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

Natalie's text here.

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

Variable Name	Variable Type
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

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

[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%

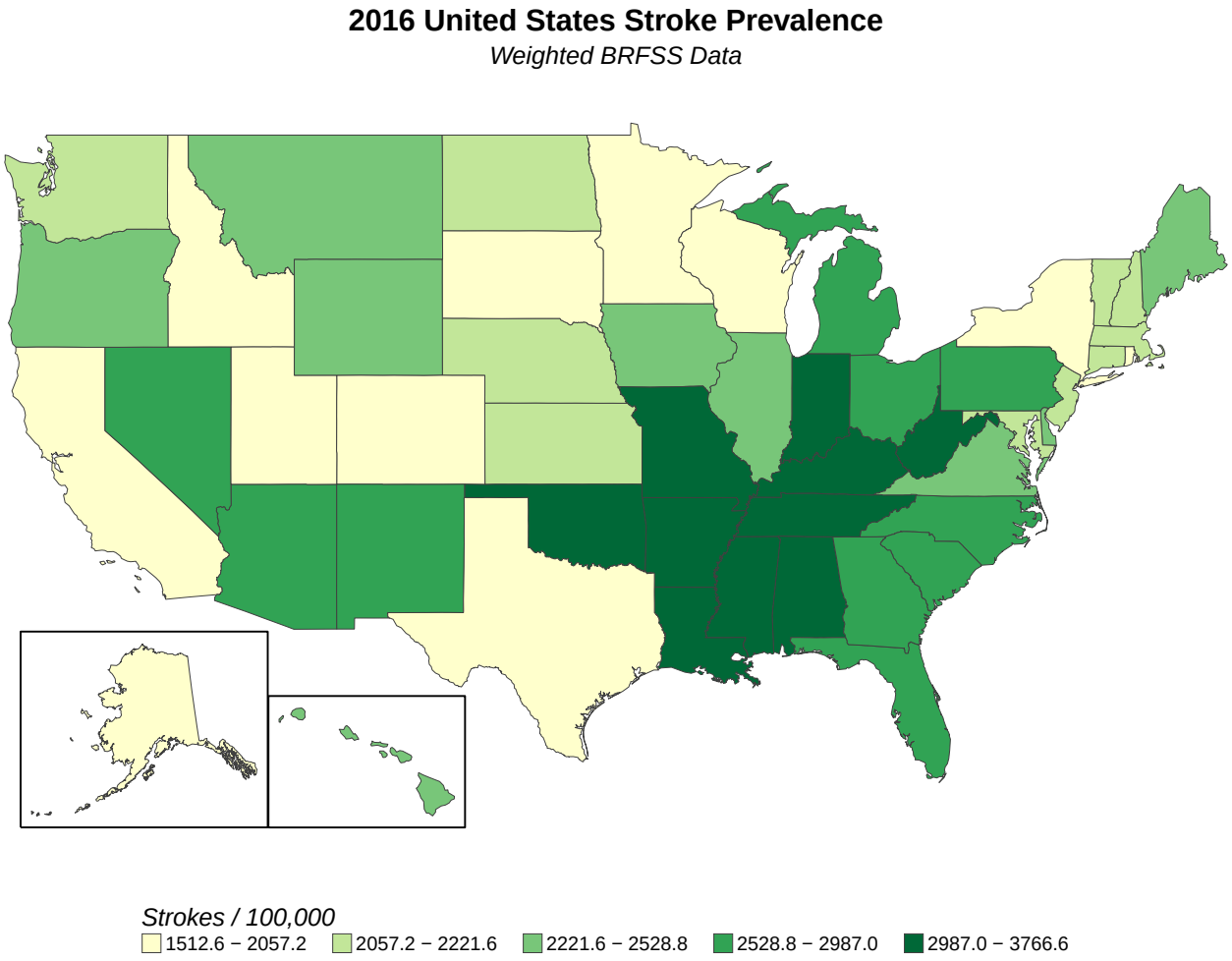

Figure 1: 2016 United States Stroke Prevalence

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.

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

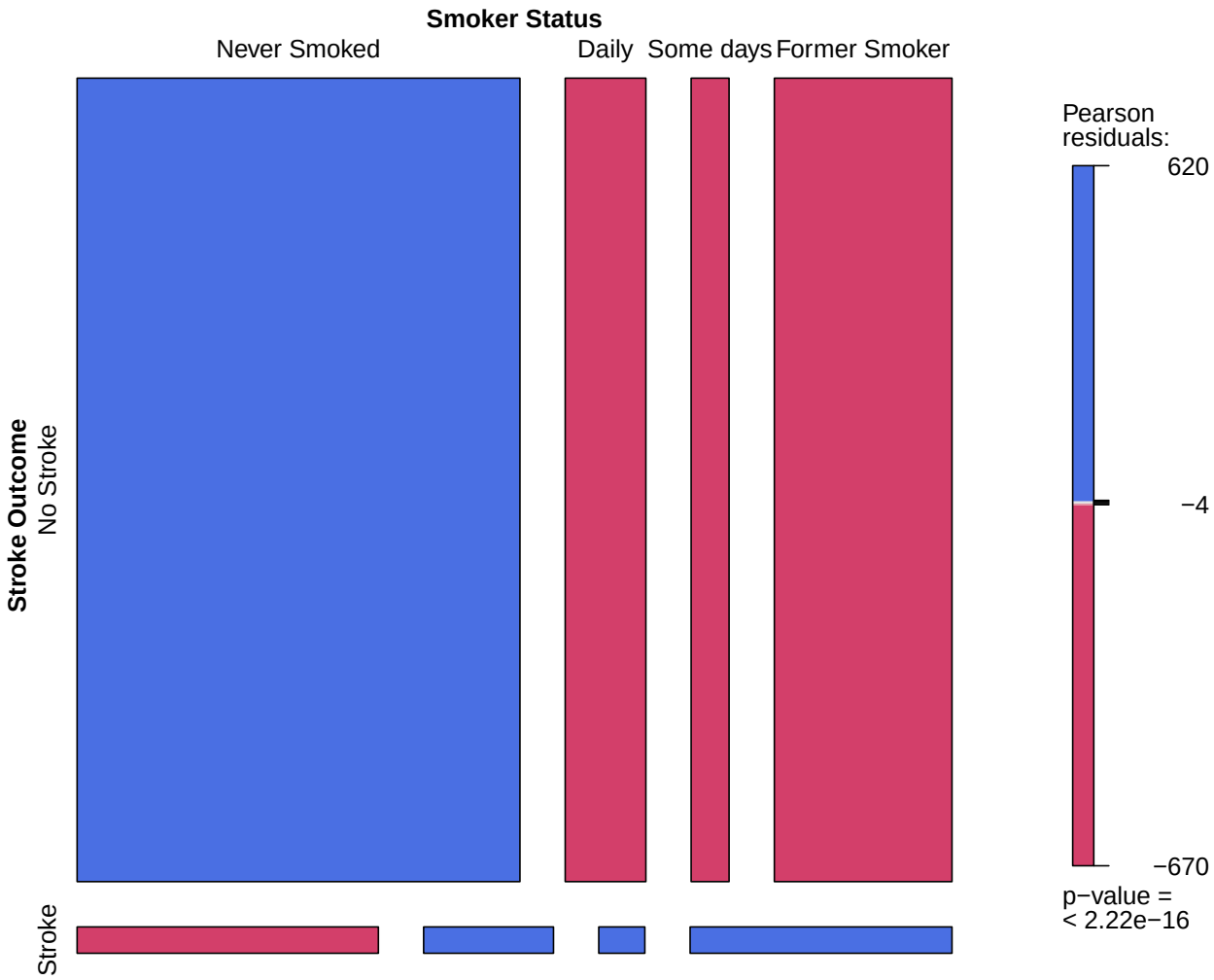

Figure 2: Proportion of Stroke Outcomes and Smoker Status

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.

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
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.

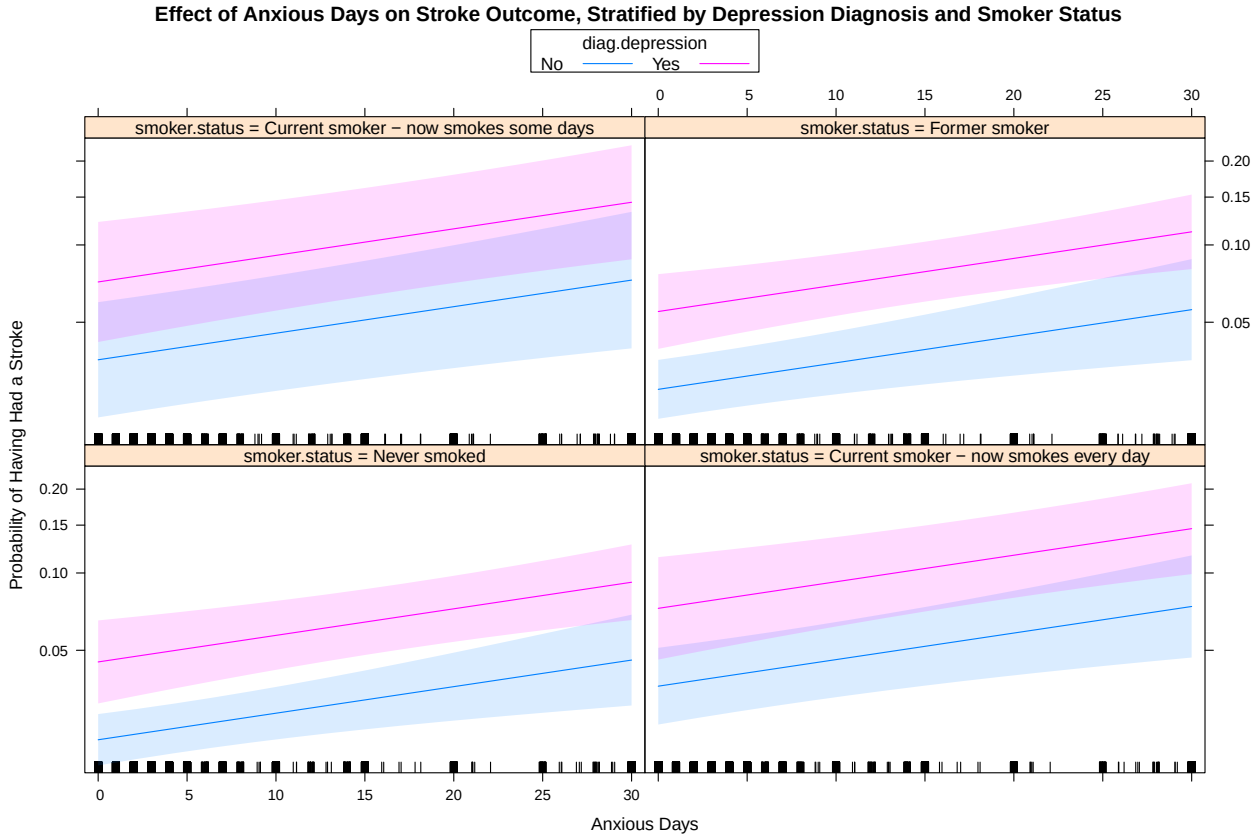

Figure 3: Effects of Anxious Days on Stroke Outcome, Stratified by Depression Diagnosis and Smoker Status

Conclusion

Beth's part here

References

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

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

Code