Old", "25␣to␣34␣Years␣Old",
    "35␣to␣44␣Years␣Old", "45␣to␣54␣Years␣Old", "55␣to␣64␣Years␣Old",
    "65+␣Years␣Old")) + ggtitle("2016␣United␣States␣Age␣Groups",
  subtitle = "Weighted␣BRFSS␣Data") + labs(x = "Age␣Group")



# Mosaic Plots
print("Creating␣mosaic␣plot...")
mosaic(results$stroke.smoker, shade = T, main = "Proportions␣of␣Stroke␣Diagnoses␣by␣Smoker␣
    Status",
  labeling_args = list(set_varnames = c(diag.stroke = "Stroke␣Outcome",
    smoker.status = "Smoker␣Status")), spacing = spacing_equal(unit(1, "lines")))

# Maps -------------------
print("Creating␣map...")
results$total.strokes.by.state <- svyby(~diag.stroke, ~STATE,
  brfss.survey.design, svytotal, na.rm = T)

results$total.strokes.by.state$STATE <- tolower(results$total.strokes.by.state$STATE)
us_states$STATE <- as.factor(tolower(us_states$NAME))
census$STATE <- tolower(census$STATE)
us_states$id <- rownames(us_states@data)
us_states@data <- join(as.data.frame(us_states), census, "STATE")
us_states@data <- join(as.data.frame(us_states), results$total.strokes.by.state,
  "STATE")
us_states$id <- rownames(us_states@data)
us_states$stroke.prevalence <- (us_states$diag.strokeStroke/us_states$Pop.2016) *
  1e+05
stroke.prevalence <- na.omit(us_states[, c("STATE", "stroke.prevalence")]@data)
stroke.prevalence$quintile <- with(stroke.prevalence, cut(stroke.prevalence,
  breaks = quantile(stroke.prevalence, probs = seq(0, 1, by = 0.2),
    na.rm = TRUE), include.lowest = TRUE, dig.lab = 5))
us_states.gg <- fortify(us_states)
us_states.gg <- join(us_states.gg, us_states@data, by = "id")
```

```r
results$colorPalette <- brewer.pal(5, name = "YlGn")
results$map.stroke.prevalence <- ggplot(stroke.prevalence, aes(map_id = STATE)) +
  geom_map(aes(fill = quintile), map = fifty_states, show.legend = T,
    colour = "grey25", size = 0.25) + expand_limits(x = fifty_states$long,
  y = fifty_states$lat) + coord_map() + scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) + labs(x = "", y = "") +
  theme(legend.position = "bottom", panel.background = element_blank()) +
  theme_Publication() + fifty_states_inset_boxes() + ggtitle("2016 United States Stroke 
      Prevalence",
  subtitle = "Weighted BRFSS Data") + scale_fill_manual(values = results$colorPalette,
  name = "Strokes / 100,000", guide = guide_legend(keyheight = unit(4,
    units = "mm"), keywidth = unit(4, units = "mm"), title.position = "top",
    reverse = F), labels = c("1512.6 - 2057.2    ", "2057.2 - 2221.6    ",
    "2221.6 - 2528.8    ", "2528.8 - 2987.0    ", "2987.0 - 3766.6    "))

# Save the results object ----
print("Saving all results...")
saveRDS(results, file = "Data/results.rds")

# Save graphs to ./Output/Graphs
ggsave("age_grp_prob.png", results$agegrp.odds.bar, device = "png",
  path = "./Output/Graphs/", scale = 1, dpi = 300, width = 10,
  height = 5, units = "in")

ggsave("sleep_depression_prob.png", results$sleep.hrs.scatter,
  device = "png", path = "./Output/Graphs/", scale = 1, dpi = 300,
  width = 12, height = 8, units = "in")

ggsave("days_anxious_depression_prob.png", results$days.anxious.scatter,
  device = "png", path = "./Output/Graphs/", scale = 1, dpi = 300,
  width = 12, height = 8, units = "in")

ggsave("stroke_prev_map.png", results$map.stroke.prevalence,
  device = "png", path = "./Output/Maps/", scale = 1, dpi = 300,
  width = 12, height = 8, units = "in")

ggsave("income_bargraph.png", results$bargraph_income, device = "png",
  path = "./Output/Graphs/", scale = 1, dpi = 300, width = 12,
  height = 8, units = "in")

ggsave("age_bargraph.png", results$bargraph_age.grp, device = "png",
  path = "./Output/Graphs/", scale = 1, dpi = 300, width = 12,
  height = 8, units = "in")
```