Mind and Body: An Investigation of Depression and Physical/Mental Health Condition

Austin Chan, Katie Evers, Jose Mawyin, Jimmy Ng

Data 621, CUNY School of Professional Studies

Abstract

In 2017-2018, the Centers for Disease Control and Prevention conducted the National Health and Nutrition Examination Survey (NHANES). The survey data contained information regarding depression, current health status, occupation, sleep disorders, and smoking. The data was collected from over 5500 individuals aged between 18 and 65. This report is a quasi-experimental research study with the goal of investigating the relationship between depression, sleep and general physical health condition. For instance, is there any relationship between one's physical health and depression? Is chronic fatigue a psychosomatic symptom of depression? Is there a linear relationship between hours of sleep and physical health? Is there any lifestyle pattern that would indicate physical health status, such as any significant difference between healthy and poor health individuals? Using logistic regression, our team examined factors that contribute significantly to depression. The results suggest that there are relationships between the following variables: depression and chronic fatigue, depression and physical health condition, and chronic fatigue and physical health condition. The most significant predictors include general health conditions, having a head or chest cold, having a stomach or intestinal illness, and type of work done in the previous week. This study suggests that maintaining a vigorous and wholesome lifestyle pattern is correlated with behavioral changes that can bring positive influence to depression.

Key words: depression, health, sleep

Introduction/Literature Review

Depression is a major psychological disorder that impacts over 6% of the US population (Lee, Stockings, Harris & Doi. 2019). In 2012, it topped ischemic heart disease as the number one cause of disability worldwide, and women had twice the risk of men (Lee et al., 2019). In addition, the comorbidity of depression and cardiometabolic disorders will be one of the primary causes of disability worldwide by 2020, with women also at twice the risk (Bartova & Pezawas, 2015; Lee et al, 2019). Depression can cause considerable physical and psychological impact. Although depressed women have more frequent suicidal thoughts, depressed men are more likely to succeed and die from suicidal behaviors (Lee et al, 2019).

Many researchers have examined the relationship between sleep and depression over the years. In a 2008 UK study of 8,580 participants, the occurrence of insomnia symptoms varied widely with age. Overall, 83% of depressed patients had at least one insomnia symptom. In the 16-to-24-year age group insomnia affected 77% and it affected 90% of patients in the 55-to-64-year age group (Nutt, Wilson, & Paterson, 2008). Hypersomnia was more prevalent in younger patients with about 40% of patients under 30 and 10% in the 50's experiencing the symptom (Nutt, Wilson, & Paterson, 2008). There is a higher incidence of sleep disorders in females of all ages. In another study, researchers examined a subgroup from 1994 and 1995 surveys of a Alameda County (California) Study comprising 2,370 residents aged over 50. They concluded that sleep disturbance exhibits strong patterns of co-occurrence with other symptoms of major depression (Roberts, Shema, Kaplan, & Strawbridge, 2000). Sleep disorders cause huge distress and negatively impact quality of life (Nutt, Wilson, & Paterson, 2008).

Promising research suggests that it may be possible to treat certain forms of depression through sleep deprivation. Pflug and Tolle (1971) found that sleep deprivation treatment can improve the health condition for individuals with endogenous depression. These techniques of sleep deprivation treatment include total sleep deprivation, partial sleep deprivation, phase advance of sleep periods, and REM (rapid eye movement) deprivation (Gillin, 1983). These treatments were administered in a very specific context, which means that sleep deprivation may not be the best treatment for individuals who are experiencing other forms of depression or who are already prone to sleep cycle irregularity. In fact, individuals who experienced sleep deprivation involuntarily may be at significant risk of mood disorders and suicide (Holsboer-trachsler & Seffritz, 2000).

Even when individuals are in good physical health, depression can influence their *perceived* physical health as well. A study conducted in 2000 explored the extent to which factors commonly associated with negative outcomes of aging also predict positive perceived health in a group of older residents. Questionnaires were randomly administered to 700 participants. A key takeaway was that fewer chronic conditions, better physical performance status, and the absence of depression all predicted positive perceived health (Bryant, Beck, & Fairclough, 2000). The studies discussed so far support the idea that sleep disorders are a symptom of depression, and that depression contributes to negative perceived health.

When an individual is depressed, there are physiological responses that can occur in the body that negatively impact their physical health. Researchers from Taiwan discovered that there is a strong positive correlation between depression and level of HDL cholesterol, and the ratios of TC/HDL and LDL/HDL (atherogenic index) might be another significant marker of the possible connection between serum lipids and anxiety or depression (Huang, et al, 2003). Essentially, they confirmed that depression could directly impact cardiovascular health in Taiwanese patients. Furthermore, these researchers found that patients suffering from depression also experienced significantly higher levels of stress and more trouble in sleeping, which could lead to decline in physical health and life satisfaction. They concluded that depression could have a long lasting impact on cardiovascular health, mood swing and sleep. These patients still constantly experienced trouble sleeping and other digestive problems after they were discharged from hospitals and no longer experienced any depressive or anxious symptom.

Depressed individuals have significantly decreased inclination to participate in physical activity, even at the suggestion of a doctor. Roshanaei-Moghaddam, Katon, and Russo (2009) found that baseline depression could be a significant risk factor for developing a sedentary lifestyle. In this study the researchers reviewed multiple medical databases and longitudinal experiments on depression to determine if depression can lead to a sedentary lifestyle. Among patients who reported baseline depression, many became more sedentary and were less likely to participate in physical exercise. This research implies a dangerous consequence for depressed people who have depression: their physical health can deteriorate as a result of a sedentary lifestyle due to depression. Luckily, other research suggests that this trend is reversible for certain populations. Jerstad et. al. (2010) found that adolescent girls who increased their physical activity could significantly reduce the risk for future increases in depressive symptoms. This research suggests that physical activity could be a way to mitigate the effects of depression for individuals who are already depressed.

More researchers found a strong connection between depression and sleep. Pilcher, Ginter and Sadowsky (1997) found that quality of sleep played a significant role in depression. Subjects were randomly assigned to participate in a 7-day sleep experiment. Their physiological and psychological conditions were measured before and after the experiment. Pilcher and colleagues found that quantity of sleep could not determine changes in psychological responses, such as feeling tension, depression, anger, anxiety, etc. However, subjects experienced better quality of sleep in general felt less depressed and angry toward others and they reported much significantly higher life satisfaction. Researchers concluded that frequent trouble sleeping, and chronic fatigue could be a dangerous sign or precursor of a major depressive episode. Furthermore, they strongly suggested that changes in lifestyle and sleep pattern, could be extremely helpful in treating depression or mood-related psychological disorders.

Researchers found that patients suffering from depression are also likely to have suffered from trouble sleeping. These patients either did not get enough sleep or reported that they had significantly longer hours of sleep (12 hours daily as an adult between 18 and 65) than the average person. Riemann, Berger and Voderholzer (2001) conducted polysomnographic sleep research and demonstrated that patients suffered from depression were all characterized by the following: reduction of slow wave sleep, disinhibition of REM sleep, shortening of REM latency, prolongation of the first REM period, and increased REM density. Basically, they experienced much shorter duration and frequency in REM sleep (a significant indicator of sleep quality) and they were more easily awakened during REM sleep. This research indicated a strong bi-directional treatment for depression, emphasizing traditional cognitive and psychiatric treatment. It also enhanced biofeedback on training in deep sleep, which was viewed as equally important in alleviating depression symptoms.

There is a strong consensus in psychiatric research that insomnia is most commonly associated with depression (Tsuno, Besset and Ritchie, 2005). It is estimated that 90% of patients with depression experienced very poor quality of sleep. Numerous sleep studies have provided extensive observations and theoretical hypotheses concerning the etiology and pathophysiology of depression since the early 1970s. These authors argued that treatment of insomnia cannot avoid treatment of depression, and vice versa. Success of anti-depressive treatment mostly depended on success of training and maintaining good sleep patterns (or as referred to “sleep architecture”). There were already numerous evidence-based studies providing a strong connection between sleep alterations and depression. These authors also argued that prevention of insomnia could help prevent or relieve impacts of depression. Sleep was viewed as a particularly important indicator for health status. Enforcing a regular and wholesome sleep pattern could be hugely beneficial to individuals with or without depression. Sleep played a very important role in a functional and healthy lifestyle.

The motivation behind this report is to investigate what factors contribute to self-reported general health condition and depression. The central ideas of the analysis are to understand whether there is a relationship among depression, sleep, and physical health condition and to understand what factors that can significantly accurately predict depression and health. This is achieved by exploring and analyzing five survey data sets sourced from the Centers for Disease Control and Prevention.

Methodology

The data used in this analysis was compiled from five 2017-2018 NHANES data sets on the Centers for Disease Control and Prevention website. NHANES is a program of studies that assess the health and nutritional status of adults and children in the United States. The five surveys contain data on Mental Health - Depression Screener (DPQ_J), Occupation (OCQ_J), Sleep Disorders (SLQ_J), Current Health Status (HSQ_J), and Smoking - Cigarette Use (SMQ_J). Table 1 in Appendix A lists all of the variables and their definitions. The 5,533 survey participants were males and females aged between 18 and 65 years old. Individual data sets were combined by using a common key (respondent sequence number) and variables with a significant number of missing values were dropped. All survey responses that were “Refused” to answer and “Don’t know” were coded as “7” and “9” in the original data. These observations were filtered out for the analysis. Data transformations were used to create the following new variables: depression score, average daily sleep hours, health status, sleepiness, depression indicator, and mental health problem indicator. These transformations were achieved by aggregating variables and assigning factor levels.

We decided to test the following null hypotheses:

1) There is no relationship between depression and chronic fatigue
2) There is no relationship between depression and physical health status
3) There is no relationship between chronic fatigue and physical health status

In addition, we explored features in lifestyle that were more likely associated with healthy individuals. We carried out ANOVA and Chi-squared tests for hypothesis testing. Additionally, we applied logistic regression models to investigate the probability of an individual having a mental health problem.

Experimentation and Results

Data Exploration

The NHANES survey data used in this analysis contains 5,533 observations of 34 variables. The number of variables compiled from each survey are as follows: nine from Mental Health - Depression Screener, seven from Current Health Status, two from Occupation, ten from Sleep Disorders, and five from Smoking - Cigarette Use. Summary statistics of all of the variables can be found in Table 2 of Appendix A. As seen in that table, the largest number of missing values for all variables is 443, which is about 8% of total observations. In the bar chart for general health (Figure 1) looks normally distributed. It also appears that as general health worsens, the greater percentage of larger depression scores there are. Depression score, seen in Figure 2, is left skewed where a majority of depression scores are zero. Depression score was calculated by summing the scores from all DPQ questions into one aggregate value. For classification purposes, we decided that individuals with a score of one or higher were considered “depressed”. For classification purposes, we decided that individuals with a score of one or higher were considered “depressed”. Figures 8 and 9 show the distribution of trouble sleeping by depression score and general health, respectively. As seen in the charts, individuals with lower depression scores are less likely to report trouble sleeping. There is a possible confounding factor between depression, general health, and trouble sleeping. As the frequency of feeling overly sleepy increases, we see that the proportion of higher depression scores also increase in Figure 10. Similarly, in Figure 11, we see that the proportion of higher depression scores increases as self reported trouble sleeping worsens. The distribution of hopelessness (Figure 12) shows that individuals with fair or better health tend to exhibit lower hopelessness scores compared to those with poor health. This observation suggests that there could be a moderate correlation between hopelessness and health condition. The correlation matrix (Figure 13) shows the strength of correlation coefficients between all of the variables. In the top left cluster, we can see that the DPQ questions tend to be positively correlated with each other. These questions correspond to the indicators of depression. When someone responds positively to one of the questions, it is likely that they will respond positively to the other questions as well. This led us to combine these variables into one “depression indicator” to prevent multicollinearity. In the bottom middle cluster, we can see that the SLQ questions tend to be positively correlated with each other. These questions correspond to smoking habits. When someone responds positively to one of the questions, it is likely that they will respond positively to the other questions as well. This finding is unsurprising given that people who smoke will tend to try multiple different smoking products.

Hypothesis Testing

To test our first hypothesis, we conducted a one-way between subject ANOVA to compare the effect of chronic fatigue associated with severity of depression (score ranged from 0 to 27). There was a significant effect of level of fatigue for the five conditions $F(4, 5055) = 204.5$, $p = 0.00$. A post hoc Tukey HSD further demonstrated that there is no difference between people who were never or rarely tired, but there is significant differences among other conditions. In general, the level of chronic fatigue is statistically, significantly associated with level of depression (as seen in Figure 3).

Second, a one-way between subject ANOVA was conducted to test the second hypothesis by comparing physical health conditions associated with the severity of depression. There was a significant effect of physical health condition for the five conditions $F(4, 5061) = 244$, $p = 0.00$. A post hoc Tukey HSD further demonstrated that there is no difference between people who reported to be in very good or excellent health, but there is significant differences among other conditions. In general, the level of physical health condition was statistically, significantly associated with levels of depression (as seen in Figure 4). People who reported to be in fair or poor health condition were more likely to be depressed.

Third, there was a significant relationship between health condition and chronic fatigue, X^2 (16, N = 5126) = 20504, $p < .00$. Essentially, people associated with poor health were more likely to feel fatigue throughout the day (as seen in Figure 5). On the other hand, individuals with excellent health conditions were less likely to feel tired. Fatigue was an indicator of poor health and it revealed a person's underlying physical and psychological problem.

It was natural to assume that healthy people tended to sleep more and that is the reason why they felt less tired throughout the day. However, there was no statistical evidence in associating hours of sleep with general health conditions (please see Figure 6). People varied in physical health conditions did not statistically differ in the number of hours of sleep per night. Everyone in this study shared a similar average hour of sleep (M = 7.9, SD = 1.5). It was speculated that people differed in the "quality" not "quantity" of sleep. The difference in the quality of sleep contributed to various degrees of physical and psychological distress.

Furthermore, we conducted odds ratio to compare people with excellent health against people with poor health conditions. There is a clear lifestyle pattern emerged in the data (as seen in Figure 7). For instances, healthy individuals were less likely to have seek any professional help for trouble sleeping (odds ratio = 2.39; 95% confidence interval [CI] = 2.09, 2.73); they were more likely to have a job (odds ratio = 2.07; 95% confidence interval [CI] = 1.82, 2.36); they were more likely to sleep at regular time between 1030 pm to 11:30 pm (odds ratio = 2.01; 95% confidence interval [CI] = 1.40, 2.85); they were more likely to wake up around 7:30 am (odds ratio = 1.79; 95% confidence interval [CI] = 1.23, 2.61); they were more likely to be non-smokers (odds ratio = 1.7; 95% confidence interval [CI] = 1.50, 1.93); they were more likely to have donated blood in the last 12 months (odds ratio = 1.7; 95% confidence interval [CI] = 1.20, 2.40), and etc. In general, they were more psychologically resilient and lived a more orderly pattern. Their lifestyle pattern would help create a buffer against any experience of depressive symptoms.

Data Preparation for Logistic Regression Model

After performing a series of statistical tests to determine if there were significant relationships between the variables of interest, we decided to model the data using a Logistic Regression Model to predict if a person would experience a mood disorder or not. RStudio, an integrated development environment for R, was used for data preparation, model training and analysis of the logistic regression model. The aim of this model was to analyze the relationship between various Mental Health indicators (DPQ_J) and potential correlators such as Occupation (OCQ_J), Sleep Disorders (SLQ_J), Current Health Status (HSQ_J) and Smoking (SMQ_J) provided by the NHANES survey data.

The dataset as downloaded contained 5,533 observations. The data had 9 Mental Health, 8 Current Health Status, 2 Occupation, 10 Sleep Disorders and 5 Smoking indicators. From this set of observations, the Mental Health indicators were selected as the response variables. The majority of the variables are factors with different levels and the rest of the observations have numerical values such as the number of hours slept.

In order to simplify the analysis, all the Mental Health indicators (with a range of 1 to 3) were aggregated into a single variable that represents if the person answering the survey is experiencing any mood disorders or not. This new variable had a range of 0 to 27 in severity of mood disorder issues. Remarkably, 34.1% of the observations had an aggregated mood disorder score of 0. We generated a second Yes/No mood disorder variable from the new aggregated mood disorder variable. Values equal to 0 were assigned the label "No" while other than 0 were assigned the label "Yes". This Yes/No variable was used as the target to predict in our logistic regression model.

As we can see in Figure 14, the data set is not complete with most of the missing entries in the responses to the Mental Health indicators (DPQ_J). Missing entries were imputed (replaced using substituted values) using the "MICE" library in R. Multivariate Imputation by Chained Equations (MICE) imputes missing values with (plausible) values drawn from a distribution specifically designed for each missing datapoint. Some of the observations had entries that were the equivalent of missing such as "Refused", "Don't know" or "Missing". Observations with these values were removed from the data set. After this process of data preparation, the data set was left with 91.8% of the original observations.

Model Building

From the set of 25 predictor variables, we trained a logistic regression model using the following predictor variables with the highest correlation, measure of the association between two variables, and calculated Yes/No mood disorder response variable:

- **HSD010** - General health condition (Factor with 5 levels Excellent, Very good,Good, Fair or Poor)
- **HSQ500** - SP have head cold or chest cold (Factor with 2 level Yes or No)
- **HSQ510** - SP have stomach or intestinal illness? (Factor with 2 level Yes or No)
- **HSQ520** - SP have flu, pneumonia, ear infection? (Factor with 2 level Yes or No)
- **HSQ571** - SP donated blood in the past 12 months? (Factor with 2 level Yes or No)
- **OCD150** - Type of work done last week (Factor with 4 levels Working at a job or business, With a job or business but not at work, Looking for work, or Not working at a job or business.)
- **SLD012** - Sleep hours - weekdays or workdays (Number from 2 to 14)
- **SLD013** - Sleep hours - weekends (Number from 2 to 14)
- **SLQ030** - How often do you snore? (Factor with 4 levels Never, Rarely, Occasionally or Frequently)

➢ **SMQ020** - Smoked at least 100 cigarettes in life (Factor with 2 level Yes or No)

We used the predictors to generate a binomial logistic regression model that could predict if a person was experiencing a mood disorder or not. Instead of predicting a continuous response, as is the case of linear regression, logistic regression generates the probability of the response being one of a categorical set of states. In our case, "Yes" or "No" experiencing a mood disorder. Based on this probability, a number from 0 to 1, the observation is assigned to either the "Yes" or "No" categorical variable.

Logistic Regression Models

Model 1: Training model using only explanatory variables

Using the "glm" package in R we trained a logistic regression model using the following parameters: 10 fold cross-validation, method="glm", family="binomial". The accuracy of the trained model is mediocre with an estimated classification accuracy of 0.66. Nevertheless, the model is useful because it outlines which predictor variables were statistically significant in explaining the aggregated mood disorder response variable.

As we can see in Figure 15, the most significant predictors from the selected set are **HSD010** (General health condition), **HSQ500** (SP have head cold or chest cold), **HSQ510** (SP have stomach or intestinal illness?) and **OCD150** (Type of work done last week). Figure 16 ranks all the predictor variables including the level of the factor itself in degree of importance. We can see that the relevant variables are those dealing with personal health rather than the lack or not of proper sleep of the person.

Model 2: Training model using explanatory variables and interaction terms

We trained a second logistic regression model to analyze the interactions between the different predictor variables. These interaction terms arise when one variable depends on the value of the other variable. From all possible interactions between factors and their levels, Figure 17 ranks them by importance in explaining the data used in the training set. Again just as in the previous regression model, we see how the predictor variables indicating personal health are the most statistically significant. Based on the absolute value of the z-score the most important interactor parameters are the following:

HSQ510 & SMQ020: Interaction between Stomach Health and Smoking
OCD150 & SLQ030: Interaction between Occupation and Snoring.
HSQ520 & SLQ030: Interaction between Flu and Snoring.
HSQ500 & OCD150: Interaction between Cold and Occupation.

The interaction between these terms is thought provoking. The most statistically significant interaction terms happen between unrelated fields (i.e. Health and Smoking or Health and Sleep). The compounding of issues may have a synergistic effect in determining how likely a person is to develop a mood disorder.

Model Evaluation and Comparison

We used AUC - ROC (Area Under the Curve - Receiver Operating Characteristics) parameter to compare the performance of both logistic regression models. This parameter, between 0 to 1, measures how well our model is capable of distinguishing between classes. The higher the value if from 0.5 (guessing at random) the more accurate is our model. Both models scored a AUC - ROC value of around 0.63 to 0.65 (shown in Figure 18). The second logistic regression model scored slightly better at the cost of increased complexity. The second model has a Fisher Score, a measure of model complexity, of 21 as compared to a score of 6 for the first model.

The results from the logistic regression models indicated that some of the predictor variables were statistically correlated to the aggregated mood disorder variable that we generated for this model. However, the predictive value of the model is quite low as the accuracy of the model was just above the threshold of a random prediction.

Discussion and Conclusions

After conducting the analysis, our team has confirmed that general health condition is the most significant factor in predicting depression among individuals. The next most important factors are specific physical ailments like head/chest colds and stomach/intestinal pain. The severity of depression can be measured and manifested through psychosomatic symptoms. Also, our finding raises an interesting implication: the prevalence of depression has less to do with specific physical ailments and more to do with the overall self-perceived well-being of an individual. In other words, if an individual is not feeling well *in general*, there is a higher likelihood of predicting depression in that individual. This finding appears to fall in line with the current understanding of depression in clinical psychology. Given that this study is observational, we were unable to determine the direction of causality between general health condition and depression. A potential avenue for future research could be determining if there is a causal link between general health condition and depression or if there is a confounding factor that contributes both to general health condition and depression.

According to the first logistic regression model, head/chest colds and stomach/intestinal pain have a *negative* impact on predicting depression after controlling for general health condition. In other words, those with head/chest colds or stomach/intestinal pain tend to have a lower incidence of depression compared to those who do not have these physical ailments. This finding is interesting because intuition would suggest that those with physical ailments have a lower quality of life, which could be a possible cause for depression. However, this does not appear to be the case. There seems to be a more complicated relationship between the *actual* physical health condition of an individual, the *perceived* physical health condition, and depression. Though, without more specific data, it will be difficult to determine the true relationships between these variables. This could be a new opportunity for research that studies the effects of actual physical health condition on perceived physical health condition. Combined with the current research on depression, researchers can determine if the gap between actual physical health condition and perceived physical health condition has an effect on depression.

In terms of the physiological effects of depression, our analysis has observed that physical health condition and depression are intertwined at many levels. Reported physical health conditions like chronic fatigue were found to be significantly associated with level of depression (as seen in Figure 3). Non-chronic fatigue was found to be an indicator of poor health and had further implications for an individual's underlying physical and psychological health. These findings appear to be intuitive with the current understanding of depression given that fatigue is a common symptom for individuals experiencing depression. Knowing that fatigue can be an early indicator for depression, doctors can be more aware of future depressive episodes in patients if they report fatigue. More research can be conducted on measuring the causal relationship between depression and fatigue.

In measuring the effect of sleep on general health conditions, we found that there was no statistical evidence in associating hours of sleep with general health conditions (see Figure 6). People varied in physical health conditions did not statistically differ in the number of hours of sleep per night. The difference in the quality of sleep contributed to various degrees of physical and psychological distress. Depression can have an interesting relationship with sleep, where individuals can experience sleep deprivation or hypersomnia, which are both indicators of depression. However, when sleep deprivation and hypersomnia are both included in a statistical model as the number of hours of sleep per night, this can result in a confounding factor where individuals with depression vary significantly in the amount of sleep they get per night. Perhaps more research can be performed on whether hypersomnia and sleep deprivation have different effects on depression.

So far, we have only discussed the effect of depression on direct, internal factors like mental and physical health. However, depression can have a significant effect on external factors in one's life as well. Our research has found that there is a clear lifestyle pattern among people who are not depressed (as seen in Figure 7). Healthy individuals had regular sleep patterns, were more likely to be employed, were less likely to smoke, and were more likely to donate blood. In general, healthy individuals were more psychologically resilient and lived a more organized life. Their lifestyle pattern would help create a buffer against any experience of depressive symptoms. It is possible that organizing one's life and priorities could be a means to treat depression. More research should be conducted on whether individuals with highly routine lifestyle patterns are less likely to experience depression.

8. References

Bartova, L. & Pezawas, L. (2015). Recurrent brief depressive disorder. In Stolerman, IP and Price, LH (eds). *Encyclopedia of Psychopharmacology*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg, pp. 1456–1459.

Begdache, L., Kianmehr, H., Sabounchi, N., Marszalek, A. , Dolma, N. (2019). Principal component regression of academic performance, substance use and sleep quality in relation to risk of anxiety and depression in young adults. *Trends in Neuroscience and Education*, 15, 29–37.

Bryant, L. L., Beck, A., & Fairclough, D. L. (2000, May). Factors that Contribute to Positive Perceived Health in an Older Population. *Journal of Aging and Health*. https://journals.sagepub.com/doi/pdf/10.1177/089826430001200202?casa_token=ymsMLNsoZjkAAAAA:rAGWW3u4TraE-4igX-wj2_KXBbgGrkfF6oJLdOGHGNlMdYK2ma1kL4WBiieG2tGhhNy73k710rBQOw

Christian Gillin, J. (1983). The sleep therapies of depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 7(2-3), 351-364.

Fava, M. (2004). Daytime Sleepiness and Insomnia as Correlates of Depression. *The Journal of Clinical Psychiatry*, 65(Suppl16), 27–32.

Holsboer-trachsler, E., Seffritz, E. (2000). Sleep in Depression and Sleep Deprivation: A Brief Conceptual Review. *The World Journal of Biological Psychiatry,* 1(4), 180-186.

Huang, T., Wu, S., Chiang, Y., & Chen, J. (2003). Correlation between serum lipid, lipoprotein concentrations and anxious state, depressive state or major depressive disorder. *Psychiatry Research*, 118, 147-153.

Jerstad, S. J., Boutelle, K. N., Ness, K. K., & Stice, E. (2010). Prospective reciprocal relations between physical activity and depression in female adolescents. *Journal of Consulting and Clinical Psychology*, 78(2), 268–272.

Nutt, D., Wilson, S., & Paterson, L. (2008). Sleep Disorders as Core Symptoms of Depression. *Dialogues in Clinical Neuroscience*, 10(3), 329-336. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181883/

Pilcher, J., Ginter, D. R., & Sadowsky, B. (1997). Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. *Journal of Psychosomatic Research*, 42, 583-596.

Pflug, B., Tolle, R. (1971). Disturbance of the 24-Hour Rhythm in Endogenous Depression and the Treatment of Endogenous Depression by Sleep Deprivation. *International Pharmacopsychiatry*, 6, 187-196.

Riemann, D., Berger, M., Voderholzer, U. (2001). Sleep and depression — results from psychobiological studies: an overview. *Biological Psychology*, 57, 67-103.

Roberts, R. E., Shema, S. J., Kaplan, G. A., & Strawbridge, W. J. (2000). Sleep Complaints and Depression in an Aging Cohort: A Prospective Perspective. *The American Journal of Psychiatry,* 157(1), 81-88. https://ajp.psychiatryonline.org/doi/full/10.1176/ajp.157.1.81

Roshanaei-Moghaddam, B., Katon, W., Russo, J. (2009). The longitudinal effects of depression on physical activity. *General Hospital Psychiatry*, 31(4), 306-315.

Tsuno, N., Besset, A., & Ritchie, K. (2005). Sleep and Depression. *The Journal of Clinical Psychiatry*, 66(10), 1254–1269.

9a. Appendix A: Supplemental Tables and Figures

Table 1: Variable Names and Definitions

Variable	Definition
SEQN	Respondent sequence number
DPQ010	Have little interest in doing things
DPQ020	Feeling down, depressed, or hopeless
DPQ030	Trouble sleeping or sleeping too much
DPQ040	Feeling tired or having little energy
DPQ050	Poor appetite or overeating
DPQ060	Feeling bad about yourself
DPQ070	Trouble concentrating on things
DPQ080	Moving or speaking slowly or too fast
DPQ090	Thought you would be better off dead
HSD010	General health condition
HSQ500	SP have head cold or chest cold
HSQ510	SP have stomach or intestinal illness?
HSQ520	SP have flu, pneumonia, ear infection?
HSQ571	SP donated blood in past 12 months?
HSQ590	Blood ever tested for HIV virus?
HSAQUEX	Source of Health Status Data
OCD150	Type of work done last week
OCD390G	Kind of work you have done the longest
SLQ300	Usual sleep time on weekdays or workdays
SLQ310	Usual wake time on weekdays or workdays
SLD012	Sleep hours - weekdays or workdays
SLQ320	Usual sleep time on weekends
SLQ330	Usual wake time on weekends
SLD013	Sleep hours - weekends
SLQ030	How often do you snore?
SLQ040	How often do you snort or stop breathing
SLQ050	Ever told doctor had trouble sleeping?
SLQ120	How often feel overly sleepy during day?
SMQ020	Smoked at least 100 cigarettes in life
SMQ890	Ever smoked a cigar even 1 time?
SMQ900	Ever used an e-cigarette?
SMQ910	Ever used smokeless tobacco?
SMAQUEX2	Questionnaire Mode Flag

Table 2: Summary Statistics

seqn	dpq010	dpq020	dpq030	dpq040
Min. : 93705	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.: 95945	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median : 98257	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000
Mean : 98276	Mean :0.3993	Mean :0.3593	Mean :0.6466	Mean :0.7618
3rd Qu.:100590	3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :102956	Max. :9.0000	Max. :9.0000	Max. :9.0000	Max. :9.0000
NA	NA's :439	NA's :440	NA's :440	NA's :441

Table: Table continues below

dpq050	dpq060	dpq070	dpq080	dpq090
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000
Mean :0.3995	Mean :0.2567	Mean :0.2722	Mean :0.1791	Mean :0.0615
3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.0000
Max. :9.0000	Max. :9.0000	Max. :9.0000	Max. :9.0000	Max. :9.0000
NA's :441	NA's :442	NA's :442	NA's :442	NA's :443

Table: Table continues below

hsd010	hsq500	hsq510	hsq520	hsq571	hsq590
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.00
1st Qu.:2.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:1.00
Median :3.000	Median :2.000	Median :2.000	Median :2.000	Median :2.000	Median :2.00
Mean :2.862	Mean :1.872	Mean :1.938	Mean :1.962	Mean :1.957	Mean :1.86
3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:2.00
Max. :9.000	Max. :9.000	Max. :9.000	Max. :9.000	Max. :9.000	Max. :9.00
NA's :396	NA's :396	NA's :396	NA's :396	NA's :396	NA's :396

Table: Table continues below

hsaquex	ocd150	ocd390g	slq300	slq310	sld012
Min. :2	Min. :1.000	Min. :1.000	23:00 :1130	06:00 : 951	Min. : 2.000
1st Qu.:2	1st Qu.:1.000	1st Qu.:1.000	22:00 :1089	07:00 : 810	1st Qu.: 6.500
Median :2	Median :1.000	Median :1.000	00:00 : 710	05:00 : 600	Median : 7.500
Mean :2	Mean :2.342	Mean :1.451	21:00 : 491	08:00 : 452	Mean : 7.621
3rd Qu.:2	3rd Qu.:4.000	3rd Qu.:2.000	22:30 : 365	06:30 : 427	3rd Qu.: 8.500
Max. :2	Max. :9.000	Max. :9.000	(Other):1713	(Other):2256	Max. :14.000
NA	NA	NA	NA's : 35	NA's : 37	NA's :44

Table: Table continues below

slq320	slq330	sld013	slq030	slq040	slq050
23:00 :1066	07:00 : 963	Min. : 2.0	Min. :0.000	Min. :0.0000	Min. :1.00
00:00 :1065	08:00 : 913	1st Qu.: 7.0	1st Qu.:0.000	1st Qu.:0.0000	1st Qu.:1.00
22:00 : 858	09:00 : 671	Median : 8.0	Median :2.000	Median :0.0000	Median :2.00
01:00 : 460	06:00 : 622	Mean : 8.3	Mean :2.071	Mean :0.9058	Mean :1.73
21:00 : 358	10:00 : 431	3rd Qu.: 9.0	3rd Qu.:3.000	3rd Qu.:1.0000	3rd Qu.:2.00
(Other):1701	(Other):1907	Max. :14.0	Max. :9.000	Max. :9.0000	Max. :9.00
NA's : 25	NA's : 26	NA's :51	NA	NA	NA

Table: Table continues below

slq120	smq020	smq890	smq900	smq910	smaquex2
Min. :0.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:1
Median :2.000	Median :2.000	Median :2.000	Median :2.000	Median :2.000	Median :1
Mean :1.797	Mean :1.597	Mean :1.644	Mean :1.803	Mean :1.855	Mean :1
3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:1
Max. :9.000	Max. :2.000	Max. :9.000	Max. :9.000	Max. :9.000	Max. :1
NA	NA	NA	NA	NA	NA

Figure 1: General Health Bar Chart

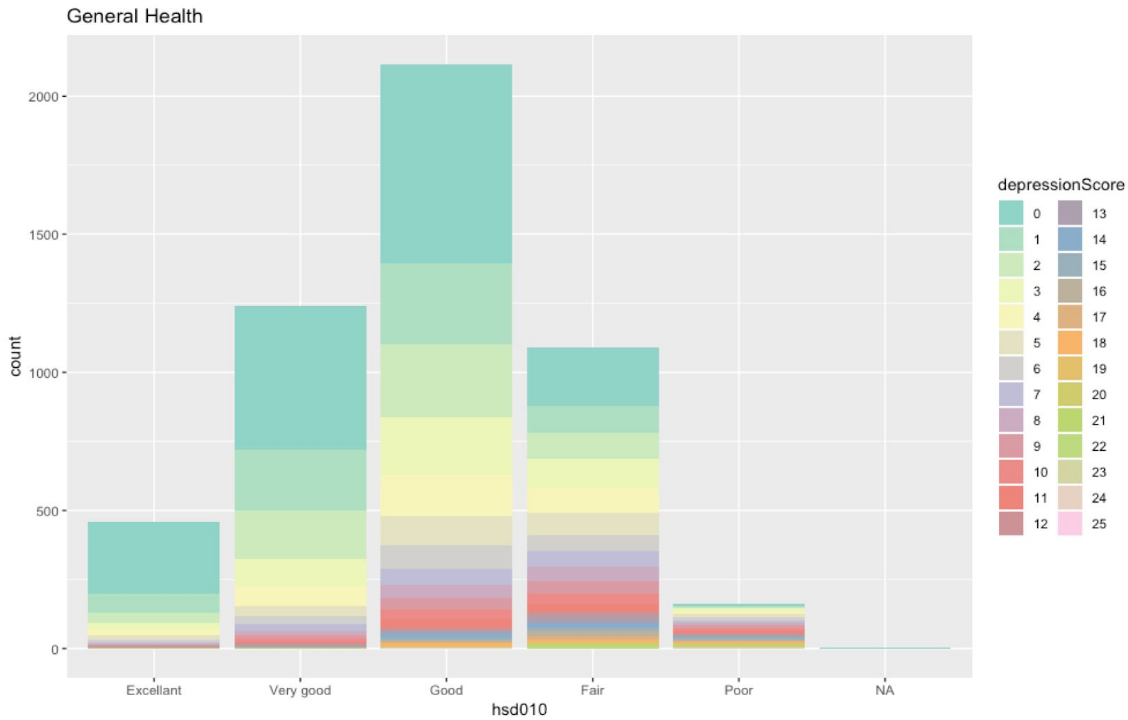


Figure 2: Depression Scores Bar Chart

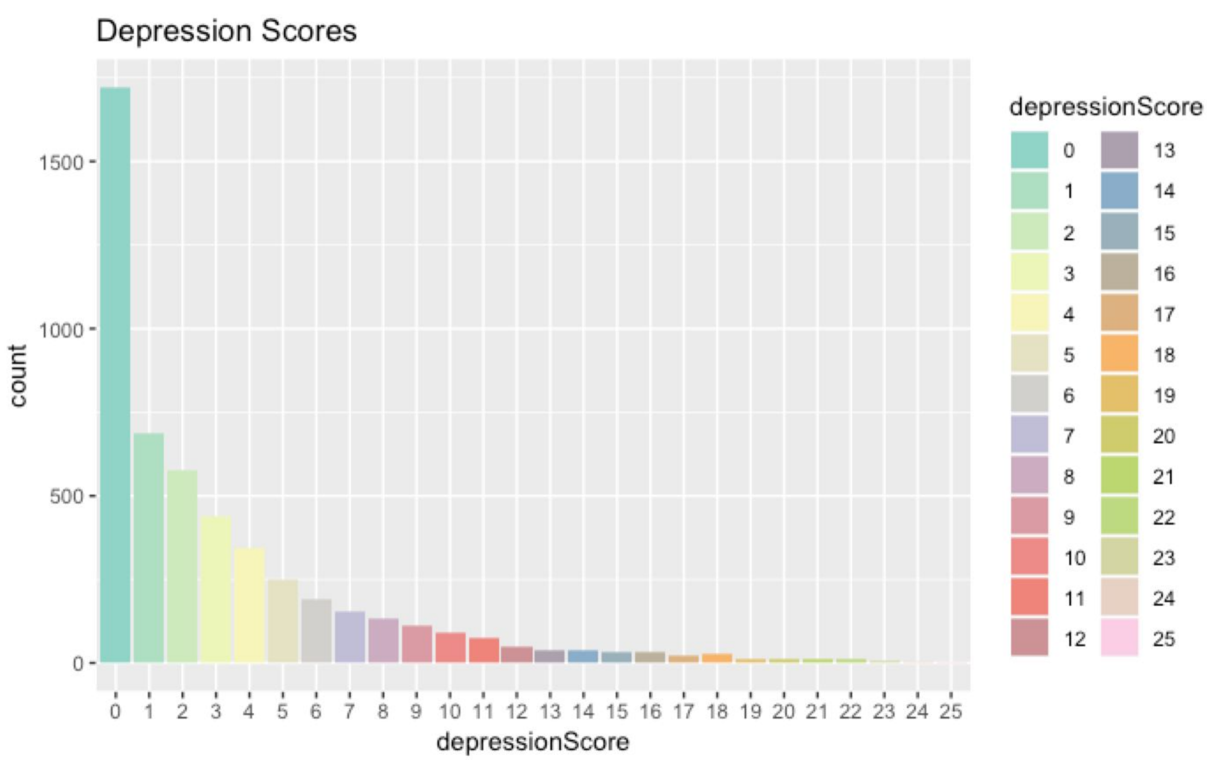


Figure 3. Depression levels varied by chronic fatigue

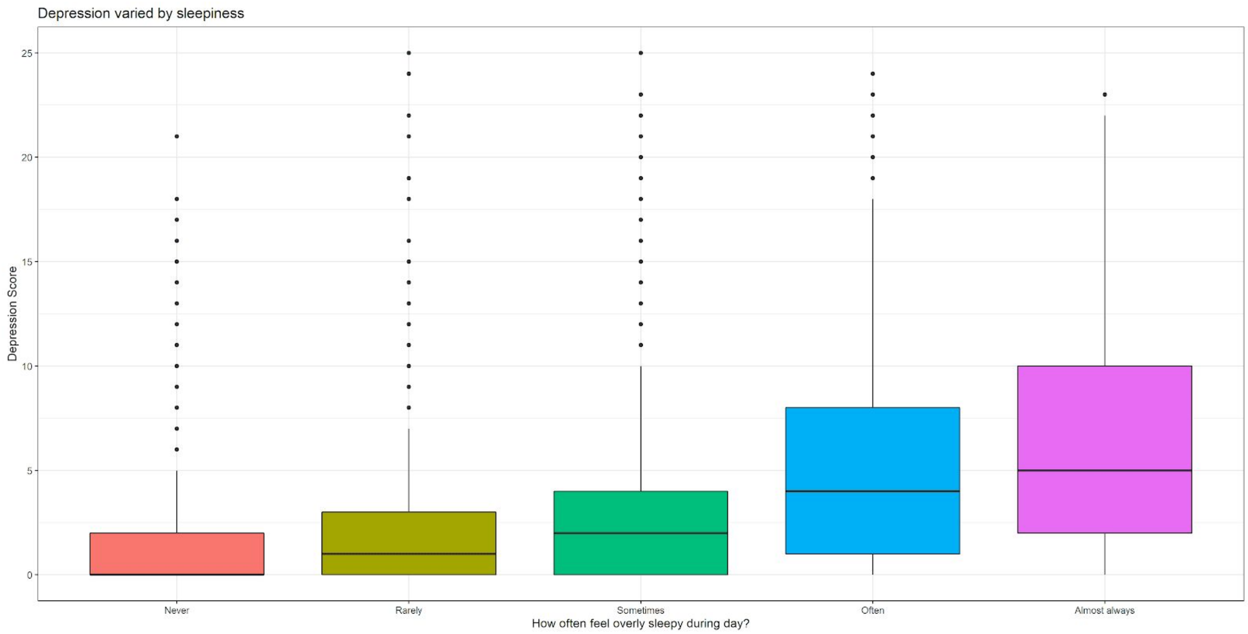


Figure 4. Depression levels varied by physical health condition

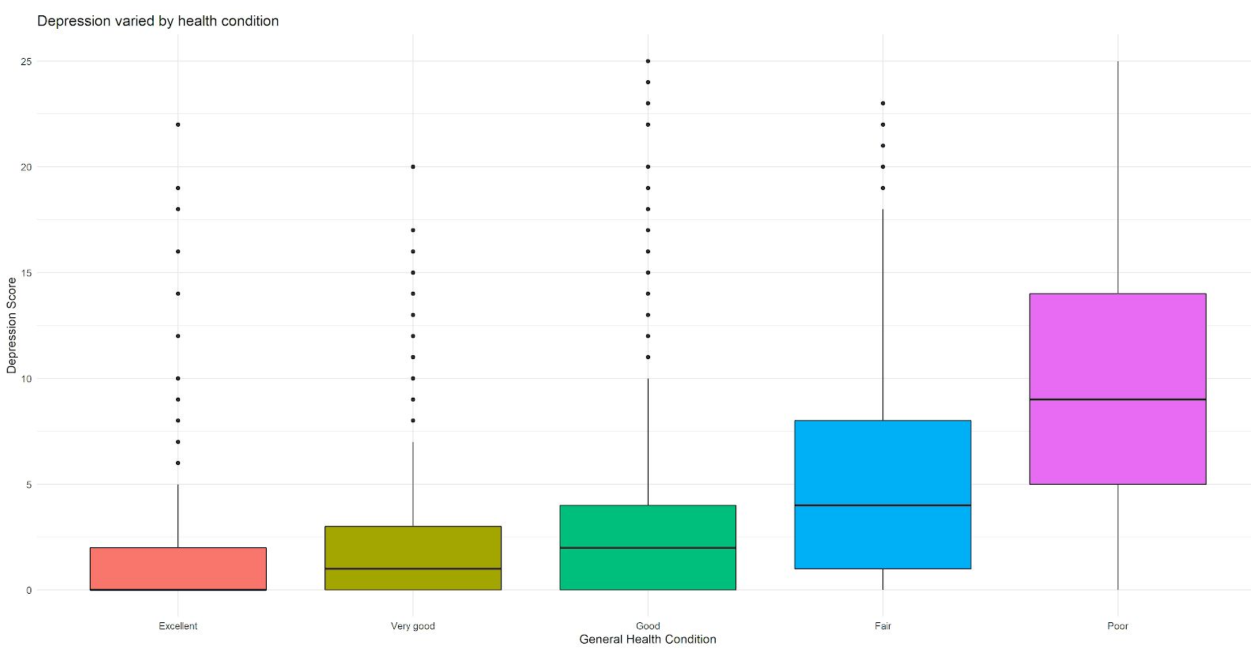


Figure 5. Health condition and fatigue

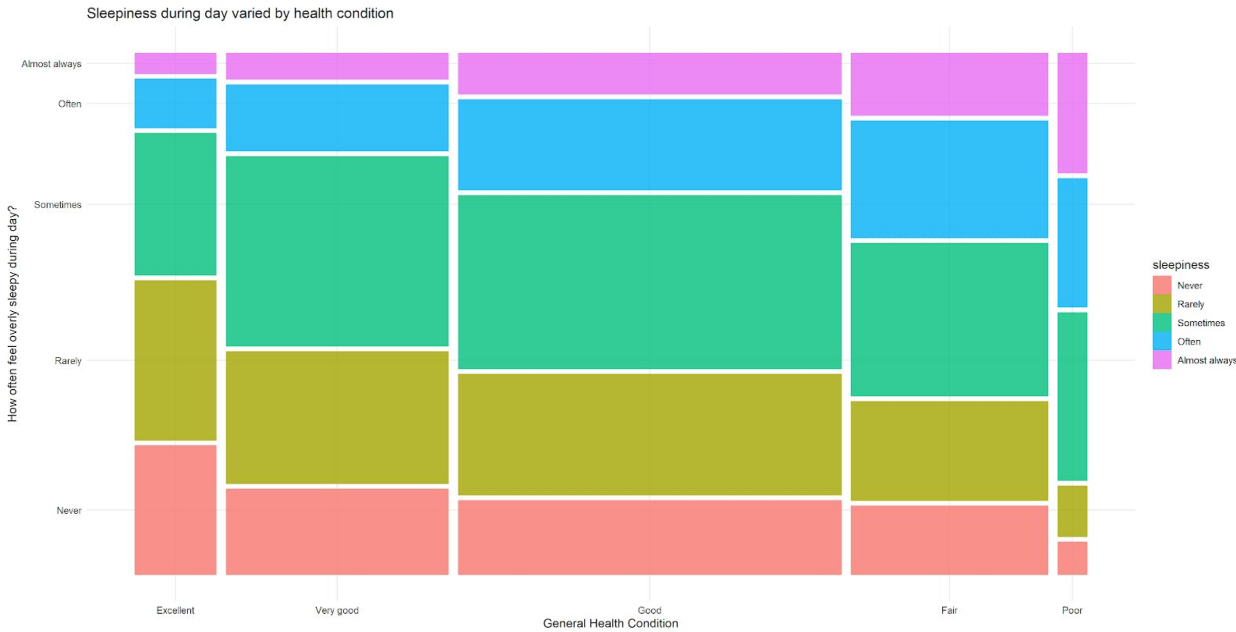


Figure 6. Distribution of hours of sleep by health condition

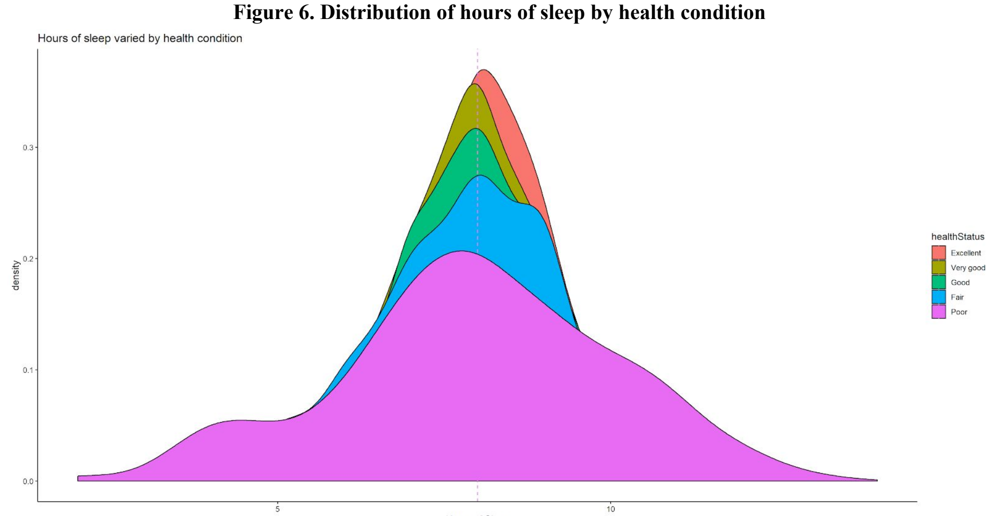


Figure 7. Odds ratio about features associated with people with excellent health condition

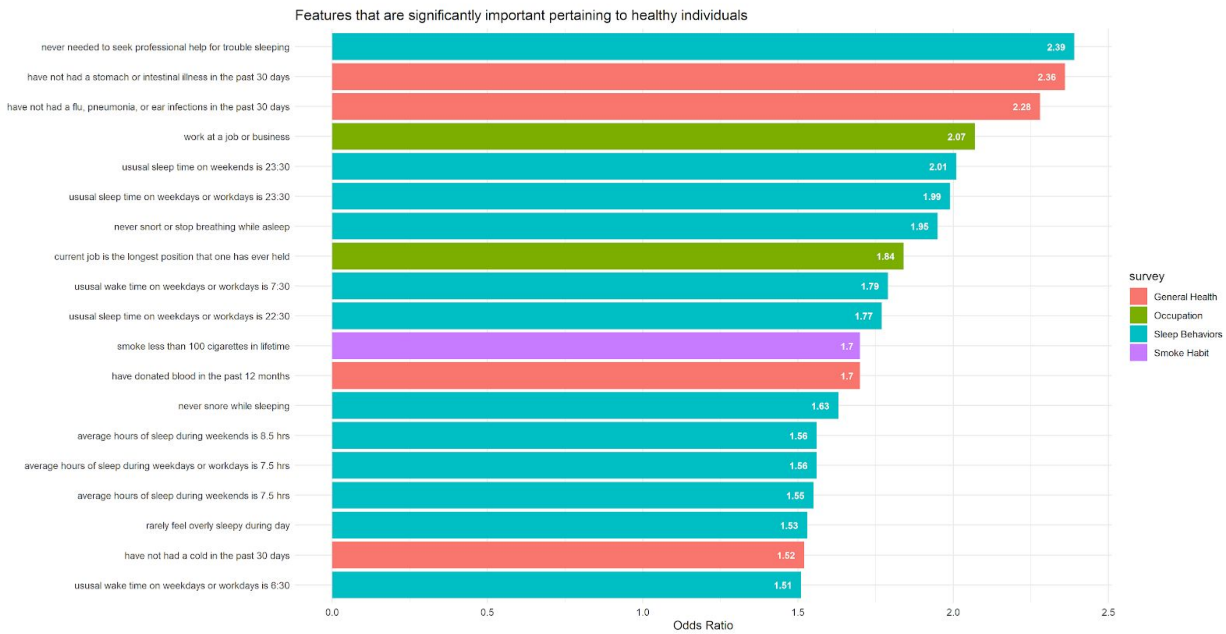


Figure 8: Ever Told Doctor Have Trouble Sleeping Bar Chart

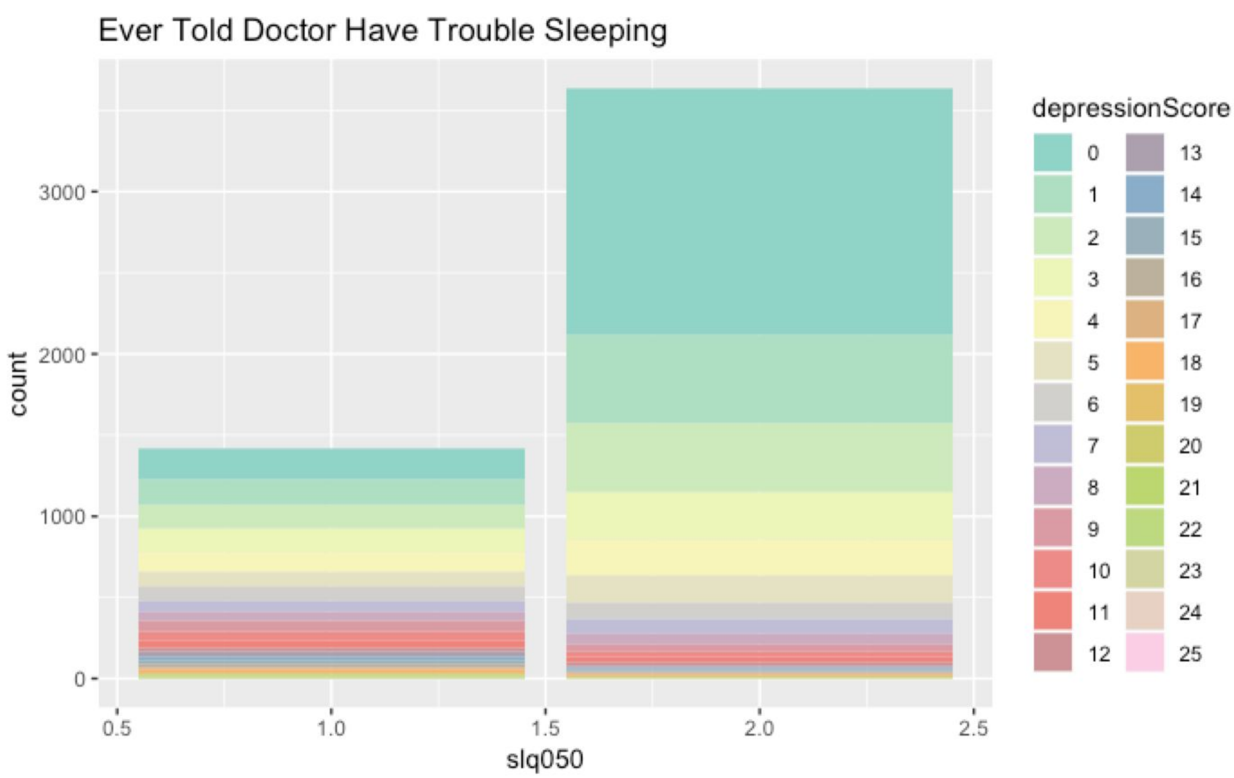


Figure 9: Ever Told Doctor Have Trouble Sleeping Bar Chart

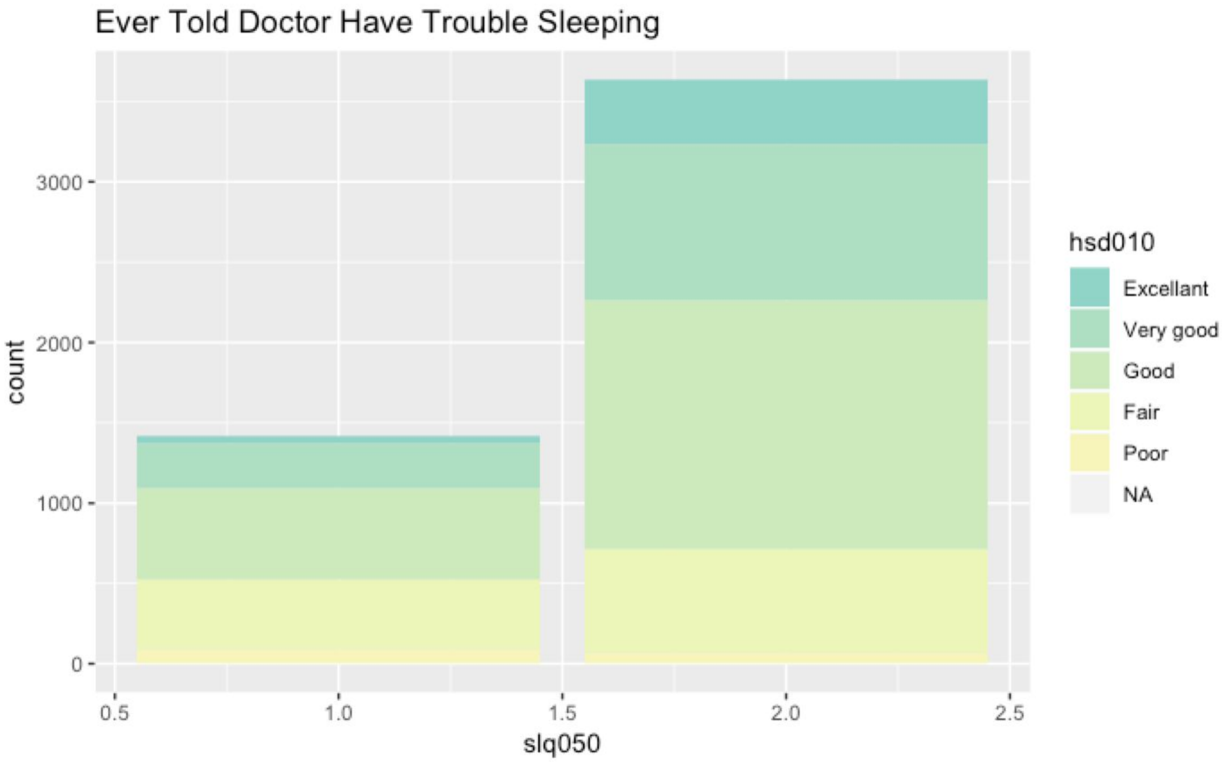


Figure 10: Feel Overly Sleepy Bar Chart

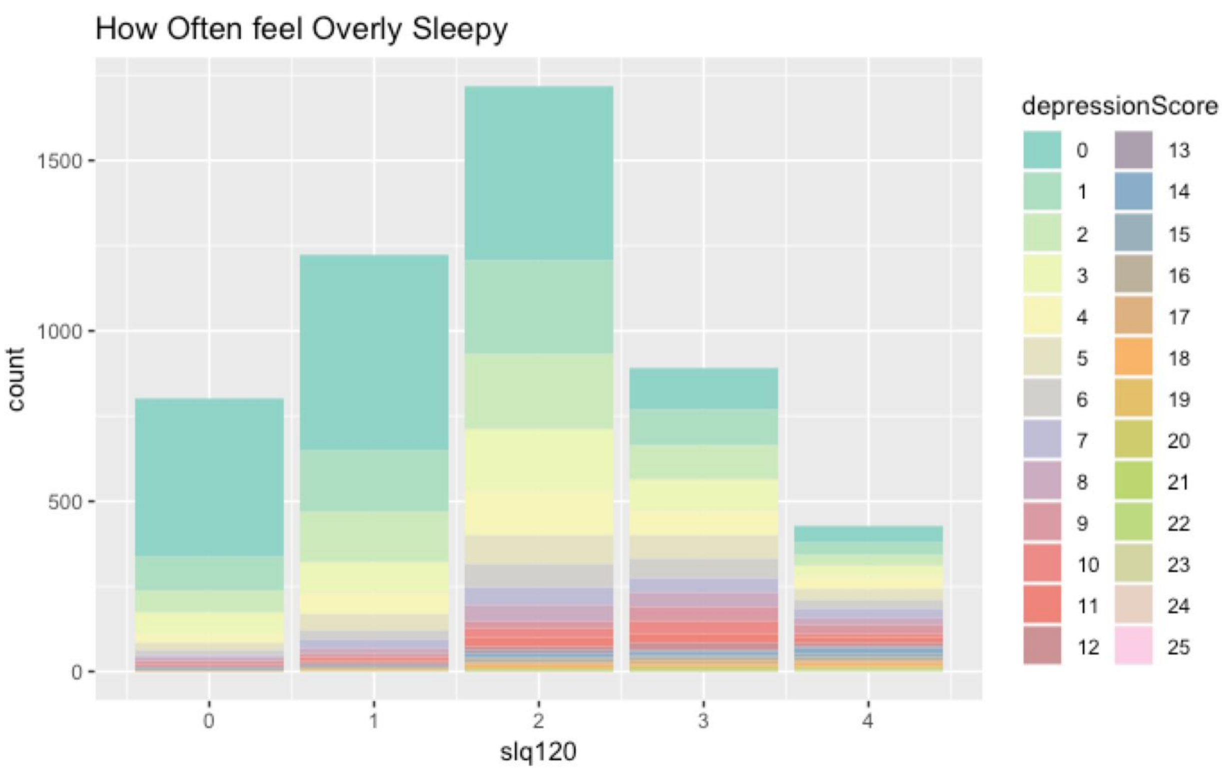


Figure 11: Trouble Sleeping Bar Chart

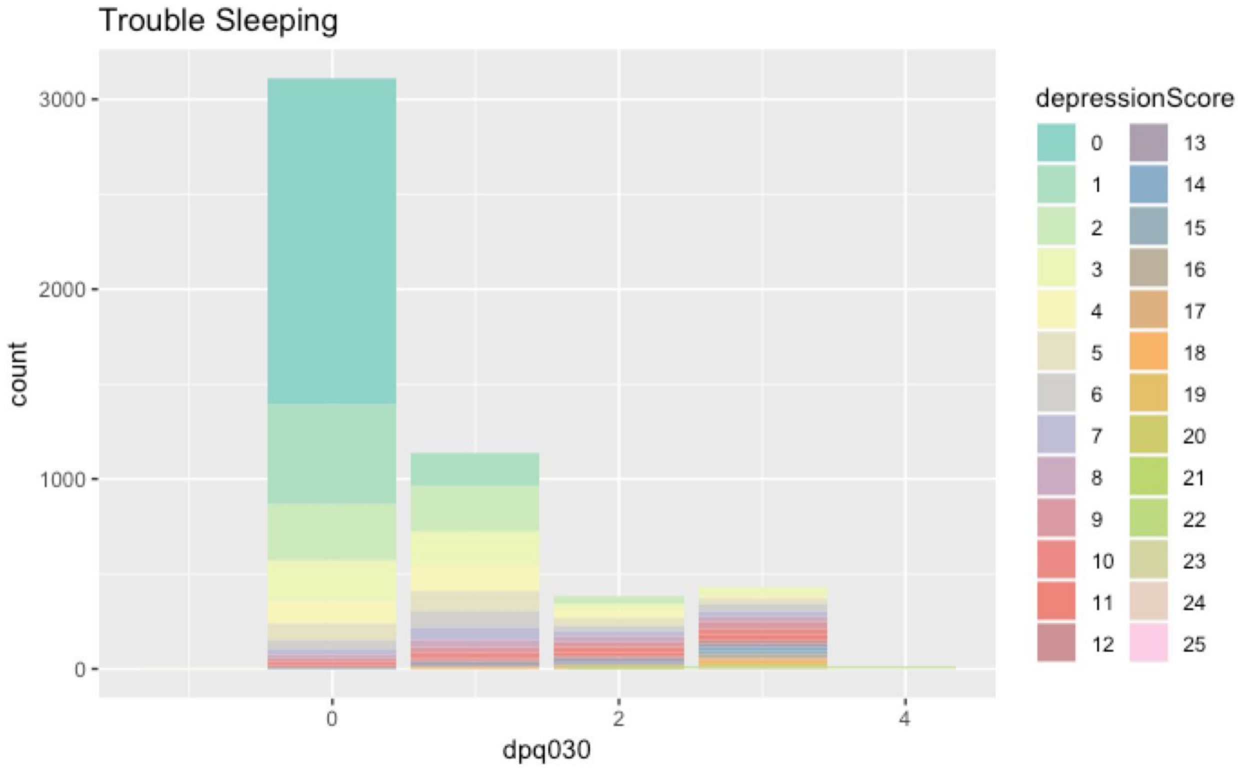


Figure 12: Feeling Down, Depressed, or Hopeless Bar Chart

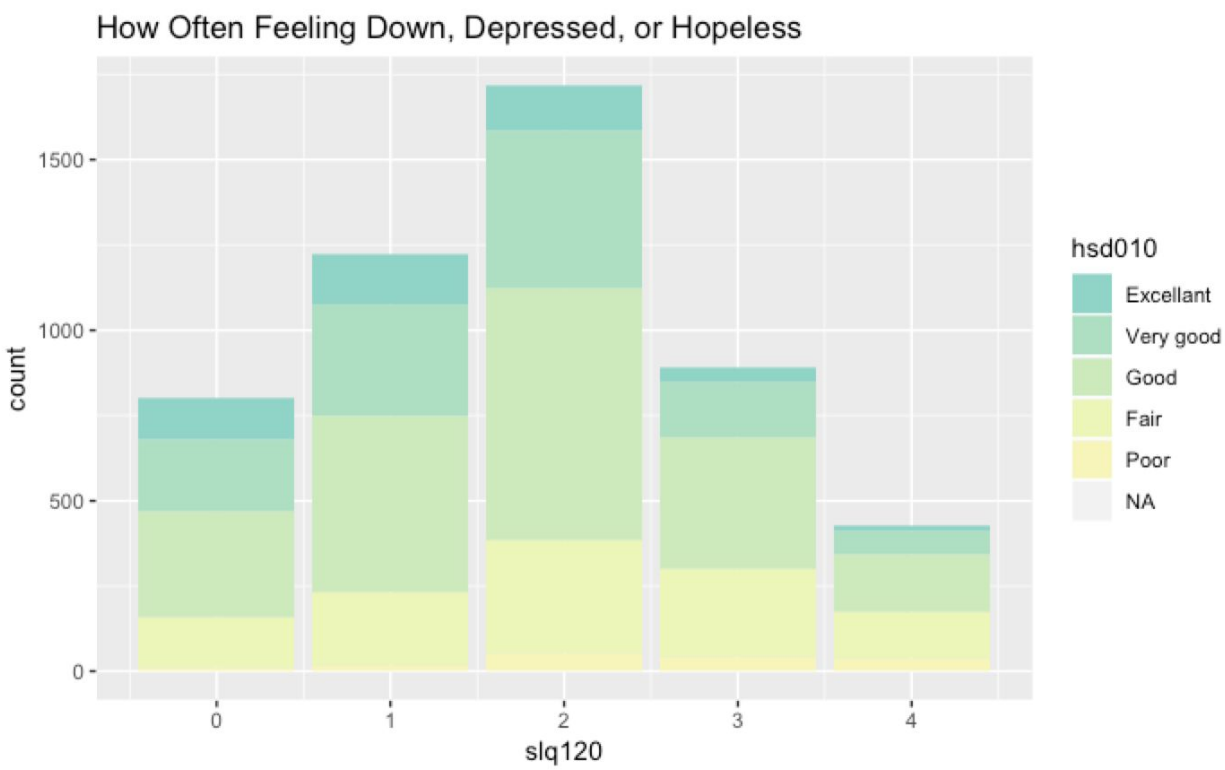


Figure 13: Correlation Plot

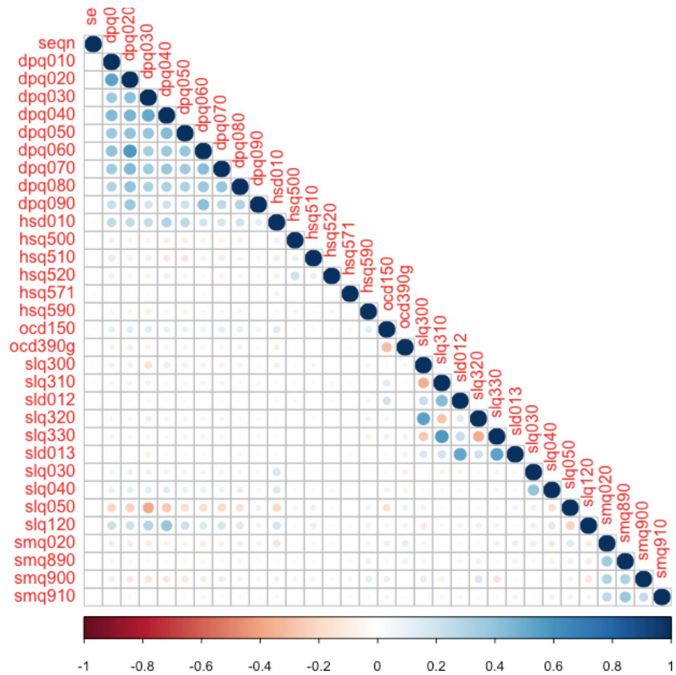


Figure 14: Visualizing missing data in

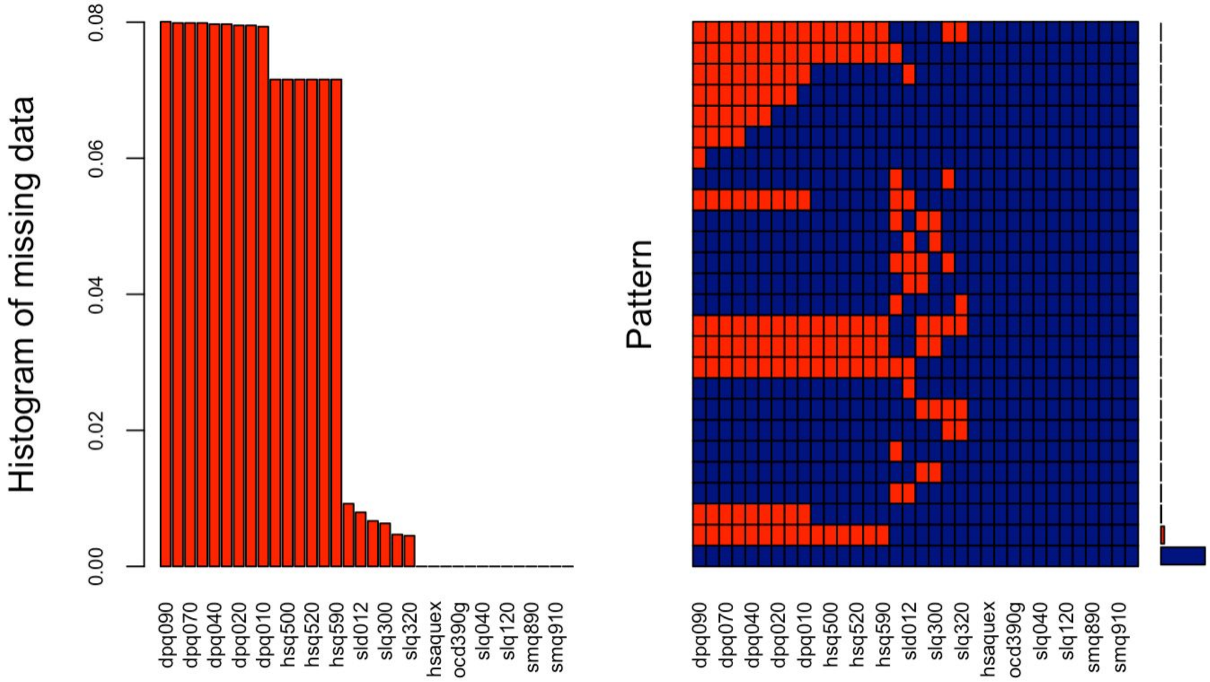


Figure 15: Statistical significance output from logistic regression model

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.8287394  0.4337947   4.216 2.49e-05 ***
hsd0102      0.5326858  0.1386995   3.841 0.000123 ***
hsd0103      0.8186117  0.1324898   6.179 6.46e-10 ***
hsd0104      1.5997623  0.1573878  10.164  < 2e-16 ***
hsd0105      3.8782062  0.7272327   5.333 9.67e-08 ***
hsq5002     -0.4320300  0.1178240  -3.667 0.000246 ***
hsq5102     -1.0627806  0.1989147  -5.343 9.15e-08 ***
hsq5202     -0.1897752  0.2322343  -0.817 0.413831    
hsq5712     -0.3845820  0.2010952  -1.912 0.055820 .  
ocd1502     -0.0991283  0.2571580  -0.385 0.699885    
ocd1503      0.5663047  0.2036985   2.780 0.005434 ** 
ocd1504      0.1419734  0.0858434   1.654 0.098155 .  
sld012      -0.0493770  0.0304012  -1.624 0.104338    
sld013       0.0205661  0.0274373   0.750 0.453515    
slq0301      0.1159531  0.1083937   1.070 0.284736    
slq0302     -0.0006311  0.1133569  -0.006 0.995558    
slq0303      0.0491379  0.1062812   0.462 0.643839    
smq0202     -0.0087796  0.0811835  -0.108 0.913880    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 16: Ranking of predictor variables by importance

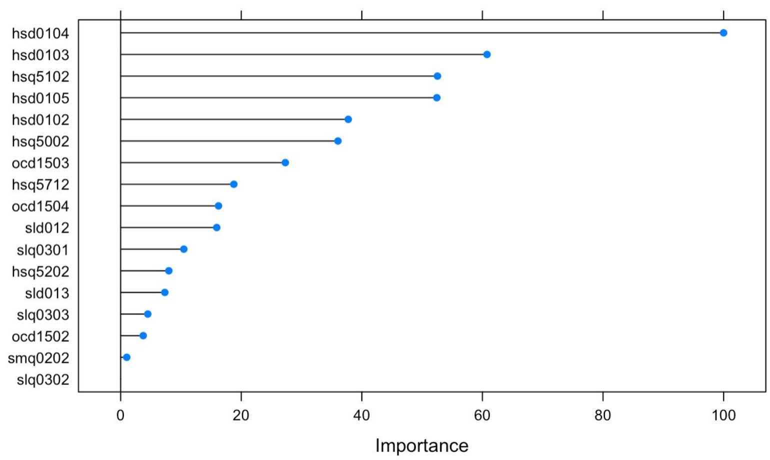
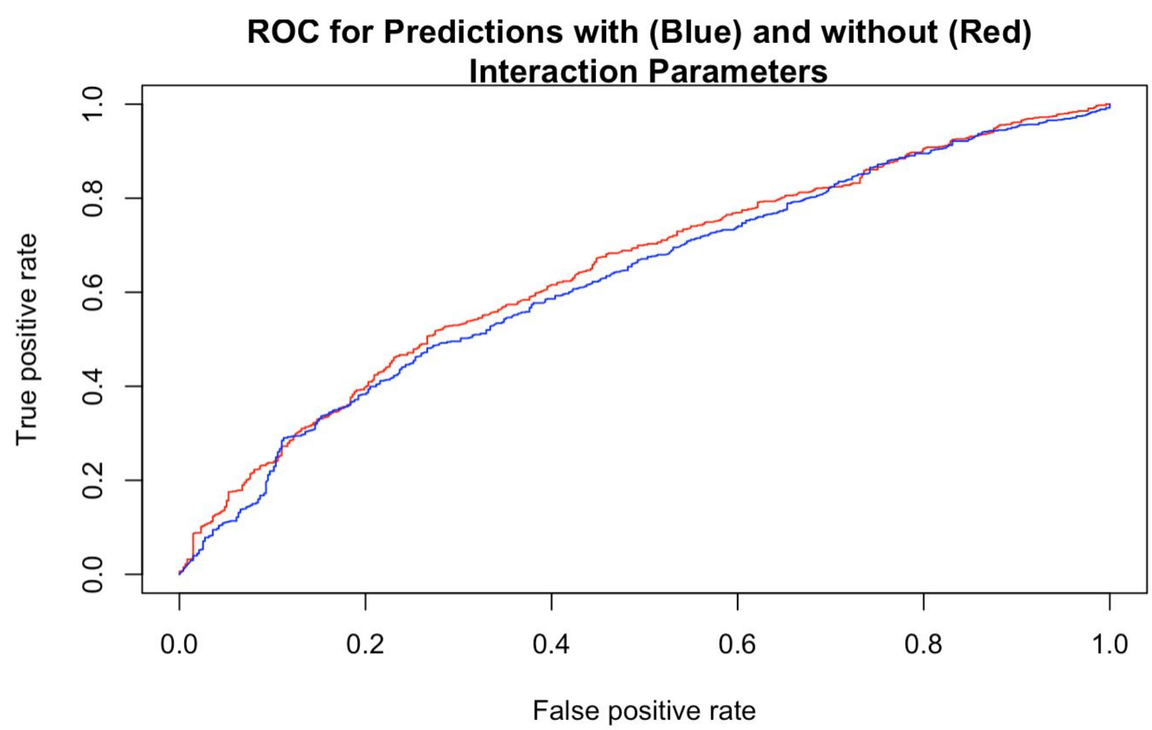


Figure 17: Ranking of interaction parameters by importance

	Overall <dbl>
`hsq5102:smq0202`	3.668845
smq0202	2.634119
`ocd1502:slq0302`	2.393827
`hsq5202:slq0303`	2.384584
hsq5202	2.183447
`hsq5002:ocd1502`	2.140832
`hsd0104:ocd1504`	2.114830
`hsq5202:sld012`	2.029916
`hsd0104:hsq5002`	1.905756
slq0303	1.882614

Figure 18: Comparing ROC AUC for two different logistic regression models



9b. Appendix B: R Statistical Programming Code

```r
########################################
########### data preparation ###########
########################################


# load packages
if(!require(pacman)){install.packages("pacman"); require(pacman)}
packages <- c("tidyverse", "sqldf", "broom", "Hmisc", "SASxport", "janitor",
"sqldf", "ggmosaic", "htmlwidgets", "wrapr", "plotly")
pacman::p load(char = packages)


# get files
surveyNames <- grep(pattern = ".XPT", dir(), ignore.case = TRUE, value = TRUE) %>%
       lapply(., function(x) gsub(".XPT", "", .)) %>% unlist %>% unique
files <- grep(pattern = ".XPT", dir(), ignore.case = TRUE, value = TRUE)


# read XPT data
for(i in 1:length(files)){
       assign(bquote(.(surveyNames[i])), value = read.xport(files[i]))
}


# combine togther
df <- DPQ J %>%
       inner join(HSQ J, by = c("SEQN", "SEQN")) %>%
       inner join(OCQ J, by = c("SEQN", "SEQN")) %>%
       inner join(SLQ J, by = c("SEQN", "SEQN")) %>%
       inner join(SMQ J, by = c("SEQN", "SEQN"))


# clean up field names
df <- df %>% janitor::clean names()


# meta data
metaData <- data.frame(fields = names(df),
                      missing = colSums(is.na(df)),
                      unique = sapply(df, function(x) length(unique(x)))) %>%
       dplyr::mutate(attribute = stringr::str_extract_all(fields, pattern =
"^[a-z]+") %>% unlist)


# get a list of variables that do not have significant number of missing values
fields <- metaData %>%
```

```
        dplyr::filter(missing < 1000) %>%
        .$fields %>%
        as.character


# subset fields
df <- df %>%
        dplyr::select(fields)


# output data for downstream analysis
write.csv(df, "df.csv", row.names = FALSE)


# run meta data again
metaDataUpdated <- data.frame(fields = names(df),
                              missing = colSums(is.na(df)),
                              unique = sapply(df, function(x) length(unique(x))),
                              type = sapply(df, class)) %>%
        dplyr::mutate(attribute = stringr::str_extract_all(fields, pattern =
"^[a-z]+") %>% unlist)


##################################
########### odds ratio ###########
##################################


# set variables
variables <- metaDataUpdated %>%
        dplyr::filter(attribute %in%
c(metaDataUpdated$attribute[metaDataUpdated$attribute %nin% c("seqn", "hsaquex",
"smaquex")] %>% unique)) %>%
        dplyr::select(fields) %>%
        .$fields %>%
        as.character


groupingVariable = "hsd010"


# transform data into long format
dfGather <- df %>%
        tidyr::gather(., fields, value, -seqn, -groupingVariable) %>%
        dplyr::inner_join(metaDataUpdated, by = c("fields", "fields"))


dfGathered <- dfGather[complete.cases(dfGather), ] %>%
```

```
        dplyr::select(attribute = fields,
                      value,
                      class = bquote(.(groupingVariable))) %>%
        dplyr::filter(attribute %in% variables) %>%
        dplyr::filter(class <=5) %>%
        dplyr::mutate(class = case when(class <=3 ~ 1,
                                        TRUE ~ 0))


# calculate odds ratio using sql
OR <- sqldf::sqldf("
        select attribute
        , value
        , interest group yes
        , control group yes
        , interest group no
        , control group no
        from (

                select x.attribute
                , x.value
                , x.interest group yes
                , x.control group yes
                , y.interest_group_total - x.interest_group_yes as
interest group no
                , y.control group total - x.control group yes as control group no
                from (
                        select attribute
                        , value
                        , sum(case when class = 1 then 1 else 0 end) as
interest group yes
                        , sum(case when class = 0 then 1 else 0 end) as
control group yes
                        from dfGathered
                        group by 1, 2
                ) x
                join (
                        select attribute
                        , sum(case when class = 1 then 1 else 0 end) as
interest group total
                        , sum(case when class = 0 then 1 else 0 end) as
control group total
                        from dfGathered
                        group by 1
                ) y on x.attribute = y.attribute
```

```
        ) z
        where interest group yes >= 25
        and interest group no >= 25
        and control group yes >= 25
        and control group no >= 25
        "
)


# Odds Ratio
ORplusCI <- OR %>%
        dplyr::mutate(OR = round((interest_group_yes / interest_group_no) /
(control group yes / control group no), 2))


# Lower 95% CI
Lower CI = with(ORplusCI,
               exp(1) ^(log(OR) -
                                 (1.96 * sqrt(1/interest group yes +
                                                     1/interest group no +
                                                     1/control group yes +
                                                     1/control group no))))


# Upper 95% CI
Upper CI = with(ORplusCI,
               exp(1) ^(log(OR) +
                                 (1.96 * sqrt(1/interest group yes +
                                                     1/interest group no +
                                                     1/control group yes +
                                                     1/control group no))))


# display only significant results and sort by OR
ORplusCI <- ORplusCI %>%
        dplyr::mutate(lower ci = round(Lower CI, 2),
                      upper ci = round(Upper CI, 2),
                      survey = stringr::str_extract_all(attribute, pattern =
"^[a-z]+") %>% unlist) %>%
        dplyr::filter(Upper CI >1 & Lower CI >1) %>%
        dplyr::select(survey, item = attribute, value, odds_ratio = OR, lower_ci,
upper ci) %>%
        arrange(desc(odds ratio))


#####################################
########### data analysis ###########
```

```
#####################################

### data transformation ###

# depression score
depressionScore <- df %>%
        dplyr::select at(vars(contains("seqn"), starts with("dpq"))) %>%
        dplyr::filter at(vars(starts with("dpq")), all vars(. <=3)) %>%
        dplyr::mutate(depressionScore = rowSums(select(., starts with("dpq")))) %>%
        dplyr::select(seqn, depressionScore)

# average daily hours of sleep
sleepHrs <- df %>%
        dplyr::select(seqn, sld012, sld013) %>%
        dplyr::filter(!is.na(sld012) & !is.na(sld013)) %>%
        dplyr::mutate(sleepHrs = (sld012 + sld013) / 2) %>%
        dplyr::select(seqn, sleepHrs)

# health status
healthStatus <- df %>%
        dplyr::select(seqn, hsd010) %>%
        dplyr::filter(hsd010 <= 5) %>%
        dplyr::mutate(healthStatus = dplyr::case when(hsd010 == 1 ~ "Excellent",
                                                  hsd010 == 2 ~ "Very good",
                                                  hsd010 == 3 ~ "Good",
                                                  hsd010 == 4 ~ "Fair",
                                                  hsd010 == 5 ~ "Poor") %>%
                          factor(., levels = c("Excellent", "Very good",
"Good", "Fair", "Poor"))) %>%
        dplyr::select(seqn, healthStatus)

# sleepiness
sleepiness <- df %>%
        dplyr::select(seqn, slq120) %>%
        dplyr::filter(slq120 <= 4) %>%
        dplyr::mutate(sleepiness = dplyr::case when(slq120 == 0 ~ "Never",
                                                slq120 == 1 ~ "Rarely",
                                                slq120 == 2 ~ "Sometimes",
                                                slq120 == 3 ~ "Often",
                                                slq120 == 4 ~ "Almost always")
%>%
```

```
                                 factor(., levels = c("Never", "Rarely", "Sometimes",
"Often", "Almost always"))) %>%
       dplyr::select(seqn, sleepiness)


#################################################################################
### data analysis using one-way subject anova


#################################################################################
# null hypothesis 1) there is no relationship between depression and sleepiness #


# get data
d1 <- merge(depressionScore, sleepiness)


# one-way subject anova
d1 aov <- aov(depressionScore ~ sleepiness, data = d1)
summary(d1 aov)


# post-hoc test
TukeyHSD(d1 aov)


##################################################################################
###########
# null hypothesis 2) there is no relationship between depression and physical
health status #


# get data
d2 <- merge(depressionScore, healthStatus)


# one-way subject anova
d2 aov <- aov(depressionScore ~ healthStatus, data = d2)
summary(d2 aov)


# post-hoc test
TukeyHSD(d2 aov)


#################################################################################
### data analysis using chisq.test
```

```
###################################################################################
###########
# null hypothesis 3) there is no relationship between sleepiness and physical
health status #


# get data
d3 <- merge(sleepiness, healthStatus)


# cross tab
round(prop.table(ftable(d3$sleepiness, d3$healthStatus), 2), 2)


# chisq.test
chisq.test(table(d3$healthStatus, d3$healthStatus))


##########################################
########### data visualization ###########
##########################################


# bar chart on odds ratio
OR bar <- ORplusCI %>%
       dplyr::filter(survey != "dpq") %>%
       dplyr::filter(odds ratio >1.5) %>%
       select(survey, item, value, odds ratio) %>%
       dplyr::mutate(survey = dplyr::case when(survey == "hsq" ~ "General Health",
                                               survey == "slq" ~ "Sleep
Behaviors",
                                               survey == "sld" ~ "Sleep
Behaviors",
                                               survey == "ocd" ~ "Occupation",
                                               survey == "smq" ~ "Smoke Habit"))
%>%
       arrange(item, value)

OR bar$label <- c('have not had a cold in the past 30 days',
                 'have not had a stomach or intestinal illness in the past 30
days',
                 'have not had a flu, pneumonia, or ear infections in the past 30
days',
                 'have donated blood in the past 12 months',
                 'work at a job or business',
                 'current job is the longest position that one has ever held',
                 'average hours of sleep during weekdays or workdays is 7.5 hrs',
                 'average hours of sleep during weekends is 7.5 hrs',
                 'average hours of sleep during weekends is 8.5 hrs',
```

```
                 'never snore while sleeping',
                 'never snort or stop breathing while asleep',
                 'never needed to seek professional help for trouble sleeping',
                 'rarely feel overly sleepy during day',
                 'ususal sleep time on weekdays or workdays is 22:30',
                 'ususal sleep time on weekdays or workdays is 23:30',
                 'ususal wake time on weekdays or workdays is 6:30',
                 'ususal wake time on weekdays or workdays is 7:30',
                 'ususal sleep time on weekends is 23:30',
                 'smoke less than 100 cigarettes in lifetime')


OR bar chart <- OR bar %>%
       ggplot(., aes(reorder(label, odds ratio), odds ratio)) +
       geom bar(stat = "identity", aes(fill = survey)) +
       xlab("") +
       ylab("Odds Ratio") +
       coord flip() +
       theme minimal() +
       geom text(aes(x = label, y = odds ratio, label = round(odds ratio, 2)),
                 position = position dodge(width = 1), hjust = 1.5,
                 col = "white", fontface = "bold", size = 3) +
       ggtitle("Features that are significantly important pertaining to healthy
individuals")


# distribution of hours of sleep by health status
sleep by healthStatus <- merge(sleepHrs, healthStatus) %>%
       ggplot(., aes(sleepHrs, fill = healthStatus)) +
       geom density() +
       geom_vline(data = aggregate(sleepHrs ~ healthStatus, data = merge(sleepHrs,
healthStatus), FUN = median),
                  aes(xintercept = sleepHrs,
                      color = healthStatus),
                  linetype = "dashed") +
       xlab("Hours of Sleep") +
       theme classic() +
       ggtitle("Hours of sleep varied by health condition")


# boxplot on depression varied by sleepiness
depression by sleepiness <- d1 %>%
       ggplot(., aes(sleepiness, depressionScore, fill = sleepiness)) +
       geom boxplot() +
       theme bw() +
```

```
        xlab("How often feel overly sleepy during day?") +
        ylab("Depression Score") +
        theme(legend.position = "none") +
        ggtitle("Depression varied by sleepiness")


# boxplot on depression varied by health status
depression_by_healthStatus <- d2 %>%
        ggplot(., aes(healthStatus, depressionScore, fill = healthStatus)) +
        geom_boxplot() +
        theme_minimal() +
        xlab("General Health Condition") +
        ylab("Depression Score") +
        theme(legend.position = "none") +
        ggtitle("Depression varied by health condition")


# mosaic plot - sleepiness x health status
sleepiness_by_healthStatus <- ggplot(data = d3) +
        geom_mosaic(aes(x = product(sleepiness, healthStatus), fill = sleepiness),
na.rm = TRUE) +
        labs(x = "General Health Condition", y = "How often feel overly sleepy
during day?") +
        theme_minimal() +
        ggtitle("Sleepiness during day varied by health condition")


################# save output in png and html #################
# chart objects
chartObjs <- list(OR_bar_chart, sleep_by_healthStatus, depression_by_sleepiness,
depression_by_healthStatus, sleepiness_by_healthStatus)


names(chartObjs) <- wrapr::qc(OR_bar_chart, sleep_by_healthStatus,
depression_by_sleepiness, depression_by_healthStatus, sleepiness_by_healthStatus)


# save as png
sapply(1:length(chartObjs), function(x) ggsave(filename =
paste0(names(chartObjs[x]), ".png"),
                                              plot = chartObjs[[x]])) %>%
        invisible()


# save as html
sapply(1:length(chartObjs), function(x) htmlwidgets::saveWidget(widget =
as_widget(chartObjs[[x]] %>% ggplotly),
                                                               file =
paste0(names(chartObjs[x]), ".html"))) %>%
        invisible()
```