Ivy Nguyen
ANTH3454

Associations Between Major Depressive and Substance Abuse Disorder

I. Research Problem.

In this research analysis, I will be investigating the connections between illegal substance use (marijuana, cocaine/crack, hallucinogens, heroin, inhalants, pain relievers, tranquilizers, stimulants, and sedatives) and major depressive disorder. Within the past decade, the public has become much more aware about mental illness. Two of the most common disorders found among individuals are depression and substance abuse. The extent to how correlated these two disorders are interests me, as from experience, those who have one illness are easily able to develop another as a consequence of their current disorder. My goal in performing this investigation is to deduce whether those who have depression are more likely to also have a drug abuse disorder. By knowing the correlation between these two disorders, mental health departments may be more prepared to treat and prevent individuals at risk for developing each corresponding disorder. For example, if someone has depression, they can monitor the patient to ensure they do not start abusing drugs as a method of self-treatment. To explore this correlation, I am going to use survey data on citizens' drug use and mental health within the year of 2020. Based on the data, I can investigate how often citizens have both substance abuse disorder and depression.

II. Description of the Database

The database I have chosen to perform research upon is from the Substance Abuse and Mental Health Data Archive, a collection of data sources found by the US Department of Health and Human Services. Specifically, I am using self-reported survey data conducted in 2020 called the National Survey on Drug Use and Health given out to 32893 individuals over the age of 12 in America. The goal of the dataset is to give statistical information

on the use of illicit drugs along with mental health issues in the US. In terms of strength, the database has a big sample size, inclusive of all populations within America. It is also recent, and thus can be generalized to the current US population. Weaknesses with the database include the fact that the data was collected via a self-reported survey–individuals might not be honest with the researchers about their personal details. Furthermore, for my analysis, I am using imputed values from the researchers for measuring whether an individual has a substance use or major depressive disorder within the past year. Thus, my analysis is dependent on how accurate their imputations are based on the participants' survey data.

In terms of my research problem, the data is incredibly relevant to the question: it contains data on both substance abuse and mental illness within the US. From here, I can analyze the data to see how correlated substance abuse is with experiencing depression within the past year. From this data, I can answer the question of how connected having substance use disorder is to having depression. I can also answer which demographics are most vulnerable to having substance abuse, depression, or both. I cannot answer the question of whether substance abuse results in depression, or depression results in substance abuse directly–I can only see how these disorders are correlated, especially since the timeline between disorders is fully presented.

III. Research Design

Within the database, there were originally 2890 variables included, each being a participant's answer to a survey question (yes/no). I condensed it down to only 5 variables that specifically deal with illicit drug disorder, depressive feelings, and basic demographics about education level and age category. Since I intend to employ statistical procedures on categorical variables to determine their association with each other, this file design is appropriate for my use. In terms of cases, I ran several procedures using different sets of cases. For the

Chi-Square tests, since the number of cases exceeds 1000, I factored the cases by age and education level. For residual analysis, however, I decided to use all cases present.

My hypothesis is that there will be an association between substance abuse and depression. That is, someone who has depression within the past year is more likely to have a substance abuse disorder as well. On the other hand, someone who has depression but has taken medication for it within the past year will be less likely to have a substance abuse disorder.

To test my hypotheses, I will be using Crosstabs analysis. My database includes only categorical variables, and so Crosstabs would be the best procedure to use to analyze it. Specifically, I will be analyzing the patterns of residuals between my depression and substance abuse variables after looking at a descriptive table about the data. Since the database has so many cases, even when factored by demographic variables, evaluating significance is not as useful, however I do perform Chi-Square tests as well. I will also be using Cramer's V as a measure of strength in association between variables, alongside residuals for pattern analysis.

IV. Statistical Description

To describe the variables, I made a descriptive table using Crosstabs and also graphed bar charts for univariate summaries. I chose these procedures because categorical variables have no inherent order and moving around values does not change the information given. Therefore, using a box plot does not make sense in this case. Descriptives such as mean and standard deviation would not really give us any useful information either, since there is really only the frequency of categorical variables given. Thus it is more beneficial to include a descriptive table of the count of each case, along with the column percentage.

Illegal Substance Use Disorder vs Major Depressive Disorder

Observing the descriptive tables, there are a lot of individuals who do not have an illegal substance use disorder and major depressive disorder (95.9%). There are, however, a lot of people who have major depressive disorder without an illegal substance use disorder (88.2%). We can see that most individuals do not have an illegal substance use disorder. Of those that do, there are 11.8% of individuals that have both in comparison to those that have an illegal substance use disorder but not major depressive disorder (4.1%). The general trend is that most individuals in the sample do not have illegal substance use disorder; more individuals have both disorders than those that only have illegal substance use disorder. In comparison to the bar charts, this impression is not as obvious. In fact, we observe the opposite within the bar chart–there are more people who have an illegal substance use disorder and do not have major depressive disorder. This is precisely because the bar chart is only graphing the count of individuals within each category; the descriptive table gives percentages of how many individuals have substance use disorder in the major depressive disorder variable. Indeed, there are more individuals that have an illegal substance use disorder and do not have major depressive disorder in count.

Illegal Substance Use Disorder vs Treating Depressive Feelings w/ Medication

Observing the descriptive tables, there are a lot of individuals who do not have an illegal substance use disorder and did not use medication to treat their depression (91.8%). There are also a lot of people who have used medication to treat depression without an illegal substance use disorder (87.5%). We can see that most individuals do not have an illegal substance use disorder. Of those that do, there are only 8.2% that have both an illegal substance use disorder and have used medication for depression in comparison to those that only have an illegal substance use disorder (12.5%). The general trend is that most individuals in the sample do not have illegal substance use disorder, and those that do have not used medication to treat depression within the past year. The bar chart shows the same information, although the counts between those that have illegal substance use disorder without using medication and those that do is so close that it is hard from just looking at the graph. It is

evident that most individuals do not have an illegal substance use disorder nor have they used medication to treat depression.

Comparing these two factors of depression, it seems as though there are more individuals that have an illegal substance use disorder and major depressive disorder, however less individuals with illegal substance use disorder that use medication to treat their depressive feelings. It would be curious to investigate whether this trend is due to self-medication–in which individuals who have major depressive disorder choose to use substances instead of a prescription to treat their illness.

V. Hypothesis Testing

There are generally two hypothesis tests to be done: one to determine the association between major depressive and substance use disorder, another to determine the association between using medication for depression and substance use disorder. Both of these tests will also be factored by age and education level to see whether there are any differences in these groups. Since the dataset consists of categorical variables, it is best to use Crosstabs analysis. In all tests, alpha is equal to 5% and the critical oft value is 3.841 with a degree of freedom of 1.

Illegal Substance Use Disorder vs Major Depressive Disorder

Hypothesis Test 1_A

Null hypothesis: Major depressive and substance abuse disorder within the past year are not related, and there is no association between these variables.

Alternative hypothesis: Major depressive and substance abuse disorder within the past year are related, and there is an association between these variables.

The observed p value is less than 0.001 and the Pearson Chi-Square is equal to 321.924. Since X^2's magnitude is more than 3.841, we have criteria to reject the null hypothesis and accept the alternative hypothesis. Since p is also less than 0.05, we have enough evidence to reject the null hypothesis. However, despite how there is evidence to reject the null hypothesis, the sample size when performing the X^2 test is above 2000, and so its significance is not that meaningful. We must observe strength measures instead.

Looking at symmetric measures, we see that Cramer's V has a value of 0.111, a value close to 0. Looking at directional measures, we see that Lambda has a value of 0. These values are close to 0, indicating a weak relationship between major depressive and substance abuse disorder.

When observing the residuals, there are several big outliers we must take note of. There is a standardized residual of -3.8 for those that have major depressive disorder and do not have an illegal substance use disorder, meaning there are less individuals that have major depressive disorder but do not have illegal substance use disorder. There are less individuals that do not have major depressive disorder but have an illegal substance use disorder (-5.8) by a larger extent. Considerably, the biggest standardized residual observed is for those that have both–there are much more individuals observed than expected for this case (16.5). Looking at these residual outliers, we can see that there are much more individuals with both major depressive disorder and illegal substance use disorder than the theoretical expectation, suggesting that a major portion of the population suffers from both disorders.

Hypothesis Test 1_B

Null hypothesis: Major depressive disorder and substance abuse disorder within the past year are not related, and there is no association between these variables when factored by age.

Alternative hypothesis: Major depressive disorder and substance abuse disorder within the past year are related, and there is an association between these variables when factored by age.

The observed p value is less than 0.001 and the Pearson Chi-Square is equal to 67.675 for 18-25 year olds and 206.748 for 26 and older individuals. Since X^2's magnitude is more than 3.841 in both cases, we have criteria to reject the null hypothesis and accept the alternative hypothesis. Since p is also less than 0.05 for both cases, we have enough evidence to reject the null hypothesis. Despite how there is evidence to reject the null hypothesis, the sample size when performing the X^2 test is above 2000, and so its significance is not that meaningful. We must observe strength measures instead.

Looking at symmetric measures, we see that Cramer's V has a value of 0.094 for 18-25 year olds and 0.106 for 26 and older individuals, both values close to 0. Looking at directional measures, we see that Lambda has a value of 0 for both cases as well. These values are close to 0, indicating a weak relationship between major depressive and substance abuse disorder when factored by age.

When observing the residuals, there are several big outliers we must take note of, but only for those that do have an illegal substance use disorder. There is a standardized residual of -3.3 for those that do not have major depressive disorder but do have an illegal substance use disorder from 18-25 years old. Considerably, this portion of the sample has a bigger standardized residual for 26 or older year olds with a residual of -4.1. There are much less individuals within these age groups that do not have major depressive disorder but have an illegal substance use disorder. When looking at those with both disorders, the standardized residuals for 18-25 year olds is 7.2 and 13.5 for 26 year olds and up. There are much more older individuals with both disorders than expected, suggesting that a major portion of the older population suffers from both.

Hypothesis Test 1_C

Null hypothesis: Major depressive disorder and substance abuse disorder within the past year are not related, and there is no association between these variables when factored by education level.

Alternative hypothesis: Major depressive disorder and substance abuse disorder within the past year are related, and there is an association between these variables when factored by education level.

The observed p value is less than 0.001 and the Pearson Chi-Square is equal to 39.801 for less than high school, 56.685 for high school grads, 102.451 for some college/associate degree, and 115.408 for college graduates. Since X^2's magnitude is more than 3.841 in all cases, we have criteria to reject the null hypothesis and accept the alternative hypothesis. Since p is also less than 0.05 for all cases, we have enough evidence to reject the null hypothesis. Despite how there is evidence to reject the null hypothesis, the sample size when performing the X^2 test is above 2000, and so its significance is not that meaningful. We must observe strength measures instead.

Looking at symmetric measures, we see that Cramer's V has a value of 0.138 for less than high school, 0.103 for high school grads, 0.113 for some college/associate degree, and 0.105 for college graduates–all values close to 0. Looking at directional measures, we see that Lambda has a value of 0 for all cases as well. These values are close to 0, indicating a weak relationship between major depressive and substance abuse disorder when factored by education levels.

When observing the residuals, there are several big outliers we must take note of. There is a standardized residual of 5.8 for those that have major depressive disorder and illegal substance use disorder with less than high school education–more undereducated individuals have both disorders. The same trend is shown for high school graduates with a standardized residual of 6.9, some college/associate degree at 9.1, and college graduate at 10.1. It seems as though education level does not play a factor in having both disorders–there are more individuals than expected at all of these education levels, however there are a considerable amount of college graduates that have

both disorders than expected. There are also less individuals than expected that do not have major depressive order but do have illegal substance use disorder for some college/associate degrees and college graduates with residuals of -3.7 and -3.3 respectively.

Illegal Substance Use Disorder vs Treating Depressive Feelings w/ Medication

Hypothesis Test 2_A

Null hypothesis: Treating depressive feelings with medication within the past year is not related to having substance abuse disorder within the past year, and there is no association between these variables.

Alternative hypothesis: Treating depressive feelings with medication within the past year is related to having substance abuse disorder within the past year, and there is an association between these variables.

The observed p value is less than 0.001 and the Pearson Chi-Square is equal to 26.693. Since X^2's magnitude is more than 3.841, we have criteria to reject the null hypothesis and accept the alternative hypothesis. Since p is also less than 0.05, we have enough evidence to reject the null hypothesis. Despite how there is evidence to reject the null hypothesis, the sample size when performing the X^2 test is above 2000, and so its significance is not that meaningful. We must observe strength measures instead.

Looking at symmetric measures, we see that Cramer's V has a value of 0.071, a value close to 0. Looking at directional measures, we see that Lambda has a value of 0.009. These values are close to 0, indicating a weak relationship between treating depressive feelings with medication and having substance abuse disorder within the past year.

When observing the residuals, there are few outliers we must take note of. There is a standardized residual of -3.1 for those who have illegal substance use disorder but did not use medication to treat depression, and 3.8 for

those that have illegal substance use disorder and used medication. This suggests that there are less people who have illegal substance use disorder and do not use medication. Yet, there are more people who have the disorder but have used medication than expected, suggesting that medication does not help the mental illness as well as it should–depressed individuals still resort to abusing drugs.

Hypothesis Test 2_B

Null hypothesis: Treating depressive feelings with medication within the past year is not related to having substance abuse disorder within the past year, and there is no association between these variables when factored by age.

Alternative hypothesis: Treating depressive feelings with medication within the past year is related to having substance abuse disorder within the past year, and there is an association between these variables when factored by age.

The observed p value is less than 0.001 and the Pearson Chi-Square is equal to 14.333 for 18-25 year olds and 16.944 for 26 and older individuals. Since X^2's magnitude is more than 3.841 in both cases, we have criteria to reject the null hypothesis and accept the alternative hypothesis. Since p is also less than 0.05 for both cases, we have enough evidence to reject the null hypothesis. Despite how there is evidence to reject the null hypothesis, the sample size when performing the X^2 test is above 2000, and so its significance is not that meaningful. We must observe strength measures instead.

Looking at symmetric measures, we see that Cramer's V has a value of 0.085 for 18-25 year olds and 0.072 for 26 and older individuals, both values close to 0. Looking at directional measures, we see that Lambda has a value of 0 for 18-25 year olds and 0.010 for 26 and older individuals. These values are close to 0, indicating a weak

relationship between treating depressive feelings with medication and having substance abuse disorder within the past year.

When observing the residuals, there is only one outlier we must take note of. For 26 year olds and older, there is a standardized residual of 3.0 for those that have substance use disorder and have used medication, meaning there are more individuals than expected in this category. As mentioned above, perhaps patients are choosing to supplement their medication with another drug.

Hypothesis Test 2_C

Null hypothesis: Treating depressive feelings with medication within the past year is not related to having substance abuse disorder within the past year, and there is no association between these variables when factored by education level.

Alternative hypothesis: Treating depressive feelings with medication within the past year is related to having substance abuse disorder within the past year, and there is an association between these variables when factored by education level.

The observed p value is 0.688 for less than high school, less than 0.0001 for high school grads, 0.046 for some college/associate degree, and less than 0.001 for college graduates. The Pearson Chi-Square is equal to 0.161 for less than high school, 13.922 for high school grads, 3.999 for some college/associate degree, and 19.318 for college graduates.

Since X^2's magnitude is more than 3.841 for high school grads, some college/associate degree, and college graduates, we have criteria to reject the null hypothesis and accept the alternative hypothesis for these cases; not for those who received less than high school education, as we must accept the null hypothesis for this factor.

Since p is also less than 0.05 for all cases except for less than high school education, we have enough evidence to reject the null hypothesis in these cases. Despite how there is evidence to reject the null hypothesis for these cases, the sample size when performing the X^2 test is above 2000, and so its significance is not that meaningful. We must observe strength measures instead.

Looking at symmetric measures, we see that Cramer's V has a value of 0.122 for high school grads, 0.046 for some college/associate degree, and 0.094 for college graduates–all values somewhat close to 0. Looking at directional measures, we see that Lambda has values of 0 for the cases discussed as well. These values are all close to 0, indicating a weak relationship between treating depressive feelings with medication and having substance abuse disorder within the past year. There are no significant standardized residuals to take note of.

VI. Summary and Conclusions

Through my analysis, I investigated the association between illegal substance abuse and depression. Specifically, I looked into how major depressive disorder correlates with illegal substance use disorder, and how illegal substance use disorder is related to using medication to treat depression. My database consisted of categorical variables, and thus I chose to mainly use Crosstabs as a method of analysis.

From my analysis, I mainly found that there are weak relationships between illegal substance use disorder and major depressive disorder. From the descriptives, there are more individuals that have an illegal substance use disorder and major depressive disorder, however less individuals with illegal substance use disorder that use medication to treat their depressive feelings. When investigating patterns, it was found that there are less individuals that have major depressive disorder but do not have illegal substance use disorder, alongside individuals that do not have major depressive disorder but have an illegal substance use disorder. This trend

suggests that more individuals are expected to have both disorders simultaneously within the past year. However, it was also seen that there are a lot more individuals with both illegal substance use disorder and major depressive disorder than expected, suggesting that more of the population is suffering from both disorders at once. Specifically, this is evident for older individuals from ages 26 and up, along with those who receive a graduate level education. I also observed a weak relationship between illegal substance use disorder and using medication to treat depression. There are less individuals who have illegal substance use disorder but did not use medication to treat depression, and more who did use medication than expected. The trend suggests that more individuals who have depression are willing to treat it with medication, however are also willing to abuse drugs as well.

I believe the quantitative techniques I employed were quite useful, specifically the Crosstabs residual analysis along with the descriptive Crosstabs table. The X^2 test was included, however due to the very large sample size, significance was a minor detail for this research problem. The strength tests did inform me of how strong each association was, and the residuals showed common patterns in the population's drug use and mental health. I am not surprised by the results of my analysis, however I did believe the strength of the relationship between all variables considered would be moderate instead of weak. This discrepancy might be due to using the imputed variables from the survey and how survey data was collected through self-reporting.

After performing the analysis, I still have many questions about the dataset. For example, I would like to further explore why these trends are coming up in the data: are individuals abusing drugs whilst using medication? Is medication not potent enough to help their depression, or do individuals choose to use illegal drugs even if the medicine works for them? There is still much more to be learned from studying these variables.

VII. Appendix

Dataset:

https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2020-nsduh-2020-ds0001

Variables:

	Name	Type	Width	Decimals	Label
1	UD5ILLNOALC	Numeric	3	0	RC-ILL DRUG USE DISORDER, NO ALC USE DISORDER - PAST YEAR
2	AMDEYR	Numeric	4	0	RC-ADULT: PAST YEAR MAJOR DEPRESSIVE EPISODE (MDE)
3	ARXMDEYR	Numeric	4	0	RC-ADULT: USED RX MEDICATION FOR DEPRESSIVE FEELINGS IN PY
4	EDUHIGHCAT	Numeric	3	0	RC-EDUCATION CATEGORIES
5	CATAG2	Numeric	3	0	RC-AGE CATEGORY RECODE (3 LEVELS)

Output:

- univariate.spv
    - Bar charts for IV
- mde.spv
    - Descriptive table, bar charts, analytical table for Major Depressive Disorder vs Illegal Substance Use Disorder
- meds.spv
    - Descriptive table, bar charts, analytical table for Illegal Substance Use Disorder vs Using Medication to Treat Depressive Feelings