Depression and Regression: A Deep-Dive into the Impacts of Health Insurance Status and Other Covariates

Karina Bellavia, Asja Hamzic, Peyton Smith

11 December 2023

ABSTRACT

BACKGROUND

In the United States, depression and lack of health insurance are major public health concerns. Health insurance is a social determinant of access to care, with access to such care having an association with depression.

OBJECTIVE

The primary objective of this study is to determine if having health insurance has an association with the odds of depression among individuals in the United States using NHANES data, adjusting for age, gender, race, citizenship status, education status, and income level. Our secondary objective of this study is to see if age, gender, race, citizenship, education status, or income level has a statistically significant impact on having insurance.

METHODS

Our analyses were completed in R Studio version 4.3.1. We used simple logistic regression, multivariate logistic regression, ordinal logistic regression, and multinomial regression techniques to answer our primary question. A multivariate logistic regression model was used to answer our secondary question.

RESULTS

We found that the simple, multivariate, ordinal, and multinomial logistic models supported that having health insurance is associated with a decreased risk of depression. Additionally, we found that access to health care varies within different demographics of the population. Specifically, the odds of being insured increases with education and income level, and the odds of being insured decreases if one is not a citizen of the United States or refused to answer the NHANES question.

DISCUSSION

Future research can expand upon means by which to address disparities in access to healthcare based on health insurance status, impacting the prevalence and effects of depression within the United States population. Overall, the key to a healthier United States population is health care access for all.

INTRODUCTION

Depression is one of the most common mental health conditions in the United States and its prevalence is reaching an all-time high. The National Institute of Mental Health reports that 21 million adults experienced at least one major depressive episode in 2021, with the prevalence of such episodes tending to be higher in females than males, those aged 18 to 25, and multiracial individuals.[1] Common risk factors for depression include family history, trauma or stress, chronic, fatal, or severe medical conditions such as cancer or stroke, and substance use.[2] The severity of depression an individual has also varies. Using a depression scoring index, one can have: no, mild, moderate, moderately severe, or severe depression. However, it should be noted that depression is not necessarily ordinal. While it can progress or regress, a person who experiences depression may not be initially diagnosed with mild depression; sudden grief can make an otherwise non-depressed person be immediately diagnosed with severe depression.

In the United States, health insurance is vital to receiving proper medical care and affording routine or emergency doctor's visits. The Centers for Disease Control and Prevention (CDC) reports that 27.6 million, or 8.4%, of all Americans did not have health insurance in 2022, which is a decrease from 10.3% in 2021.[3] Additionally, since the passage of the Affordable Care Act, the amount of people who have health insurance coverage in the United States has increased drastically, specifically among those aged 19 to 25.[4] Despite the increase in health insurance subscriptions, the amount of Americans who remain uninsured is significant. This lack of health insurance corresponds to a lack of access to care, which can also contribute to depression.[5,6,7]

People experiencing depression may experience sadness, memory loss, physical pain, poor academic or work performance, fatigue, loss of appetite, or suicidal thoughts[8]. Similarly, lack of health insurance can negatively impact both physical and mental health, with the added stress of being unable to access or afford treatment for illnesses and injuries making an otherwise treatable condition worse or prolonged, which, consequently, can also be a risk factor for depression. Both depression and lack of health insurance are barriers to healthcare access. Previous studies have found that depression is associated with a lack of insurance and access to affordable healthcare, though we did not find existing analyses that compared the odds of depression between people with various types (public and private) of health insurance.[9,10]

The National Health and Nutrition Examination Survey (NHANES) is a program through the National Center for Health Statistics. The goal of NHANES is to assess the health and nutritional status of those living in the United States. This program is unique, as data is collected through both interviews and physical examinations across all demographic groups with the goal of representing the entire population[11]. NHANES includes data on characteristics of individuals, including household income, race and ethnicity, income, education, health insurance status, age, and gender.

Through our analysis of NHANES results from 2013 to 2018, we aim to determine the odds of depression among American individuals who have health insurance compared to individuals without health insurance. Additionally, we will assess the relationship between health insurance status and the severity of depression, as well as examine the characteristics of the individuals with and without health insurance. The motivation for conducting this analysis is to bring awareness to mental health and health insurance disparities while contributing to the ongoing discourse of the social and demographic factors that contribute to both of these major public health issues.

RESEARCH AND ANALYSIS METHODS

In order to run regression models on the NHANES dataset, we used R Studio, version 4.3.1. The following R packages were installed to complete our analysis: gtools, splitstackshape, ResourceSelection, ROCR, jtools, Publish, nnet, ggplot2, knitr, dplyr, and nhanesA. These packages allowed us to preprocess the data, run our regression models, and generate visualizations and other forms of expressing our findings.

We began first with a simple logistic regression model, examining the relationship between depression, a binary variable equaling zero for an individual with no depression and equaling one for an individual with depression ranging from mild to severe, and health insurance, a binary variable equaling zero for an individual with no health insurance and equaling one for an individual with health insurance. This model, though neglecting other predictors of depression level, allowed us to visualize the simple relationship between our two main variables of interest: depression and health insurance status. To complement this model, we also ran a multivariate logistic regression with the same outcome and predictor variable as the simple logistic regression model, but adjusting for the covariates age, income level, specifically annual household income, gender, race, citizenship status, and education status. Based on domain expertise and previous literature on the topic, we chose to adjust for these covariates in our model in order to determine the extent of the statistically significant relationship between each covariate and the outcome.[1, 6, 9] The listed covariates all satisfy the classical definition of a confounder, meaning they are each associated with health insurance status, they are associated with depression in the absence of health insurance status, and they are not a downstream consequence of health insurance status. We therefore included these covariates in numerous models, as it is crucial to adjust for each of them given their status as confounders. The categories within each covariate used in the multivariate logistic regression model, excluding the health insurance predictor, can be found in Table 1, and the first value for each covariate listed is the reference category for that variable.

To test for collinearity between covariates, we used a variance inflation factor (VIF) test. VIF describes the amount that a certain covariate explains, or is correlated with, another covariate in the model. If a VIF is greater than 2 for a specific covariate, then it raises concern for multicollinearity in a model. After testing generating VIFs for our model, we found that each covariate had a VIF less than 2, thus we are not worried that there is multicollinearity in our model (Table 2).

In order to examine the relationship between individual levels of depression and health insurance status, we constructed two multinomial logistic regression models. The first model, which had depression as the outcome, used having no depression as a reference category, comparing each other level of depression (mild, moderate, moderate-severe and severe) with the reference category and using only the covariate of health insurance status. By generating this model, we are able to better grasp how having health insurance is related to varying levels of the severity of depression. In the second multinomial model, we divided the dataset into three groups: the first group consisted of individuals with no depression, the second group consisted of individuals with mild or moderate depression, and the third group consisted of individuals with moderate-severe or severe depression. We created these divisions after noticing that there were very few individuals with more severe levels of depression, therefore generating these groups allowed our model results to have more power. This second multinomial logistic regression model uses these three groupings, with the no depression group being the reference category, and has the covariate health insurance. Although the multinomial logistic regression models with the outcome of depression and covariate health insurance provide us insight into the relationship between the two variables, there are other factors that impact depression level that are not captured in the health insurance status predictor. To preserve parsimony and for ease of interpretation, we only included the covariates age, income level, gender, citizenship status, race, and education level in the multivariate logistic regression model.

An ordinal logistic regression model is appropriate in this scenario if the proportional odds assumption is met. If this assumption is met, that implies that there may be a natural ordering to the levels of our outcome of depression. The proportional odds assumption was verified using the generalized ordinal model. We obtained a p-value of 0.631, meaning we fail to reject our null hypothesis, concluding the proportional odds model is adequate.

With the proportional odds assumption being met, we ran two ordinal logistic regression models. The first model included the covariates health insurance, defined by having health insurance or not, age, income level, education level, citizenship status, gender, and race, and had the outcome of depression, with the cumulative probability of having no depression being compared to the cumulative probability of having some level of depression (mild, moderate, moderate-severe, or severe). The second model included the same covariates as the first model, but the outcome of depression compared the cumulative probability of having no depression or mild depression to the cumulative probability of having either moderate, mild-moderate, or severe depression. We chose to generate the second model for a similar reason as why we ran another multinomial logistic regression model above, as there were very few individuals with more severe levels of depression.

To answer our secondary question, we ran a multivariate logistic regression model. The outcome was health insurance status, coded as either having health insurance or not, and the covariates we included in the model were age, income level, education level, gender, citizenship status, and race. This model allowed us to more closely examine the relationship between each of the covariates and health insurance without considering their impact on depression.

In order to evaluate the relative goodness of fit of each of our models, the Akaike Information Criteria (AIC) was used. This was done by combining measures of model fit and measures of model complexity. We will calculate this value for each model, though this will be combined with domain expertise and subject matter knowledge in evaluating the usefulness and application of each of the models discussed.

In our dataset, we are missing 14,679 people who did not answer the depression screener or have missing data elsewhere. We removed 1,664 people who chose “$20,000 and over” and 376 people who chose “$20,000 and below” as their income level because we found it was redundant when accounting for the other income level shown in Table 1. Furthermore, we removed 10,609 children under the age of 18, all of whom were not permitted to take the depression screener, and 1,642 adults who did not take the depression screener. Lastly, we removed 4 people who did not provide their citizenship status and 335 people who did not provide their education level. Based on the previous information, we think that the missing data would be classified as MCAR, therefore we are not concerned that our data will need special attention in terms of addressing missingness. It should be noted that among those who did answer the depression screener questions, there is no pattern of missingness among the data. After removing such individuals above, our dataset contained 13,382 people.

FINDINGS AND ANALYSIS

PRIMARY QUESTION

Simple Logistic Regression

According to this data, the odds of having depression amongst people who have health insurance are 0.81 times the odds of having depression amongst people who do not have health insurance, on average (95% CI = [0.73, 0.89]; p-value = 2.16e-05).

$$(1)\ logit(\widehat{p}_{depression}) = -0.97106 - 0.21662X_{Insurance}$$

The AIC value for this specific model is 14,992.78 (Table 4).

Multivariate Logistic Regression

According to this data, the odds of having depression amongst people who have health insurance are 0.90 times the odds of having depression amongst people who do not have health insurance, on average, holding all other covariates constant (95% CI = [0.79, 1.00]; p-value = 0.052).

$$\begin{aligned}(2)\ logit(\widehat{p}_{depression}) = &-0.20 - 0.10X_{Insurance} - 0.00023X_{Age} - 0.51X_{Gender} \\ &+ 0.04\,X_{Non-Hispanic\ Black} + 0.23\,X_{Non-Hispanic\ White} + 0.11X_{Other\ Hispanic} \\ &+ 0.01X_{Multirace} + 1.74X_{Ctzn\ -\ Don't\ Know} - 0.24\,X_{Not\ a\ Ctzn} - 0.45\,X_{Ctzn-Refused} \\ &- 0.62\,X_{College\ Grad\ or\ Above} + 2.06\,X_{Educ\ -\ Don't\ Know} - 0.25X_{HS\ Grad} - 0.08\,X_{<\ than\ 9th\ grade} \\ &- 0.29\,X_{Some\ college/AA\ deg.} - 0.40\,X_{Income\ -\ 25k\ to\ 55k} - 0.49\,X_{Income\ -\ 55k\ to\ 75k} \\ &- 0.83\,X_{Income\ -\ 75k\ and\ up}\end{aligned}$$

The AIC value for this model is 14,202.5 (Table 4). Given this is the lowest AIC value, this suggests that this model is the best fit for our data, though this value alone should not be the only information taken into account when choosing the best model for our data.

Ordinal Logistic Regression

For those with health insurance, the cumulative odds of depression in the category greater than or equal to k versus those below the category k are 0.82 times the cumulative odds for those without health insurance, on average and according to this data (95% CI =[0.73, 0.90]; p-value = 3.24e-0.5), where k is the depression category (k=1 corresponds to no depression, k=2 corresponds to mild depression, k=3 corresponds to moderate depression, k=4 corresponds to moderate-severe depression, and k=5 correspond to severe depression).

$$(3)\ log(\frac{\widehat{P}(Y\geq 2)}{\widehat{P}(Y<2)}) = -0.98 - 0.21X_{health\ insurance}$$

$$(4)\ log(\frac{\widehat{P}(Y\geq 3)}{\widehat{P}(Y<3)}) = -1.26 - 0.21X_{health\ insurance}$$

$$(5)\ log(\frac{\widehat{P}(Y\geq 4)}{\widehat{P}(Y<4)}) = -2.68 - 0.21X_{health\ insurance}$$

$$(6)\ log(\frac{\widehat{P}(Y\geq 5)}{\widehat{P}(Y<5)}) = -4.43 - 0.21X_{health\ insurance}$$

The AIC value for this model is 21,774.81 (Table 4).

Multinomial Logistic Regression (Full)

The reference category for each of these models is no depression. The values for each category, k=1, 2, 3, 4, 5, refer to the same depression level as in the ordinal logistic model above.

According to this data, the relative risk of having mild depression versus no depression for those with health insurance is 0.75 times that of those without health insurance, on average (95% CI = [0.63, 0.92]).

$$(7)\ log(\frac{\widehat{P_2}}{\widehat{P_1}}) = -2.52 - 0.28X_{health\ insurance}$$

According to this data, the relative risk of having moderate depression versus no depression for those with health insurance is 0.84 times that of those without health insurance, on average (CI 95% = [0.74, 0.95]).

$$(8)\ log(\frac{\widehat{P_3}}{\widehat{P_1}}) = -1.57 - 0.17X_{health\ insurance}$$

According to this data, the relative risk of having moderate-severe depression versus no depression for those with health insurance is 0.77 times that of those without health insurance, on average (95% CI = [0.63, 0.95]).

$$(9)\ log(\frac{\widehat{P_4}}{\widehat{P_1}}) = -2.63 - 0.25X_{health\ insurance}$$

According to this data, the relative risk of having severe depression versus no depression for those with health insurance is 0.72 times that of those without health insurance, on average (95% CI = [0.48, 1.1]).

$$(10)\quad log(\frac{\widehat{P_5}}{\widehat{P_1}}) = -4.07 - 0.33X_{health\ insurance}$$

In comparison to other models, the AIC value for this model is 21,779.08 (Table 4).

Multinomial Logistic Regression (Simplified)

The relative risk of having mild or moderate depression versus no depression for those with health insurance is 0.82 times that for those without health insurance, on average and according to this data (95% CI = [0.73, 0.91]).

$$(11)\quad log(\frac{\widehat{P_2}}{\widehat{P_1}}) = -1.24 - 0.20X_{health\ insurance}$$

The relative risk of having moderate-severe or severe depression versus no depression for those with health insurance is 0.77 times that for those without health insurance, on average and according to this data (95% CI = [0.64, 0.92]).

$$(12)\quad log(\frac{\widehat{P_3}}{\widehat{P_1}}) = -2.42 - 0.27X_{health\ insurance}$$

The AIC value for this model is 18,209.19 (Table 4).

SECONDARY QUESTION

Multivariate Logistic Regression

We found that the odds of having health insurance associated with a one-year increase in age are 1.04 times the odds of not having health insurance for those at the current age, on average according to these data, and holding all other covariates constant (95% CI = [1.04, 1.05]; p-value = < 2e-16).

We also found that the odds of having health insurance if an individual is male are 0.704 times the odds of having health insurance if an individual is female, on average according to these data, and holding all other covariates constant (95% CI = [0.64, 0.78]; p-value = 2.90e-11).

To expand further, we found that the odds of having health insurance for non-hispanic white individuals are 1.39 times the odds of having health insurance if the individual is of another race, on average according to these data, and holding all other covariates constant (95% CI = [1.17, 1.65]; p-value = 9.28e-05). Furthermore, we also found that the odds of having health insurance if a multiracial individual are 1.77 times the odds of having health insurance if the individual is of another race, on average according to these data, and holding all other covariates constant (95% CI = [1.45, 2.14]; p-value = 6.02e-09).

We also found that the odds of having health insurance if not a citizen of the United States are 0.28 times the odds of having health insurance if an individual is a citizen of the U.S. or refused to answer the question on citizenship status, on average according to these data, and holding all other covariates constant (95% CI = [0.24, 0.32]; p-value = < 2e-16). We also found that the odds of having health insurance if an individual refused to share their citizenship status are 0.15 times the odds of having health insurance if the individual is either a citizen or not a citizen of the U.S., on average according to these data, and holding all other covariates constant (95% CI = [0.04, 0.50]; p-value = 0.003).

We also found that the odds of having health insurance if an individual is at least a college graduate are 3.25 times the odds of having health insurance if the individual is not at least a college graduate, on average according to these data, and holding all other covariates constant (95% CI = [2.66, 3.97]; p-value = < 2e-16). We also found that the odds of having health insurance if an individual is a high school graduate are 1.30 times the odds of having health insurance for an individual with any other education level, on average according to these data, and holding all other covariates constant (95% CI = [1.11, 1.54]; p-value = 0.001). We also found that the odds of having health insurance if an individual completed some college are 1.60 times the odds of having health insurance for an individual with any other education level, on average according to these data, and holding all other covariates constant (95% CI = [1.36, 1.86]; p-value = 6.92e-09).

Furthermore, we found that the odds of having health insurance if an individual has an income between $25,000 and $54,999 are 1.24 times the odds of having health insurance if an individual is at any other income level, on average according to these data, and holding all other covariates constant (95% CI = [1.09, 1.39]; p-value = 0.00051). We also found that the odds of having health insurance if an individual has an income between $55,000 and $74,999 are 2.38 times the odds of having health insurance if an individual is at any other income level, on average according to these data, and holding all other covariates constant (95% CI = [1.97, 2.86]; p-value = 2e-16). Lastly, we found that the odds of having health insurance if an individual has an income greater than $75,000 are 3.89 times the odds of having health insurance if an individual is at any other income level, on average according to these data, and holding all other covariates constant (95% CI = [3.29, 4.62]; p-value = 2e-16).

$$(13) \quad logit(\widehat{p}_{health\ insurance}) =- 1.07 + 0.04X_{Age} - 0.35X_{Gender} + 0.08X_{Non-Hispanic\ Black} + 0.33\,X_{Non-Hispanic\ White} + 0.12\,X_{Other\ Hispanic} + 0.57\,X_{Multirace} - 1.36\,X_{Ctzn\ -\ Don't\ Know} - 1.28\,X_{Not\ a\ ctzn} - 1.90\,X_{Ctzn-Refused} + 1.18X_{College\ Grad\ or\ Above} + 11.68X_{Educ\ -\ Don't\ Know} - 0.27X_{HS\ Grad} + 0.11X_{<\ than\ grade\ 9} + 0.47\,X_{Some\ college/AA\ deg.} + 0.21X_{Income\ -\ 25k\ to\ 55k} + 0..87\,X_{Income\ -\ 55k\ to\ 75k} + 1.36\,X_{Income\ -\ 75k\ and\ up}$$

We found that the AIC value of this model is 9,945.178 (Table 4).

DISCUSSION

Result Implications

The purpose of our study was to see if having health insurance impacts the odds of depression for American adults. We found that the simple, multivariate, ordinal, and multinomial logistic models supported that having health insurance is associated with a decrease in the odds of depression, which follows results from previous studies that only looked at public insurance such as Medicaid. This indicates that having health insurance is protective against depression.

Additionally, we wanted to explore which demographics are most likely to have health insurance. First, we found that American adults are more likely to obtain health insurance as they age (OR = 1.04). This is likely due to older adults being able to be insured through Medicare[12]. Additionally, we found that males are less likely to have health insurance than females (OR = 0.704). This may be because women are more likely to be eligible for public insurance because of their lower average income level[14]. We also found that caucasians are more likely to have health insurance (OR = 1.39). This is likely because caucasian people are more likely to have a higher education level than individuals of most other races, and therefore are more likely to have higher paying jobs that provide them with the income to afford health insurance. As a result, caucasians people have a higher health insurance rate than their counterparts[13]. Similarly, we found that those who are multi-race are even more likely to have health insurance than other races (including caucasians) (OR = 1.77). There is limited literature providing background on why this is the case, but we may have obtained this result because of the small number of multiracial individuals in our sample. Furthermore, we found that people are less likely to have health insurance if they are not a citizen of the United States (OR = 0.28). This is likely because non-U.S. citizens typically occupy paid positions with lower wages, which tend to have lower rates of employer-based coverage[14]. We also found that people are even less likely to have health insurance if they refused to share their citizenship status on the survey (OR = 0.15). We believe many people refused to provide their citizenship status because it is likely that such individuals were not citizens of the United States. We also found that the odds of having health insurance increased with each increase in the level of education attained (ORs = 1.30 (HS level), 1.60 (some college), 3.25 (college)). This is likely because positions in the workforce offering a health insurance plan are typically associated with requiring at least a college degree in terms of education level[14]. Lastly, we found that the odds of having health insurance increases with each incremental increase in income level (ORs = 1.24 ($25k to $55k), 2.38 ($55k to $75k), 3.89 ($75k and up)). This is likely due to the fact that an individual with a greater income is often in higher paid careers with health insurance packages, as well as the fact that individuals with more income can afford health insurance[13].

Limitations

Limitations of our project include complex groupings of race and ethnicity that made it difficult to analyze due to the fact that people may fit into multiple groups and the fact that a different coefficient is needed for every racial or ethnic identity added complexity to models (2) and (13). Combining or removing groups may leave out crucial information, considering that multiracial individuals are more at risk for depression, grouping this with other races, as was done in the "Other Race - Including Multi-Racial" variable, may minimize the effect we saw.

The nature of NHANES data is also a limiting factor, as it was all self-reported by individuals taking the survey. None of our data is able to be verified via medical or financial records, so we must rely on and trust in the accuracy of each survey participant, with no way to validate their responses. This demonstrates that there is likely some response bias embedded within the results we obtained.

We reached out to three subject-matter experts, two health economists, Dr. Benjamin Sommers and Dr. Katherine Baicker, and the Director of the Center for Depression, Anxiety, and Stress Research, Dr. Diego Pizzagalli, but were unable to meet with them to dive deeper into our topics. Therefore, the background information on depression, insurance, and the other covariates of interest were gleaned from literature and our own domain knowledge. The knowledge and experience from these individuals would have enhanced our project further and may have altered certain aspects of our approaches.

In creating bins, or categories, for the covariate of income level, we found it important to note the impact of the chosen bins, as there may be individuals who fall into a certain income level category, but may have characteristics that better identify them with the individuals in the income level category above or below them. We followed federal recommendations for what is considered the poverty line for households of 2 and 3 people, which are both under $25,000 annually[14], as the reference group. This is a limitation of using categorical data in comparison to continuous data, as we lost precision in our estimation when we generated these categories, and a decision that would have benefited greatly with a discussion with one of our chosen health economists.

Lastly, evaluating goodness of fit for our models was severely limited in that we could not employ the Hosmer-Lemeshow Test due to negative degrees of freedom. A chi-square test and ROC curve were also unsuccessful. We relied mostly on AIC and VIF's, but we recognize these methods are more useful for model selection than goodness of fit.

Future Extensions

While our project filled a gap in research by assessing the odds of depression in people by combining those with public and private insurance (that is, having any insurance at all compared to none), we were time-constrained in evaluating the odds of depression in people with different types of insurance. We hope that our work inspires other statisticians to compare the outcome of depression between people with private, public, and other or multiple types of insurance.

Further research could expand upon our project by conducting a cost-benefit analysis of health insurance and its coverage of treatment for depression. This could be taken in various directions, evaluating the impacts of insurance covering treatment for depression, for example, and would add greater depth to the findings from our project.

Conclusion

The results of our project are worrisome in that they indicate the potential for a large portion of the United States' population to be at risk for depression. Depression is a substantial public health crisis as its prevalence in the United States continues to rise[15], which, in recent times, can largely be attributed to the after-effects of the COVID-19 pandemic, as well as high-stress lifestyles and the de-stigmatization of mental health. Our results also show that if an individual is more educated, wealthy, a United States citizen, and caucasian, they will tend to have health insurance and consequently be less at risk for depression than other demographic groups. We therefore must consider that these "other" demographic groups also tend to include marginalized or vulnerable populations, and therefore already at increased risk of depression[16] by virtue of circumstances they cannot control. This, coupled with fewer of these individuals having health insurance, further exacerbates the already growing crisis.

This is a multifaceted issue that cannot be solved through passing a single law. Although there are current initiatives in place to improve mental health, as companies and universities add counseling resources and implement mental health days and Medicaid continues to expand following the Affordable Care Act[4], millions of Americans are still not being reached. We hope the results from our paper encourage collaboration between public health officials and policymakers on this issue.

Finally, though the number of people without health insurance decreased slightly from 2021 to 2022, this number does not consider people who are “underinsured”, which can pose similar risks. The key to a healthier population is access to health insurance for all.

TABLES AND FIGURES

Age (RIDAGEYR)	**Annual Household Income (INDHHIN2)**	**Gender (RIAGENDR)**	**Race (RIDRETH)**	**Citizenship Status (DMDCITZN)**	**Education Level (DMDEDUC2)**
Continuous 0+	$0 to $24,999	Male	Mexican American	Citizen by birth or naturalization	Less than 9th grade
-	$25,000 to $54,999	Female	Other Hispanic	Not a citizen of the US	9-11th grade (Includes 12th grade with no diploma)
-	$55,000 to $74,999	- -	Non-Hispanic White	Refused	High school graduate/GED or equivalent
-	$75,000 and Over	-	Non-Hispanic Black	Don't Know	Some college or AA degree
-	Refused	-	Other Race - Including Multi-Racial	-	College graduate or above
-	Don't know	-	-	-	Refused
-	-	-	-	-	Don't Know

Table 1: Levels of Covariates (Excluding Health Insurance) For Multivariate Logistic Regression Model

	GVIF	Df	GFIV^(1/(2*Df))
Health Insurance	1.217019	1	1.103186
Age	1.151932	1	1.073281
Gender	1.014207	1	1.007078
Race	1.432647	4	1.045966
Citizenship	1.380978	3	1.055272
Education	1.530470	5	1.043476
Income	1.249177	3	1.037777

Table 2:Variance Inflation Factors (VIFs) for model covariates

	Age (< 18 years*) (n = 28,061)	Age (> 18 years) (n = 28,061)	Annual Household Income (n = 28,061)	Gender (n = 28,061)	Race (n = 28,061)	Citizenship Status (n = 28,061)	Education Level (n = 28,061)
Total number of responses	0	13,382	13,382	13,382	13,382	13,382	13,382
Number of missing or removed data points	10,609	1,664	2,040	0	0	4	335
% Missing or removed	37.8	5.93	7.27	0	0	0.0143	1.19

Table 3: Summary of missing data points for depression status

**Depression screener is not offered to survey participants under the age of 18. Analysis was completed using only adults*

Model	AIC Value
(1) Simple Logistic Model	14,761.81
(2) Multivariate Logistic Model	13,686.35
(3, 4, 5, 6) Ordinal Logistic Model	21,774.81
(7, 8, 9, 10) Multinomial Logistic Model (Full)	21,779.08
(11, 12) Multinomial Logistic Model (Reduced)	18,209.19
(13) Multivariate Logistic Model (Secondary)	9,498.526

Table 4: AIC Values

REFERENCES

1. "Major Depression." *National Institute of Mental Health*, U.S. Department of Health and Human Services, www.nimh.nih.gov/health/statistics/major-depression. Accessed 3 Nov. 2023.
2. "Mental Health Conditions: Depression and Anxiety." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 14 Sept. 2022, www.cdc.gov/tobacco/campaign/tips/diseases/depression-anxiety.html.
3. "U.S. Uninsured Rate Dropped 18% during Pandemic." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 16 May 2023, www.cdc.gov/nchs/pressroom/nchs_press_releases/2023/202305.htm#:~:text=Highlights%20from%20the%20report%20include,or%203.7%20million%20in%202019.
4. Sommers, BD. Number of Young Adults Gaining Insurance Due to the Affordable Care Act Now Tops 3 Million. Office of the Assistant Secretary for Planning and Evaluation. 2012.https://aspe.hhs.gov/reports/number-young-adults-gaining-insurance-due-affordable-care-act-now-tops-3-million-0
5. Balasuriya L, Quinton JK, Canavan ME, Holland ML, Edelman EJ, Druss BG, Ross JS. The Association Between History of Depression and Access to Care Among Medicare Beneficiaries During the COVID-19 Pandemic. J Gen Intern Med. 2021 Dec;36(12):3778-3785. doi: 10.1007/s11606-021-06990-4

6. Coombs NC, Meriwether WE, Caringi J, Newcomer SR. Barriers to healthcare access among U.S. adults with mental health challenges: A population-based study. SSM Popul Health. 2021 Jun 15;15:100847. doi: 10.1016/j.ssmph.2021.100847
7. Khattar, J., Griffith, L.E., Jones, A. *et al.* Symptoms of depression and anxiety, and unmet healthcare needs in adults during the COVID-19 pandemic: a cross-sectional study from the Canadian Longitudinal Study on Aging. *BMC Public Health* 22, 2242 (2022). https://doi.org/10.1186/s12889-022-14633-4
8. "Depression (Major Depressive Disorder)." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 14 Oct. 2022, www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007.
9. Baicker K, Taubman SL, Allen HL, Bernstein M, Gruber JH, Newhouse JP, Schneider EC, Wright BJ, Zaslavsky AM, Finkelstein AN. The Oregon Experiment — Effects of Medicaid on Clinical Outcomes. N Engl J Med 2013;368:1713-22. DOI: 10.1056/NEJMsa1212321
10. Jun H, Aguila E. Private Insurance and Mental Health among Older Adults with Multiple Chronic Conditions: A Longitudinal Analysis by Race and Ethnicity. Int J Environ Res Public Health. 2021 Mar 5;18(5):2615. doi: 10.3390/ijerph18052615
11. "Nhanes - about the National Health and Nutrition Examination Survey." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 31 May 2023, www.cdc.gov/nchs/nhanes/about_nhanes.htm.
12. Cohen RA, Cha AE, Terlizzi EP, Martinez ME. Demographic Variation in Health Insurance Coverage: United States, 2019. National Health Statistics Report. 29 June 2019, https://www.cdc.gov/nchs/data/nhsr/nhsr159-508.pdf.
13. "Coverage Matters: Insurance and Health Care." Institute of Medicine (US) Committee on the Consequences of Uninsurance, 2009, https://www.ncbi.nlm.nih.gov/books/NBK223654/.
14. "Federal Poverty Level (FPL) - Glossary." *Federal Poverty Level (FPL) - Glossary | HealthCare.Gov*, www.healthcare.gov/glossary/federal-poverty-level-fpl/. Accessed 6 Dec. 2023.

15. Goodwin, Renee D et al. "Trends in U.S. Depression Prevalence From 2015 to 2020: The Widening Treatment Gap." *American journal of preventive medicine* vol. 63,5 (2022): 726-733. doi:10.1016/j.amepre.2022.05.014
16. Bailey, Rahn Kennedy et al. "Racial and ethnic differences in depression: current perspectives." *Neuropsychiatric disease and treatment* vol. 15 603-609. 22 Feb. 2019, doi:10.2147/NDT.S128584