Fall 12-15-2021

Smoking, Alcohol Consumption, and Depression in Association with Incidence of Type 2 Diabetes among Mexican Americans in Starr County, Texas

Gabriela Rubannelsonkumar
grubannelsonku@mail.stmarytx.edu

Smoking, Alcohol Consumption, and Depression in Association with Incidence of Type 2 Diabetes among Mexican Americans in Starr County, Texas

Gabriela Rubannelsonkumar

Mentored By: Dr. Lori Boies

Department of Biological Scienes

HO4301OL Senior Thesis

St. Mary’s University

I. ABSTRACT

Previous studies on conditions like obesity, hypertension, and type 2 diabetes mellitus (T2DM) have explored the correlations between them and various other human conditions, including aortic stiffness, left ventricular hypertrophy and sleep apnea, as they predict possibilities of developing certain diseases in Mexican Americans. This study aims to observe the correlation between lifestyle decisions that could relate to the onset of the depression in normal, prediabetic, and diabetic individuals. These include smoking habits and alcohol consumption. Many papers have previously conducted research on these lifestyle habits as they relate to obesity, hypertension, diabetes, however, have done so in a singular analysis approach. For example, they only focused on alcohol consumption or smoking, whereas this study takes a more holistic approach that combines all the variables. Understanding the relationship that these conditions have with each other, through use of a case-control study for individual exposures and logistic regression methods for multiple exposures, heightens the chances of catching the development of these more serious long-term diseases.

With smoking, ethanol, and mental health scales, we visualized their relation to diagnosis of T2DM. Data was collected on the Mexican American population in South Texas because unusually higher records of these major disease categories but without a solid explanation of what factors contribute the most to this increase. We found that smoking and alcohol consumption could not be considered significant predictors of depression in diabetic individuals, however a positive association between heavy physical activity and depression while holding bodyfat, smoking, alcohol, and weight constant.

II. INTRODUCTION

An important approach in research when studying diseases and their onset is to investigate different variables that may be the cause of the development of a certain disease, or variables serve as biological confounders. These variables may be other health conditions in the individuals as well as lifestyle habits that affect human health. A widely known example of this is the correlation between smoking cigarettes and developing lung cancer. The number one factor for developing lung cancer is smoking cigarettes, constituting almost 80-90% of deaths to due to lung cancer ("What are the risk…"). Studies have shown that even inhalation of secondhand smoke increases the risk ("What are the risk…"). This is a simple example of how studying possible correlations could help in spotting disease risk. The purpose of this study is to identify if smoking and alcohol consumption are risk factors for developing depression in normal, prediabetic, and diabetic patients.

From extensive research on the negative effects that smoking has on the body, it has been found that this lifestyle habit acts as risk factor of incident type 2 diabetes (Chang). Smoking is a preventable leading cause of disease and takes responsibility for a large portion of American deaths every year ("Cigarette Smoking…."). It increases the chances of getting several conditions of the lungs and the heart, as well as cancer ("Cigarette Smoking…"). This behavior, particularly, is strange to observe, because it isn't beneficial in any way to the body, and is often described as unpleasant in taste, yet people continue to do it. It appears that the effects of nicotine, which influences the release of neurotransmitters regulating mood, has such a strong grip on the smoking population, where they choose to overlook the lethal effects (Jiloha). In addition to the effects of nicotine on the body, the withdrawal that results from trying to quit are difficult to get over. These symptoms include having cravings to smoke, getting easily irritated, feeling restless, trouble sleeping, having a difficult focusing on a task, gaining weight because of increased hunger, and feeling depressed ("7 Common").

Cigarette smoking has also previously been identified as a risk factor of type 2 diabetes. The specific mechanism behind smoking exposure increasing the risk of diabetes has not been made clear, however, there is plenty of evidence to show that smoking increases insulin resistance (Chang). To briefly describe the general mechanism behind this claim, smoking inarguably causes harm in many ways to the body by interfering with cell's normal function. Increased inflammation can disrupt the effectiveness of the natural levels of insulin produced in the body, which already increases the risk of developing the disease ("Cigarette Smoking"). In addition to reducing the effectiveness of the insulin already produced in the body, smoking also decreases the efficacy of the treatment for type 2 diabetes. Patients who don't have history of smoking are treated for the condition with extra dosage of insulin, because the body is not able to produce it, which has shown to effectively maintain the balance. Smoking alters that balance to where the normal course of treatment of the disease is not as effective as in non-smoking patients. Overall, studying smoking as a modifiable risk factor for type 2 diabetes could be an effective approach to decreasing the prevalence of type 2 diabetes, or T2DM.

Alcohol consumption is another addicting behavior that has been found to be associated with developing type 2 diabetes (Zeratsky). Like smoking cigarettes, the negative effects of alcohol have been proven, especially long-term. Early exposure to alcohol increases the likelihood of alcohol abuse later in life, which increases the risk for several other health-related conditions as people age. One of these commonly is liver disease, as the liver is the main site for ethanol metabolism. It has been found that 35% of the population of problem drinkers, or people who experience negative effects from drinking but isn't considered to be dependent on alcohol, develop advanced liver disease due to modifications in the disease progression (Osna et al.). Studies have also shown how alcohol consumption affects the neuroendocrine system and bone. Further research could help in understanding whether severe chronic disorders are a result of excessive drinking as an adolescent or the accumulation of drinking over the years ("The effects of alcohol…"). Although drinking does directly damage some organs more than others in the short-term, the habit of excess drinking over a long period of time eventually reaches all organs in the body and more negative effects are observed all around ("The effects of alcohol…"). This possible cascade effect of alcohol on the body could be a reason to consider alcohol consumption as a risk factor for developing depression in individuals with and without type 2 diabetes, which is what this study seeks to find.

Alcohol as it relates specifically to type 2 diabetes has some interesting findings. It is believed that alcohol consumption in moderate amounts may actually decrease the risk for developing T2DM, however, the opposite is true for consuming excess amounts in the long term (Zeratsky). As stated previously, alcohol intake may not have a direct effect on the organs involved in the mechanism behind diabetes, mainly the pancreas, but eventually its effects reach all organs in the body. Chronic inflammation of the pancreas because of alcohol abuse could weaken its function of secreting insulin, thereby developing diabetes (Zeratsky). Various studies have concluded that moderate alcohol consumption, which is about one drink a day for women and two drinks a day for men, is associated with reduced incidence of type 2 diabetes in the long term (Carlsson et al.). This however isn't definitive for the world population in general, various other factors including weight, age, and gender, play a significant role in whether alcohol consumption will be beneficial or detrimental. For example, high alcohol consumption was found to increase type 2 diabetes incidence in women with lower BMIs, but not in overweight men and women (Carlsson et al.). In addition to studying the association of alcohol consumption to developing diabetes directly, it may be interesting to study if these drinking habits have an association to developing depression with diagnosis of type 2 diabetes in the Starr County population.

The last non-traditional risk factor of T2DM that will be analyzed is patient mental health status, specifically depression status. In the world, an estimated 5% of adults suffer from depression ("Depression."), and about 6.3% are affected by type 2 diabetes (Khan et al.), and the goal of the study is to determine if there is an overlap between these two populations. Depression is a condition that affects daily functioning of individuals, where they don't have the energy or the motivation to want to accomplish the tasks they need. It is more than just mood swings and sudden negative emotional responses. At the worst, depression can lead to suicide. The claim that depression is associated with many conditions is supported by the idea that the stress and struggle that comes with managing life with treatment for the condition can lead to symptoms of depression. Depression itself could lead to poor lifestyle decisions, including smoking and alcohol consumption, which added on top of the challenges of an illness, only results in more stress for the individual ("Diabetes and Depression:…"). Managing both conditions together take a toll on the individual and their community. So, studying the extent to which the two conditions are associated with each other, through case-control studies and multiple logistic regression analyses, will allow for treatments to target the secondary disease to potentially reduce the overall drain on the individual.

Studying the relationship of alcohol and smoking exposure on depression in populations of patients with diabetes, pre-diabetes, and normal patients requires the use of appropriate statistical analyses tools to gather significant insights. The study design typically to test for associative characteristics between a disease outcome and risk factors is a case-control study. In this type of study, specific exposures are tested against the cases and controls. The cases are usually the group of individuals that have the disease, in this case depression, while the controls are the group without the disease. Within those groups, some subjects may have had the exposure in question or not (eg. smoking, alcohol etc.). With this case, the frequency of subjects in the cases that had the exposure can be compared to the frequency of subjects in the controls that had the exposure to test association.

III. LITERATURE REVIEW

This research proposal utilizes data obtained from a collaboration study with Dr. Craig Hannis at UT Health Houston School of Public Health, titled *Beyond type 2 diabetes, obesity, and hypertension: an axis including sleep apnea, left ventricular hypertrophy, endothelial dysfunction, and aortic stiffness among Mexican Americans in Starr County, Texas*. This paper took evidence that suggested that less traditional risk factors shouldn't be ignored when looking at common conditions like diabetes, obesity, hypertension, and cardiovascular disease and performed a study to test those hypothesized correlations. Such factors include cardiac structure, aortic stiffness, impaired endothelial function, and obstructive sleep apnea. It has been previously observed that Mexican Americans carry a larger weight of major diseases than other populations, but the purpose of the study was to get a better understanding of the distribution of those untraditional risk factors in individuals with major diseases. The study was conducted by initially gathering surveys to identify individuals in the Starr County population that have been diagnosed with any of the major diseases from 2002 to 2006. Then, these same individuals and a group with type 2 diabetes were re-examined in 2010 until 2014, with the additional assessment of the less traditional risk factors. Data from the two examinations were analyzed for any noteworthy associations. They identified that from the 1200 individuals in the sample 885 have one or more of these conditions, and 50% have three or more. Looking at the group of individuals with type 2 diabetes, 74% have one or more of the non-traditional risk factors, aortic stiffness, LV mass, impaired reactive hyperemia, and sleep apnea. These high percentages indicate that even if not causative, there is still an indirect relationship. Expanding upon this study, this paper focuses on lifestyle risk factors as they relate to individuals with prediabetes, to potentially identify the increased risk of developing type 3 diabetes (Hanis et al.).

A similar study, Shi et al., focused on the association of physical activity, smoking, and alcohol with the incidence of type 2 diabetes, also referred to as T2DM, in middle-aged and elderly men of Chinese descent. With the prevalence of T2DM increasing in China, studying the contribution of lifestyle habits that could be modified is essential to prevent it from becoming more widespread than it already is. The purpose of *Shi et al*. aligns with that of this study, and it was informative to gather information on the methodology used to test the correlations as they may be utilized. In the study, type 2 diabetes was similarly identified in patients by surveys, and the lifestyle risk factors were interpreted using Cox proportional hazard analyses. The Cox proportional-hazards model is a regression model that investigates the association between patient survival time and predictor variables, which in this case, are the lifestyle risk factors. Since these risk factors do affect the body physically, it was hypothesized that there will be some trend between them and incidence of type 2 diabetes (Shi et al.). What they found was that moderate alcohol consumption and physical activity are both inversely proportional to T2DM risk, and smoking, on the other hand, is directly proportional to T2DM risk in middle-aged and elderly Chinese men (Shi et al.). The results they gathered allow healthcare professionals to target lifestyle habits of patients that could potentially change the outcome of developing T2DM, which is essentially is also the goal of this study. These results helped shape a hypothesis for this study, because similar analyses between risk factors and T2DM incidence are being done, but on a different ethnic population.

The non-traditional risk factors that are being studied in this paper include smoking, alcohol intake, and mental health. The previous study focused on the correlation between diabetes and smoking and alcohol. *Johnson et al*. looks at the relationship between prevalent mental disorders, such as depression and anxiety, and incidence of diabetes in individuals in the Kerala Diabetes Prevention Program (K-DPP). Individuals at high risk for type 2 diabetes in Kerala, India who are considered to have depression and anxiety were examined to see the relationship between mental health symptoms and incident Type 2 Diabetes over a period of two years. As mental health does not have a large direct effect on biological processes like smoking and alcohol consumption, it can be hypothesized that study of this correlation will not yield significant associations. *Johnson et al.* gathered depression and anxiety prevalence information from the 9-item Patient Health Questionnaire (9-PHQ) and the Generalized Anxiety Disorder 7-item scale. This study proposes a way for this study to go about analyzing depression prevalence in populations at high risk for T2DM, because the depression data was also measured using the 9-PHQ. The questionnaire was given to the population of 1007 high-risk individuals, and of those 7.5% were found to have depression and 5.5% anxiety. Although this prevalence of depression and anxiety seem to be higher for the general population of India, the small group of individuals in the sample with observed mental health symptoms didn’t indicate a distinct association between that and the development of T2DM (Johnson et al.). These results give insight for generating a hypothesis for this study, which is that depression status will not show significant association in relation to developing T2DM.

IV. METHODS

A. DATA DICTIONARY

The data dictionary, PMSummForBoiesData, describes information on all the data that is included in this study, all of which is from the baseline examinations. The study design from which the data is from recruited 300 individuals from Starr County, Texas, that were prediabetic and 300 individuals with normal glycemia. All subjects were examined at baseline.

The column that allows for separating different glycemia statuses is the E0DIAB attribute, which specifies whether the individual is normal (no diabetes), prediabetic, or diabetic (based on fasting glucose, 2 hours glucose and/or HbA1c or use of glucose lowering medications).

The subset of data that give general information on subjects included education and occupation information. Specifically, years of education (EDUCATION), income at the 3-month exam (INCOME), insurance status at the 3 month exam (INSURANCE), employment status at the 3-month exam (EMPLOYED), and marriage/relationship status (MARRIED). Other data points in the dataset are subject biological data like glycemia-related measures, anthropometrics, lipids, blood pressure and pulse. These are not the target variables that will be focused on; however, they give insight to individual characteristics that may potentially play a role in the status of their mental health, smoking and drinking status.

The attributes included in the dataset that will be focused on are mental health, specifically depression status as measured on a depression scale at the Baseline Exam (E0PHQSC1, E0PHQSC2, E0PHQMAJ, E0PHQOTH, and E0PHQDIF), smoking and ethanol, and physical activity. For smoking and ethanol, there is data on baseline exam smoking status (E0SMOKE), smoking exposure (E0PACKYRS and E02HNDSMK), and grams of ethanol consumed per week (E0METHWK). The column E0SMOKE specifies smoking status with values of 1 for current smoker, 2 for former smoker, and 3 for never smoker. The column E0METHWK specifies the grams of ethanol consumed per week calculated and this was stratified as moderate and high based upon CDC guidelines (CDC). For physical activity, data included is baseline exam hours per day with no activity such as sleeping (E0NOACT), hours per day in heavy activity that largely increase heart rate and breathing (E0HVYACT), hours per day in moderate activity that cause an increase in heart rate and breathing like walking (E0MODACT), hours per day in slight activity that don't increase heart rate or breathing (E0SLTACT), hours per day of sedentary activity like watching TV (E0SEDACT), total hours of activity to check if the previous variables add up to 24 hours per subject (E0TOTHRS), and hours watching TV or sitting at the computer per day (E0TVDAY).

B. DATA ANALYSIS

The first steps to begin this project was performing an exploratory data analysis for understanding the data. Both univariate and bivariate analysis was done to achieve an all-around idea of what the data looks like and what attributes correlate with each other. For univariate analysis, histograms and boxplots were created to understand the distribution of each attribute of numeric type, which allows observation of outliers as well. Using R packages, *ggplot2* (Wickham, 2016) and base R functions, histograms, boxplots, and bar plots were constructed to get a visual idea of the subjects' lifestyle. Careful consideration of outliers, if they are accurate or error, was important for analysis to avoid gathering inaccurate insights. In the dataset utilized, outliers are kept, in order to keep the data as true to the sample as possible. For example, outliers were identified in the E0GMETHWK (ethanol intake in grams per week) column that indicated individuals in the sample consuming large amounts of alcohol that seem too high to be accurate.

These values are kept in the study to ensure that all the data is considered in analysis. For missing values, either imputation of the data point will be done after analysis of the entire set, or the missing values will be excluded from the statistical tests.

After understanding the distribution of the subjects, with information regarding their age, education, income and other lifestyle aspects, a case-control study to obtain odds ratio values was run for alcohol and smoking on depression in individuals with diabetes, prediabetes, and normal glycemia. This method was used initially to observe the individual exposures (smoking and alcohol intake) on depression. The data is first subset into groups of individuals: Normal Glycemia (normalSubset), Prediabetes (prediabSubset), and Diabetes (diabSubset). Based on values from the column E0DIAB (glycemia status at baseline) in the master dataset (616 observations), the data was subset to obtain separate data frames for each. In the E0DIAB column, a value of 1 indicates an individual with normal glycemia, 2 indicates prediabetes, 3 is diabetes based on value of fasting glucose, 2-hour glucose and/or HbA1c, and 4 is diabetes based on use of glucose lowering medications. The subset of normal glycemia contains 152 observations, while the diabetes subset contains 96 observations, and the prediabetes contains 368. This imbalance in data split between the three subsets was taken note of to be the potential cause of some insignificant odds ratio values.

For a case-control study, the attributes must be binary, where there are two values (eg. 0 or 1, “yes” or “no”). As described in the data dictionary, the values of E0SMOKE are 1, 2, and 3. So, for the purpose of the study the individuals assigned 1 and 2, which indicate current smoker and former smoker, respectively, were grouped together as exposure to smoking, and individuals assigned 3, which is never smoked, as no exposure. Similarly, E0METHWK was also simplified to be binary. Alcohol consumption categorization as moderate and heavy vary based on gender, which was accounted for in this study. Moderate drinking is categorized as 2 or less drinks a day for men and 1 drink or less for women (“Alcohol Questions…”). Excessive drinking is normally defined as more than 15 drinks per week for men, and more than 8 drinks per week for women (“Alcohol Questions…”). The ethanol consumption is provided as grams of ethanol consumed. For the study, the ethanol was grouped into no drinking (0 grams for both male and female), moderate drinking (1-14 grams per week for males and 1-7 grams per week for females), and heavy drinking (>210 grams per week for males and >98 grams per week for females).

The odds ratio algorithm, using the *epi.2by2* function from the *epiR* package (Stevenson et al., 2021) in R (R Core Team, 2021) was applied to each subset, for one exposure of smoking and another of alcohol consumption. In this way, the study will output odds ratio values to help determine if individual exposure to either smoking or alcohol is associated with the outcome of depression status in diabetic, prediabetic, and normal individuals.

This was followed by a multiple logistic regression analysis to see if multiple exposures or confounding variables affect the outcome of depression. This tests for how smoking and alcohol affects the odds of depression in diabetic, prediabetic, and normal individuals. In addition to these exposures, confounding variables are also tested for association to depression status. The logistic regression model equation to test the multiple exposures of smoking and alcohol with confounders of age and exercise is as follows:

$$Logit\ diabetes\ and\ depression\ =\ \beta_0 + \beta_1(smoke) + \beta_2(alcohol) + \beta_3(age) + \beta_1(exercise)$$

Similarly, logistic regression for other confounders were also done. Such confounders include weight and BMI. These tests were done using the base R multiple logistic regression command, *glm* (R Core Team, 2021), and converted into adjusted odds ratios using the logistic.display from the epiDisplay library.

V. RESULTS

A. EXPLORATORY DATA ANALYSIS

An initial exploratory data analysis (EDA) was performed to help understand the data before making assumptions from the statistical tests.

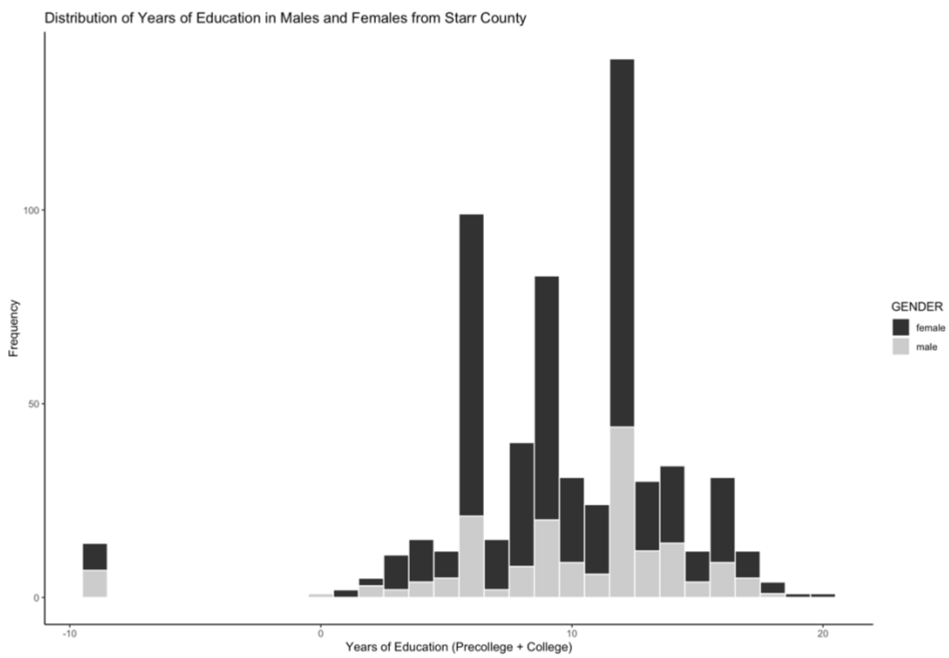

Figure 1. Comparison of Years of Education (Precollege + College) between Males and Females in Starr County

From Figure 1, initially, we can see that, of the sample studied, there is a significantly larger number of female subjects than male. Most subjects have received about 12 years of education exactly, which is consistent with what is known of the socioeconomic status of Starr County residents. Individuals may not have the same education access in the Rio Grande Valley of South Texas, which leads to individuals being more likely to not have more than 12 years of education. The outlier of -9 indicates that there are some missing values found in the EDUCATION attribute, as seen in the graph. To connect to the purpose of the study, the years of education was observed between each diabetic status (Normal Glycemia, Prediabetes, and Diabetes), seen in Figure 2 below.

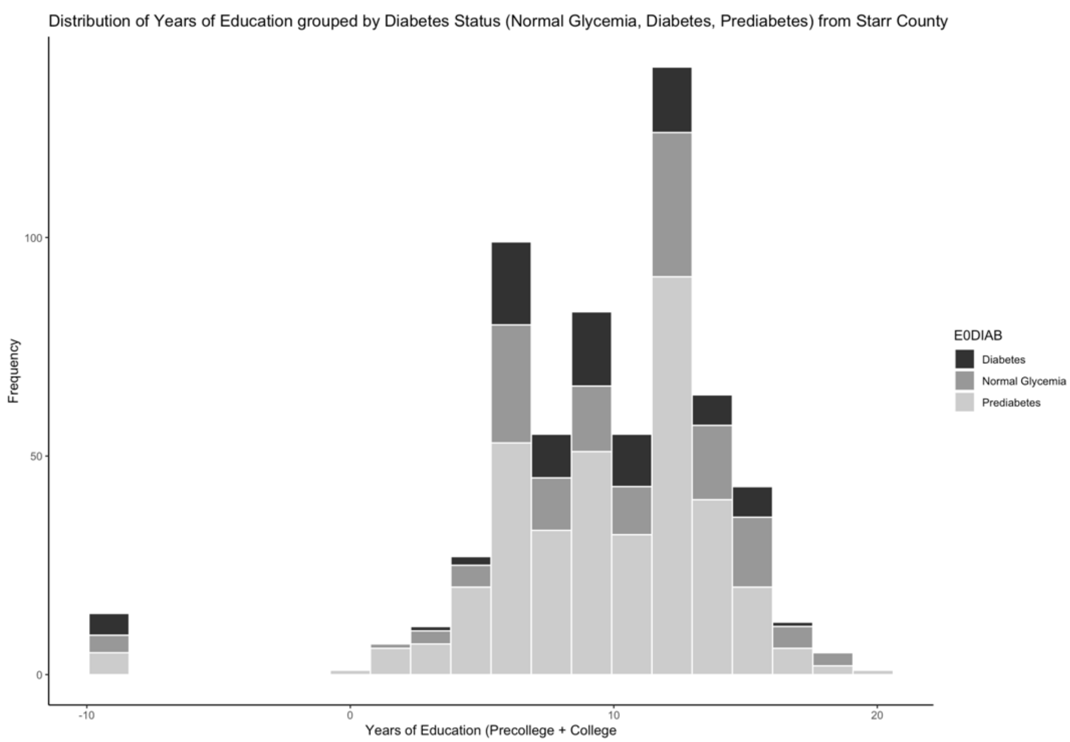

Figure 2. Comparison of Years of Education (Precollege + College) between Normal Glycemic, Prediabetic, and Diabetic Individuals in Starr County

The smallest group observed is the diabetic individuals, followed by normal, then prediabetic. Among the largest group of prediabetic individuals, most have had 12 years of education, which is also seen in normal glycemic individuals, but not seen across to diabetic individuals. Most of the diabetic individuals have had less than 12 years of education.

To continue with the preliminary data visualization, in relation to the cases of the case-control study, the distribution of both depression statuses (E0PHQSC1 and E0PHQSC2) among males and females were observed. Figure 3 and Figure 4 below show the spread of self-reported depression status between each of the diabetic status.

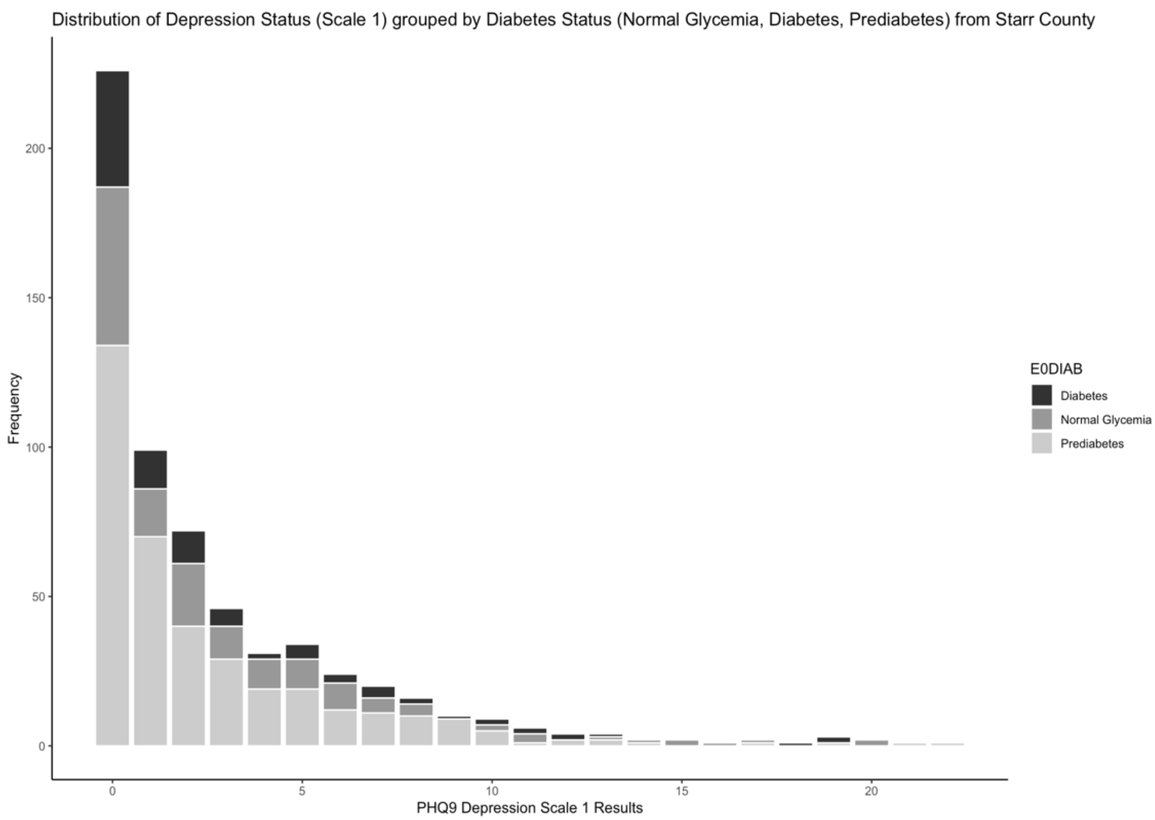

Figure 3. Comparison of the Distribution of Depression Status (Scale 1) between Normal Glycemic, Prediabetic, and Diabetic Individuals in Starr County

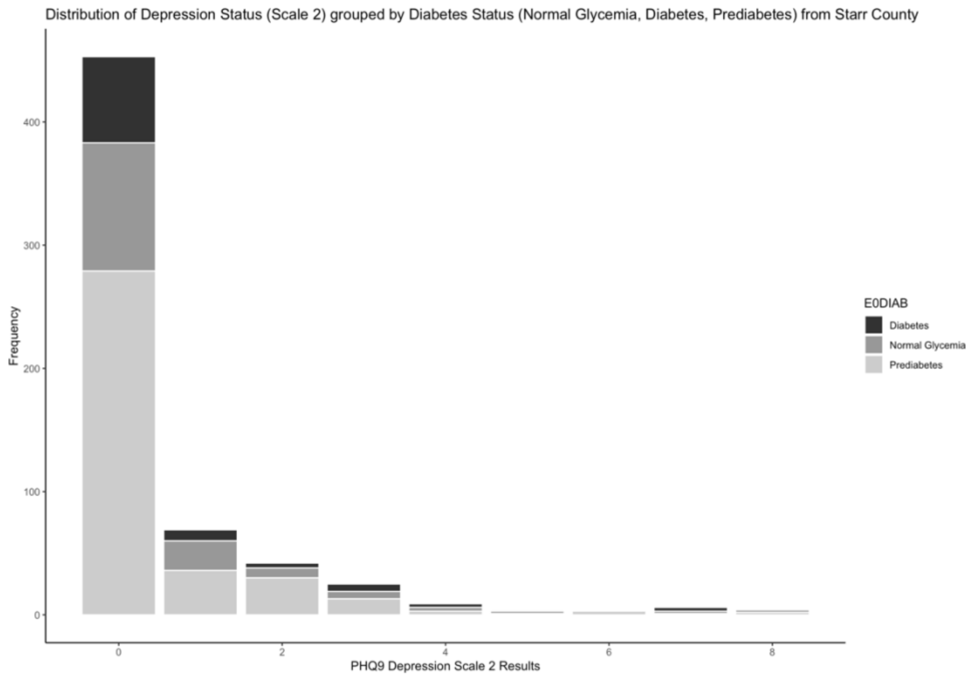

Figure 4. Comparison of the Distribution of Depression Status (Scale 2) between Normal Glycemic, Prediabetic, and Diabetic Individuals in Starr County

These histograms show a skew in the self-reported depression statuses overall. Looking at the different diabetes statuses, the same general trend is seen between all three. Most subjects, within each group, reported that they are showing no depression or minimal to mild depression.

Now, the exposures used in the case-control studies were smoking and alcohol, so bar plots to see the distribution of smokers versus non-smokers in normal, prediabetic, and diabetic subjects, were generated to make initial hypotheses for the smoking exposure. A side-by-side comparison between the groups are shown below, in Figures 5, 6, 7.

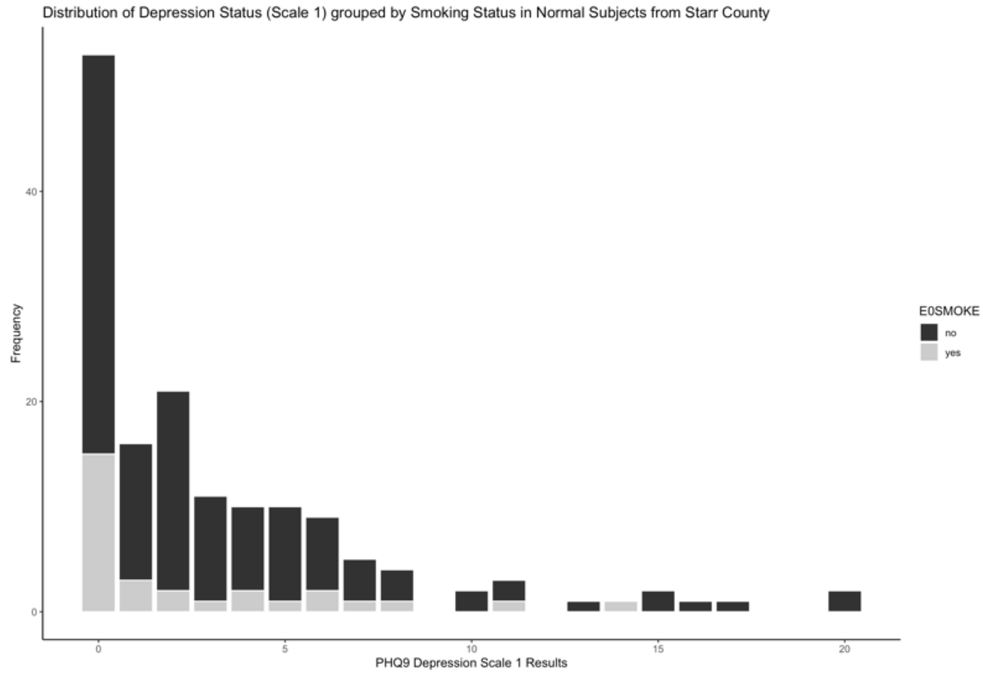

Figure 5. Comparison of the Distribution of Depression Status (Scale 2) grouped by Smoking Status in Normal Glycemic Subjects in Starr County

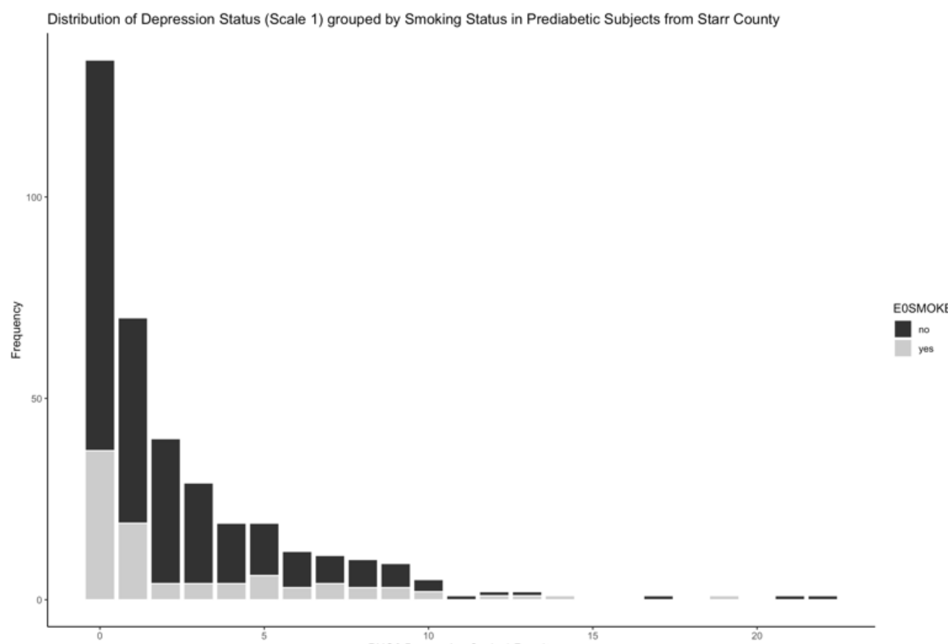

Figure 6. Comparison of the Distribution of Depression Status (Scale 2) grouped by Smoking Status in Prediabetic Subjects in Starr County

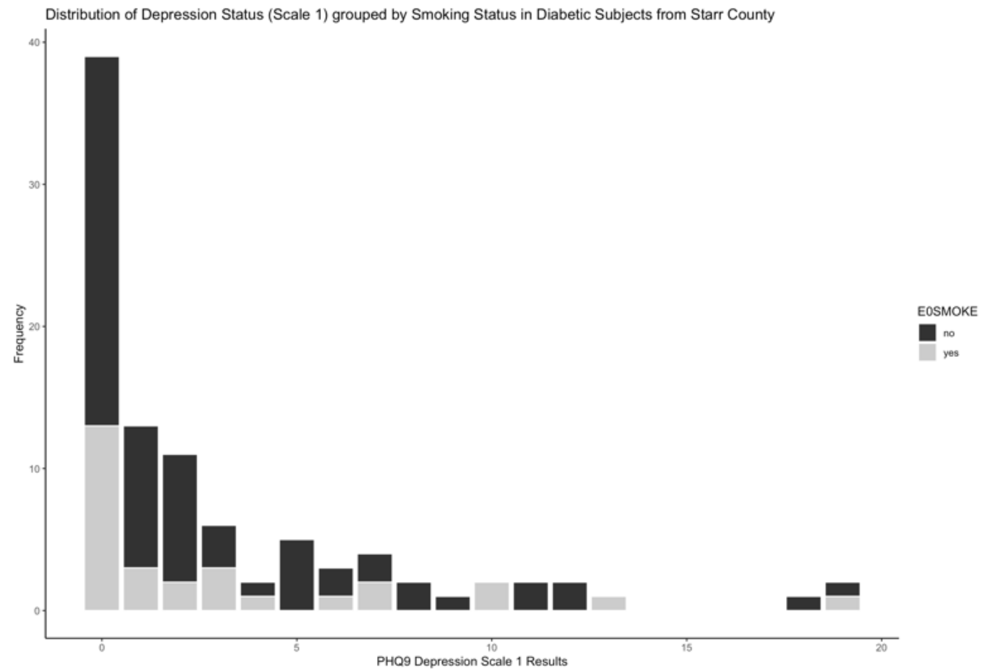

Figure 7. Comparison of the Distribution of Depression Status (Scale 2) grouped by Smoking Status in Diabetic Subjects in Starr County

It can be observed that the smoking status in relation to depression status of individuals has a similar trend across the diabetes groups. There are few observed current and former smokers in the master dataset when compared to those who have never smoked. Based on the figures, there is no trend that distinguishes normal, prediabetic, and diabetic groups, except that there is a slight appearance that most of the smokers in the dataset fall in the prediabetic group, only because they have the largest group size. A true conclusion can't be drawn without obtaining the odds ratio value from a case-control study. A case-control study on smoking exposure in each group with depression (cases) and without diabetes (control) was conducted to obtain the odds ratio values which are displayed in the table below, along with the 2 by 2 tables of cases and controls against exposures.

B. STATISTICAL TESTS

Table 1. Data of Normal, Prediabetic, and Diabetic Exposure to Smoking on Cases of Depression

	Exposure	Depression +	Depression -	**Total**
Normal	Smoking +	7	23	**30**
	Smoking -	34	88	**122**
	Total	**41**	**111**	**152**
Prediabetes	Smoking +	25	68	**93**
	Smoking -	51	224	**275**
	Total	**76**	**292**	**368**
Diabetes	Smoking +	7	22	**29**
	Smoking -	18	49	**67**
	Total	**25**	**71**	**96**

After running case-control method with the epi2by2 function in the epiR package in R, the odds ratio values and χ^2 Values that indicate true association between exposure of smoking on depression were obtained. According to Table 2 below, both normal and diabetic groups have a odds ratio of less than 1, indicating a protective attribute of smoking exposure to depression. For the normal glycemic and diabetic subjects, the odds ratios (95% CI) found were 0.79 (0.31, 2.00) and 0.87 (0.32,2.37). These confidence intervals, however, include 1, which makes the odds ratios not significant. The odds ratio found from the prediabetes group was 1.61, with confidence interval of (0.93, 2.80). Although the confidence interval is very close to excluding 1, it still falls under the range, which means the odds ratio cannot be considered significant, and there isn't sufficient evidence to suggest that smoking affects the outcome of depression in the subjects.

Table 2. Odds Ratio and χ^2 Values Regarding Smoking Exposure on Outcome of Depression in Normal, Prediabetic, and Diabetic Subject Groups

	Odds Ratio	**χ^2 Value**
Normal	0.79 (0.31, 2.00)	0.251
Prediabetes	1.61 (0.93, 2.80)	2.947
Diabetes	0.87 (0.32, 2.37)	0.078

A similar methodology was done for alcohol consumption. The below table displays the 2 by 2 table of counts for alcohol exposure and depression in each subset.

Table 3. Data of Normal, Prediabetic, and Diabetic Exposure to Drinking on Cases of Depression

	Exposure	Depression +	Depression -	**Total**
Normal	Drinking +	3	58	**61**
	Drinking -	10	81	**91**
	Total	**13**	**139**	**152**
Prediabetes	Drinking +	5	140	**145**
	Drinking -	10	213	**223**
	Total	**15**	**353**	**368**
Diabetes	Drinking +	2	34	**36**
	Drinking -	8	52	**67**
	Total	**10**	**86**	**96**

The 2 by 2 matrices for each subset was run through the case-control algorithm, where odds ratios and χ^2 values were obtained to potentially find significant associations between drinking exposure and mental health status in individuals without type II diabetes, with pre-diabetes, and those with type II diabetes. Table 4 below displays these values, and it is seen that this dataset doesn't show any significant associations. According to the table all 3 subsets have odds ratio values of less than 1, like two of the groups found in smoking, indicating that drinking exposure has a protective quality against type 2 diabetes. The odds ratios (95% CI) were 0.42 (0.11, 1.59) for the normal group, 0.76 (0.25, 2.27) for prediabetic, and 0.38 (0.08, 1.91) for diabetic, and as observed with smoking, since the confidence intervals contain 1, the odds ratios are insufficient evidence to indicate association between alcohol consumption and depression in diabetic and non-diabetic individuals.

Table 4. Odds Ratio and χ^2 Values Regarding Drinking Exposure on Outcome of Depression in Normal, Prediabetic, and Diabetic Subject Groups

	Odds Ratio	**χ^2 Value**
Normal	0.42 (0.11, 1.59)	0.190
Prediabetes	0.76 (0.25, 2.27)	0.623
Diabetes	0.38 (0.08, 1.91)	0.388

An alternate approach that was taken to observe covariates in relation to each other was multiple logistic regression. Logistic regression was run using various combinations of covariates to yield significance in correlation. As stated in the hypothesis, the attributes E0SMOKE, which is smoking exposure, and E0GMETHWK, alcohol consumption, were the variables of interest to identify if their exposure alters the presence of depression in diabetic individuals. The following table, Table 5, shows the logistic regression coefficients extracted from the initial model on each subset with the covariates E0SMOKE and E0GMETHWK.

Table 5. Adjusted Odds Ratios (95% CI) and P-values from GLM Model with Covariates Smoking and Alcohol Exposure on Depression

	Covariates	**Adj. OR (95% CI)**	**P-Value**
Normal	**Smoking**	1.56 (0.46,5.31)	0.473
	Alcohol	0.99(0.97,1.023)	0.76
Prediabetes	**Smoking**	0.6 (0.3,1.24)	0.168
	Alcohol	0.97 (0.93,1.01)	0.194
Diabetes	**Smoking**	0.95 (0.37,2.44)	0.919
	Alcohol	1.00(0.99,1.01)	0.467

Because the initial model didn't yield significant correlation, various other models were tested to identify which, if any, covariate combination yielded some significance. After generating multiple models, one regression model in the diabetes subset showed that body fat had a significant relationship with depression after controlling for other variables. This final model used body fat (E0BFAT), smoking (E0SMOKE), alcohol (E0GMETHWK), heavy physical activity (E0HVYACT), and weight in kilograms (E0WTKG). Table 6 below provides the coefficients from the final model from the Diabetes Subset.

Table 6. Adjusted Odds Ratios (95% CI) and P-values from final GLM Model with Covariates Body Fat, Smoking, Alcohol, Heavy Physical Activity, and Weight in Kg on Depression for Diabetes Subset

	Adj. OR (95% CI)	P-Value
Body Fat	1.15 (0.98,1.36)	0.084
Smoking	0.86 (0.31,2.39)	0.768
Alcohol	1.00(0.99,1.01)	0.342
Heavy Physical Activity	5.15 (1.05,25.27)	0.043 *
Weight in Kg	0.99 (0.94,1.04)	0.676

* Significant value ($p < 0.05$)

A similar model was run for the normal glycemia and prediabetes subset to test to see if these results were consistent, however no significance was found with this same combination of covariates. These results are shown in Table 7 and 8 below.

Table 7. Adjusted Odds Ratios (95% CI) and P-values from final GLM Model with Covariates Body Fat, Smoking, Alcohol, Heavy Physical Activity, and Weight in Kg on Depression for Prediabetes Subset

	Adj. OR (95% CI)	P-Value
Body Fat	1.06 (0.96,1.17)	0.257
Smoking	0.55 (0.27,1.15)	0.113
Alcohol	0.98 (0.94,1.02)	0.256
Heavy Physical Activity	0.97 (0.53,1.75)	0.915
Weight in Kg	0.993(0.958,1.030)	0.712

Table 8. Adjusted Odds Ratios (95% CI) and P-values from final GLM Model with Covariates Body Fat, Smoking, Alcohol, Heavy Physical Activity, and Weight in Kg on Depression for Normal Subset

	Adj. OR (95% CI)	**Pr(>|z|)**
Body Fat	0.94 (0.86,1.03)	0.202
Smoking	1.8 (0.52,6.25)	0.357
Alcohol	0.993 (0.963,1.024)	0.660
Heavy Physical Activity	0.96 (0.47, 1.68)	0.992
Weight in Kg	1.00665 (0.959,1.056)	0.782

Unlike the diabetes set, the group of prediabetic and normal patients showed no significant associations between heavy physical activity and development of depression in individuals without diabetes and with prediabetes.

VI. DISCUSSION

A. ANALYSIS

The purpose of this study was to identify if smoking and alcohol exposure or any other variables are related to the presence of depression in diabetic and nondiabetic individuals who are Mexican Americans from Starr County, Texas. Looking at the distribution of depression among each subset, illustrated in Figure 4, there are very few subjects that reported moderately severe to severe depression, but this is something that we can investigate further as they may exhibit more signs of diabetes than those who reported minimal depression. One limitation to take note of is that these are self-reported so there is no guarantee of accurate information. It was hypothesized that smoking and drinking habits would be reflected in the data as significantly associated with depression, especially in diabetic individuals, however, the results didn't support this hypothesis.

Initially, a case-control test was run on each subset of individuals based on diabetic status for smoking exposure, where the sample counts are indicated in Table 1. The case to control ratio for each set is 0.37 for the normal subset, 0.26 for the prediabetic subset, and 0.35 in the diabetic subset. An ideal case to control ratio would be 1 case for every 1 to 5 controls, but due to the smaller size of the dataset, the ratio was a little higher. After running the case to control algorithm, the odds ratio and χ^2 values were obtained. As seen in Table 2, the normal and diabetic groups have odds ratio values of less than 1, which would signify that smoking has a protective association to depression, or the odds of having depression are reduced with exposure to smoking. For prediabetic individuals the odds ratio was 1.61 (0.93, 2.80), indicating that smoking would increase the risk of depression. However, for all three groups, the confidence interval includes 1, which means that the results are insufficient evidence to claim that there is a correlation between smoking exposure and presence of depression in diabetic individuals. This is also seen from the χ^2 values of, 0.251, 2.947, and 0.078 for normal, prediabetic, and diabetic, respectively. One explanation for this is simple. Looking at the total number of individuals for each group (152 for normal, 368 for prediabetic, and 96 for diabetic), the drastic difference may not be truly representative of the whole population. For example, it was expected that there would at least be significance in the diabetic subset of individuals, but since there were only 92 total individuals with diabetes in the dataset it could be that it is an inaccurate representation of the true distribution of diabetic individuals who smoke.

This same problem is found with the case-control test run for drinking exposure. Looking at the 2 by 2 count data in Table 3, the case to control ratios observed are much smaller than that found in smoking. The ratios were 0.09 for the normal glycemic group, 0.04 for prediabetic, and 0.11 for diabetic. The odds ratios (95% CI) were 0.42 (0.11, 1.59) for the normal group, 0.76 (0.25, 2.27) for prediabetic, and 0.38 (0.08, 1.91) for diabetic. The odds ratio values of less than 1 would mean that drinking exposure decreases the likelihood of depression in individuals with type 2 diabetes and prediabetes, as well as normal individuals, however the confidence intervals contain 1, so the data is not sufficient evidence to suggest a significant correlation. This is also observed in the χ^2 values of 0.190, 0.623, and 0.388 for normal, prediabetic, and diabetic, respectively, in Table 4. It was expected that the drinking exposure would be most significantly associated with depression in diabetic individuals, because of having to manage two difficult conditions, but this data didn't reflect that. This again could be due to there not being enough data to truly represent the population distribution of diabetes, prediabetes, and normal glycemia.

As a result of these limitations, a multiple logistic regression to identify potential predictors was conducted instead. Many regression models were created, the first looking at the focus variables E0SMOKE and E0GMETHWK. The coefficients from this model are displayed in Table 5. As seen in the table, in all three subsets, looking at smoking when controlling for alcohol consumption, and vice versa, there is no significant association seen between both covariates and depression. Looking at the estimate coefficient of each though, it is possible to get insights on whether the variables may increase or decrease depression. The estimates values, also called the regression coefficient, in the table are the log odds of incident Diabetes for predictor. It is estimated that changes in E0BFAT, E0SMOKE, and E0HVYACT are more likely to predict depression in diabetic individuals, because their values are not as close to 0 as E0GMETHWK and E0WTKG.

After various constructing many models on each subset, one model constructed with the Diabetes subset showed significance. This was used as a final model, and results are shown in Table 6. According to the results, E0HVYACT, which is heavy physical activity, is associated with depression, when holding body fat (E0BFAT), smoking (E0SMOKE), alcohol consumption (E0GMETHWK) and weight (E0WTKG) constant in the model, indicated by a P-value of 0.0434. There is no significant association between body fat (E0BFAT) and depression, when holding all other covariates in the model constant, shown by the P-value of 0.0843. Similarly, no significant association was found for E0SMOKE, E0GMETHWK, and E0WTKG, and depression, when controlling for all other covariates, signified by P-values of 0.7677, 0.3429, and 0.676 respectively.

VII. CONCLUSION

Overall, from this study it was observed that smoking and alcohol consumption are not significant predictors of depression in diabetic individuals as was initially hypothesized, however it was found that there is a positive association between heavy physical activity and depression while holding bodyfat, smoking, alcohol, and weight constant. Although this study yielded unexpected results, further exploration of a larger sample could have better results. If the sample size is increased, there may be more conclusive associations between the variables that were looked at in this study, especially E0SMOKE and E0GMETHWK. It is possible that the dataset didn’t truly reflect the count of smokers or drinkers in the whole population, which is why a larger sample might be better for gathering insights. To expand on this project, follow-up data on the individuals from this dataset that is available to observe how the variables change the outcome over time through running paired analysis tests.

VIII. REFERENCES

Hanis, Craig L., et al. "Beyond Type 2 Diabetes, Obesity and Hypertension: An Axis Including Sleep Apnea, Left Ventricular Hypertrophy, Endothelial Dysfunction, and Aortic Stiffness among Mexican Americans in Starr County, Texas." *Cardiovascular Diabetology*, vol. 15, no. 1, 2016, https://doi.org/10.1186/s12933-016-0405-6.

Shi, Liang, et al. "Physical Activity, Smoking, and Alcohol Consumption in Association with Incidence of Type 2 Diabetes among Middle-Aged and Elderly Chinese Men." *PLoS ONE*, vol. 8, no. 11, 2013, https://doi.org/10.1371/journal.pone.0077919.

"What Are the Risk Factors for Lung Cancer?" *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 18 Oct. 2021, https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm.

"Cox Proportional-Hazards Model." *STHDA*, http://www.sthda.com/english/wiki/cox-proportional-hazards-model.

"Cigarette Smoking: A Risk Factor for Type 2 Diabetes." *U.S. Food and Drug Administration*, Center for Tobacco Products, 4 May 2020, https://www.fda.gov/tobacco-products/health-effects-tobacco-use/cigarette-smoking-risk-factor-type-2-diabetes.

Chang, Sang Ah. "Smoking and Type 2 Diabetes Mellitus." *Diabetes & Metabolism Journal*, vol. 36, no. 6, 12 Dec. 2012, p. 399., https://doi.org/10.4093/dmj.2012.36.6.399.

"7 Common Withdrawal Symptoms." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 18 June 2021, https://www.cdc.gov/tobacco/campaign/tips/quit-smoking/7-common-withdrawal-symptoms/index.html.

Osna, Natalia A et al. “Alcoholic Liver Disease: Pathogenesis and Current Management.” *Alcohol research : current reviews* vol. 38,2 (2017): 147-161.

“The effects of alcohol on physiological processes and biological development.” *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism* vol. 28,3 (2004): 125-31.

Zeratsky, Katherine. “Diabetes: Does Alcohol and Tobacco Use Increase My Risk?” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 11 July 2020, https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/expert-answers/diabetes/faq-20058540.

Carlsson, S., et al. “Alcohol Consumption and the Incidence of Type 2 Diabetes: A 20-Year Follow-up of the Finnish Twin Cohort Study.” *Diabetes Care*, vol. 26, no. 10, 2003, pp. 2785–2790., https://doi.org/10.2337/diacare.26.10.2785.

Khan, Moien Abdul, et al. “Epidemiology of Type 2 Diabetes – Global Burden of Disease and Forecasted Trends.” *Journal of Epidemiology and Global Health*, vol. 10, no. 1, 2019, p. 107., https://doi.org/10.2991/jegh.k.191028.001.

“Depression.” *World Health Organization*, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/depression.

"Diabetes and Depression: Coping with the Two Conditions." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 3 Sept. 2020, https://www.mayoclinic.org/diseases-conditions/diabetes/expert-answers/diabetes-and-depression/faq-20057904.

"What to Expect When You Quit Smoking." *What to Expect When You Quit Smoking - Better Health Channel*, https://www.betterhealth.vic.gov.au/health/healthyliving/What-to-expect-when-you-quit-smoking.

Michael Kaliszewski. "Alcohol & Diabetes: Can Alcohol Cause Diabetes?" *American Addiction Centers*, 26 Oct. 2021, https://americanaddictioncenters.org/alcoholism-treatment/alcohol-abuse-and-diabetes.

"Diabetes and Mental Health." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 7 May 2021, https://www.cdc.gov/diabetes/managing/mental-health.html.

"Insulin Resistance & Prediabetes." *National Institute of Diabetes and Digestive and Kidney Diseases*, U.S. Department of Health and Human Services, https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance.

"Prediabetes." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 22 Sept. 2020, https://www.mayoclinic.org/diseases-conditions/prediabetes/symptoms-causes/syc-20355278.

"Alcohol and Diabetes." *Alcohol and Diabetes | ADA*, https://www.diabetes.org/healthy-living/medication-treatments/alcohol-diabetes.

Sherman, J. J. "The Impact of Smoking and Quitting Smoking on Patients with Diabetes." *Diabetes Spectrum*, vol. 18, no. 4, 2005, pp. 202–208., https://doi.org/10.2337/diaspect.18.4.202.

"Alcohol Questions and Answers." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 16 Feb. 2021, https://www.cdc.gov/alcohol/faqs.htm.

RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Virasakdi Chongsuvivatwong (2018). epiDisplay: Epidemiological Data Display Package. R package version 3.5.0.1. https://CRAN.R-project.org/package=epiDisplay

Mark Stevenson, Evan Sergeant with contributions from Telmo Nunes, Cord Heuer, Jonathon Marshall, Javier Sanchez, Ron Thornton, Jeno Reiczigel, Jim Robison-Cox, Paola Sebastiani, Peter Solymos, Kazuki Yoshida, Geoff Jones, Sarah Pirikahu, Simon Firestone, Ryan Kyle, Johann Popp, Mathew Jay, Charles Reynard, Allison Cheung, Nagendra Singanallur and Aniko Szabo. (2021). epiR: Tools for the Analysis of Epidemiological Data. R package version 2.0.39. https://CRAN.R-project.org/package=epiR

Multiple Logistic Regression Analysis. sphwebbumcbuedu. https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Multivariable/BS704_Multivariable8.html.

CDC. 2019. CDC - Frequently Asked Questions - Alcohol. Centers for Disease Control and Prevention. https://www.cdc.gov/alcohol/faqs.htm.