Number of Sexual Partners and Depression

Abby Mutic

abby.mutic@emory.edu

Introduction/Background

As a PhD student actively involved in the Center for Children's Health, the Environment, the Microbiome and Metabolomics and a Certified Nurse Midwife practicing in the Grady Teen Clinic, I am interested in studies involving children or adolescents, their behaviors, and how their behaviors may influence health outcomes. I am particularly interested in mental health outcomes and better understanding specific areas to target for effective clinical intervention.

Research Question

For this project, I aim to answer the research question: Is the number of sexual partners experienced in adolescence ($\leq 12^{th}$ grade) associated with depression and/or anxiety in adulthood? I hypothesize that the more sexual partners an individual has at an early age, the greater amount of depressive or anxiety cases they will experience in their adulthood. As a part of this project, I learned how to handle longitudinal data with multiple time points, merge datasets, create new variables, and use multiple software packages in R during analysis.

Methodology/Approach

Dataset

The data used in this project is pulled from The National Longitudinal Study of Adolescent to Adult Health (Add Health) public database. It is a longitudinal study of United States adolescents in grades 7-12 during the 1994-95 school year. The Add Health cohort has been followed at various time points with the most recent being wave IV in 2008. Wave IV includes only those who were retained in the study from the original data collection period in 1994-1995.

Since the Add Health database is open to the public, the file was easily downloaded and added to my personal repository on GitHub. The repository was then pulled into a new project in R studio in preparation for data wrangling and analysis. The Add Health database file can be found at https://dataverse.unc.edu/dataset.xhtml?persistentId=doi:10.15139/S3/11900.

Analysis Plan

The statistical analysis will be performed using R. Implementation of basic statistical procedures in this platform was central to my training. Before visualizing the data, I organized the data by creating a data.frame where the rows represent each individual case and the columns represent the variables of interest. Next, I renamed the given variable names and choose names that will make coding and interpretation easier. This was done by running the attributes of the data.frame and then recoded as desired. The variable I will used for age was attempted to be recoded to display an age number rather than the year of birth as it is currently given in the dataset but the code did not work. The dichotomous dependent variables- depression and anxiety- were attempted to be combined as total$both showing a person has both mental health disorders. However, due to my limited R skills, I ended up using the variable depression as my only outcome used in the modeling.

After my variables of interest were renamed, I looked at a subset of the data and printed the subset by using knitr::kable(). Next, I looked at a summary of the numeric statistics and frequencies of the factor variables. To view summary statistics of factor variables, I converted the variables to numeric form as needed.

The analysis plan summarized and describing the sample using descriptive statistics in order to visualize the data in a meaningful way. The demographics and descriptive data were placed in Table 1. Next, inferential statistics were conducted to test the hypotheses. The continuous

independent variable- number of sexual partners- will be compared to the dependent dichotomous variables- depression and anxiety- by using logistic regression. Logistic regression was used to predict the characteristics associated with the diagnosis of depression. Odds ratio(s) were reported and used for interpretation. An odds ratio >1 indicates as the number of sexual partners increases, the odds of depression also increases. An odds ratio <1 indicates as the number of sexual partners increase, the odds of depression decreases. Results of the logistic regression are displayed in Table 2. Careful consideration of confounding variables were included in model building such as gender, race, and anxiety diagnosis. Other confounders identified a priori and included in initial exploratory analysis were personal and/or family history of mental illness, personal and parental socioeconomic status, social support, parental support, education, and sexual abuse. Missing data was addressed locally by telling R to remove missing data prior to performing the functions.

Data Summary/Results

Table 1. Characteristics of The National Longitudinal Study of Adolescent to Adult Health cohort

	Wave I 1994-95 (n = 6504)		**Wave IV 2008 (n =5114)**	
Characteristics	**n**	**(%)**	**n**	**(%)**
Age (mean, range) (m=3)	(16, 13-21)		(29, 25-34)	
Gender (m=1)				
Male	3147	(48.4)	NA	NA
Female	3356	(51.6)	NA	NA
Race (check all that apply)				
Hispanic or Latino	743	(11.4)	NA	NA
White	4294	(66.0)	NA	NA
Black or African American	1619	(24.9)	NA	NA
American Indian or Native American	236	(3.6)	NA	NA
Asian or Pacific Islander	270	(4.2)	NA	NA
Other	425	(6.5)	NA	NA
Felt depressed in past week (m=20)				
Most/all the time	193	(3.0)	108	(2.1)
A lot of the time	444	(6.8)	211	(4.1)
Sometimes	1853	(28.5)	1178	(23.0)

Never/rarely	3994	(61.4)	3616	(70.7)
Depression diagnosis (m=1)				
Ever	NA	NA	827	(16.2)
Never	NA	NA	4286	(83.8)
Anxiety diagnosis				
Ever	NA	NA	639	(12.5)
Never	NA	NA	4474	(87.5)
Ever had sexual intercourse (m=86)				
Yes	2565	39.4	NA	NA
No	3853	59.2	NA	NA
Sexual partners (mean, range) (m=4874)	(6.4, 1-500)		NA	

Table 2: Characteristics associated with Depression Diagnosis

Characteristic	**Odds Ratio**	**CI**	**p-value**
Female Gender	1.33	(0.97, 1.70)	1.27e-12
Feeling depressed in past week			
Most or all the time	2.37	(1.55, 3.21)	1.66e-08
A lot of the time	1.97	(1.32, 2.63)	2.87e-09
Sometimes	1.25	(0.88, 1.61)	3.05e-11

Figure1. Scatterplot of Interested Variables from Add Health Study

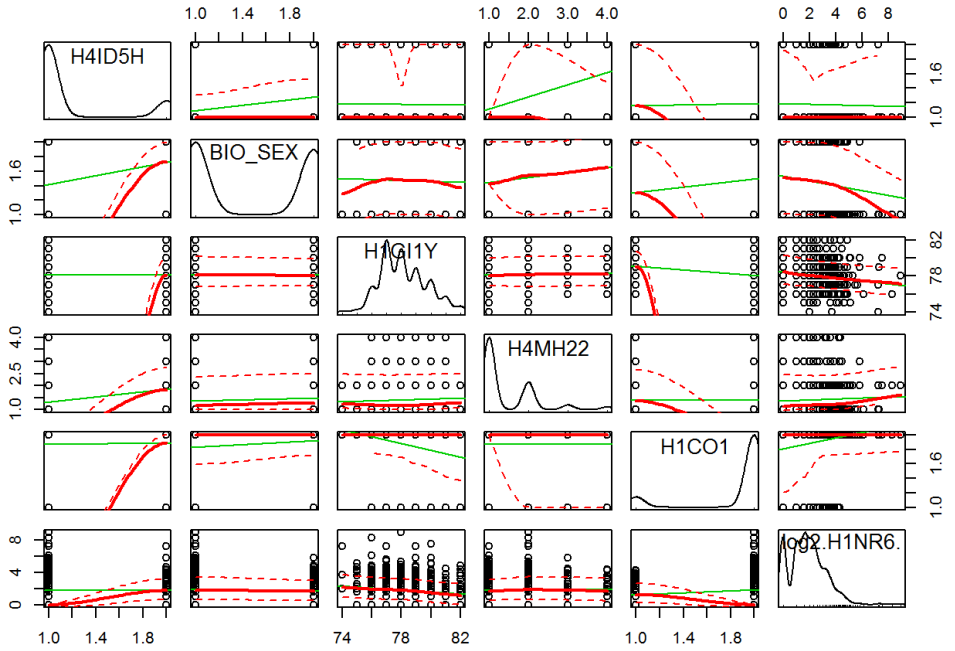


Figure 2. Distribution of Number of Sex Partners in Adolescents from Add Health Study

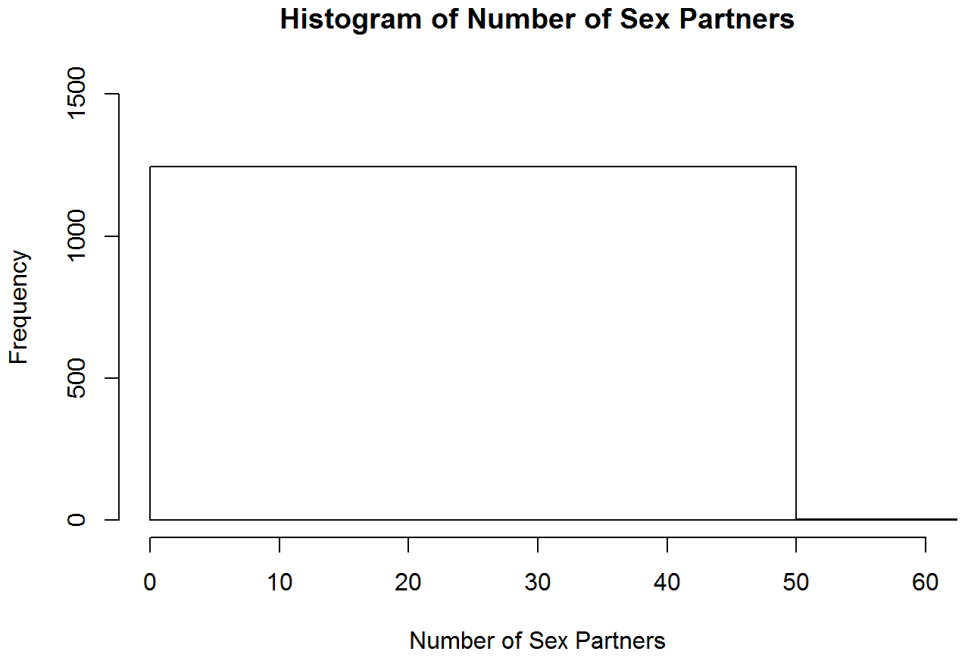
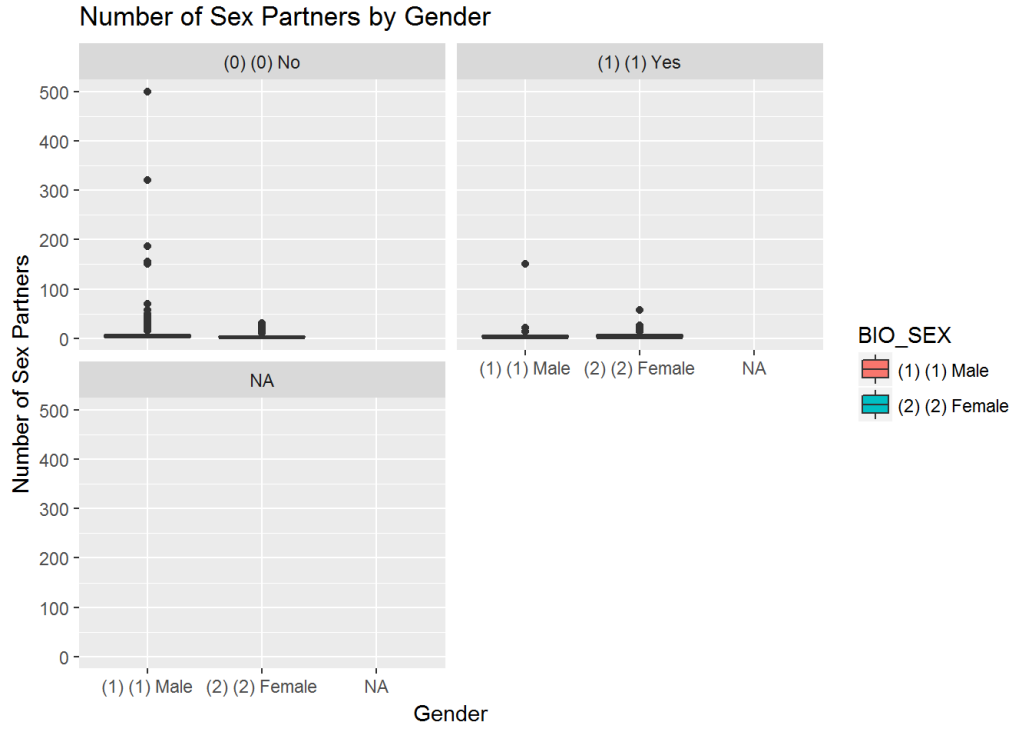


Figure 3. Distribution of Number of Sex Partners by Gender in Adolescents from Add Health Study



Interpretation, Limitations, and Conclusions

Interestingly females are 1.3x more likely to be diagnosed with depression than males. Additionally, for every unit change in being depressed in the past week, the log odds of depression increases by 2.4 (most or all the time), 2.0 (a lot of time), 1.2 (sometimes). The total number of sexual partners in adolescence doesn't seem to be associated with diagnosis of depression in adulthood.

The bar graph and histogram produced did not show the distributions as I would have expected and were ultimately not helpful in visualizing the data.

Limitations

The main limitation in this project was my limited R coding knowledge. There were many procedures I wanted to complete during the analysis but was not able to master the coding in order to produce it. Another limitation was the Add Health survey did not include data that was most likely confounders in my analysis such as history of depression or anxiety, social support, mental health counseling, or stress indices. It would have also been helpful and more comprehensive to have all variables repeated in every wave.

Code, Associated Data and Files, and Github Repository

My repository can be found at the following link.
https://github.com/amutic/AddHealth The repository includes a record of the raw data files from Add Health and a copy of the codebook is provided by the Add Health research study funded by a collaboration between the University of North Carolina at Chapel Hill and Duke University. My rmd file detailing my analysis thought and code and can be found at https://github.com/amutic/AddHealth/blob/master/Add%20health%20final%20project_Mutic.Rmd