Research question

To examine the association between education level and depression in adults 20 years or older, controlling for gender or race/hispanic origin w/ nh asian, and whether the relationship is modified by the ratio of family income to poverty.

Hypothesis

I hypothesize that adults, 20 years or older with lower education levels will be more likely to report depressive symptoms, controlling for gender, or race/hispanic origin w/ nh asian. Additionally, I expect that the relationship between education level and depressive symptomatology will be modified by the ratio of family income to poverty, such that the impact of lower education on depressive symptoms may be stronger in adults 20 years or older from low-income households.
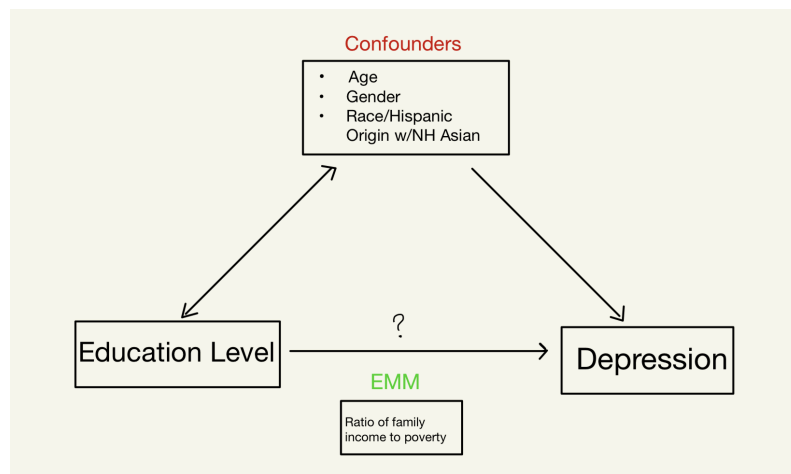

Figure 1: Directed Acyclic Graph

Abstract

Depression is a widespread mental health concern, with lower education levels being linked to increased risk. This association is influenced by various factors including gender, race/ethnicity, and socioeconomic status. Understanding these complexities is crucial for effective public health interventions. Using data from the National Health and Nutrition Examination Survey (NHANES), the association between education level and depression in adults aged 20 years or older, considering gender, race/Hispanic origin with non-Hispanic Asian, and the ratio of family income to poverty was investigated. A multistage sampling technique along with univariable, bivariable, and logistic and linear regression analyses were utilized. Among the 9,232 adults analyzed, lower education levels were correlated with higher depression prevalence and severity. Notably, individuals with some college education or above exhibited lower odds (OR=0.56, 95% CI: 0.40-0.78) and lower depression scores (-0.46

[SE=0.16], p<0.004) compared to those with less than a 9th-grade education. Additionally, females (OR=1.36, 95% CI: 1.23-1.5) and individuals from lower ratio of family income to poverty brackets had higher odds of depression. The findings highlight the inverse relationship between education level and depression risk, emphasizing the importance of educational attainment as a protective factor. While gender, race/ethnicity, and ratio of family income to poverty also play significant roles, there was no evidence of the ratio of family income to poverty modifying the relationship between education and depression. These insights underscore the need for targeted interventions addressing socioeconomic disparities to mitigate the burden of depression effectively.

Introduction/ Background

Depression is a prevalent and severe medical condition that has an adverse impact on one's emotions, thoughts, and behavior. Research indicates that there is a strong association between depression risk and lower education level. The association is further explained by many causal factors. A lower level of education may result in decreased access to support networks and resources along with financial stability which can exacerbate depression rates. Studies have shown that because education provides people with coping skills to deal with challenges, higher education can act as a protective factor against depression. However, the protective effect of education against depression is known to vary across subgroups of the population (Bauldry, et.al, 2015). Studies consistently demonstrate the intersectionality of factors such as gender, race/ethnicity, and socioeconomic status come into play with each other to influence the association of education level and depression. A plethora of studies show that women are twice as likely as men to develop depression. For instance, higher depression disorders among women was noted in studies done by Kessler, McGonagle, Swartz, Blazer, & Nelson-lifetime prevalence of 21.3%, compared with 12.7% in men (Abate, 2013). Furthermore, changing trends in the frequency of depression across different racial and ethnic groups are shown by research spanning different time periods. For example, accumulating evidence and research spanning different time periods suggests that Blacks, Hispanics, and Asians experience more depressive symptoms than their white counterparts. (Hargrove, et. al, 2020). Additionally, research has indicated an association between higher prevalence of depression and low socioeconomic status. Adults from lower-income families are more likely than people from higher-income households to experience depression. Specifically, adults living below the poverty level showed a considerably greater prevalence of depression than those above the line, according to a study conducted by Pratt et al. (2014). This unequal distribution implies that overlapping factors including education level, race/ethnicity, gender, and socioeconomic status have a direct effect of placing groups at different risks for depression. Thus, comprehending the association between education level and depression along with these covariates is essential for mitigating the prevalence of depression in the U.S. across a variety of demographics.

In the coming decades, depression is expected to be the second greatest cause of disability in the world and the leading cause of disability in the United States (Gonzalez, 2010). Information on the research and the strength of these studies' trends indicate a growing public health concern. Depression has a wide range of detrimental effects on people's quality of life and overall well-being.

Depressive symptoms are linked to significant social and personal burdens, poor health conditions, limited achievement of socioeconomic resources such as educational attainment and earnings and strained social and familial relationships (Hargrove, et.al, 2020). Untreated depression can also lead to serious consequences including suicide thoughts and actions, which increases the burden of disability and mortality. Understanding the complex effects of depression highlights the pressing need for all-encompassing public health initiatives focused on primary, secondary, and tertiary preventative methods that in turn could exponentially decrease the burden of depression.

There is a known association between depression and education level. This association highlights the complex nature of depression, which can be influenced by sociodemographic factors such as gender, race/ethnicity, and the ratio of family income to poverty, in addition to education level. Moreover, research continuously shows how education levels interact to shape the prevalence of depression in adults especially those with lower socioeconomic status. For example, Blacks and Hispanics are more likely to live in poverty and have lower incomes for a given level of education compared to their white and Asian American counterparts (Hargrove, et.al, 2020). Education has a direct correlation to increased income, job stability, and a supportive network of, all of which can greatly lessen or prevent depression symptoms. Given such importance of coping methods like increased income, job security and network, the education gradient in depression is usually prominent. Comprehending and tackling this association is crucial in formulating efficacious methods targeted at decreasing depression prevalence specifically for susceptible groups. Additionally, the changing of patterns and trends in the prevalence of depression over time and among various populations point to an increasing public health problem. My research aims to examine the association between education level and depression in adults 20 years or older, controlling for gender or race/hispanic origin w/ nh asian, and whether the relationship is modified by the ratio of family income to poverty. I hypothesize that adults, 20 years or older with lower education levels will be more likely to report depressive symptoms, controlling for gender, or race/hispanic origin w/ nh asian. Additionally, I expect that the relationship between education level and depressive symptomatology will be modified by the ratio of family income to poverty, such that the impact of lower education on depressive symptoms may be stronger in adults 20 years or older from low-income households.

Methods

1. **Study Design**

        The National Center for Health Statistics (NCHS) conducts the National Health and Nutrition Examination Survey (NHANES) to gather comprehensive data on the nutritional status and general health of the noninstitutionalized civilian population in the United States. NHANES data collection consists of two primary components: interviewer administered surveys and a clinician administered medical examination. Personal interviews are used in the surveys to gather person-level demographic, health, and nutrition information, while standardized procedures for medical examinations and specimen collection occur in specialized mobile examination centers (MECs)(Akinbami, et.al, 2022). In order to create a representative sample, NHANES employs a multistage sampling strategy that entails many selection processes. Primary sample units (PSUs) are chosen from every county in the United States during the first phase. In later phases, smaller units are sampled from these PSUs, which function as geographic clusters. NHANES uses stratified sampling, which divides the population into subgroups/strata based on population demographics (e.g., age, race/ethnicity, income) within each chosen PSU. NHANES also employs stratification in order to guarantee that the chosen PSUs are distributed equally across a range of factors, including geography, urban-rural distribution, and population demographics. Clustered sampling is another technique used by NHANES, in which groups of people are sampled at every step of the selection process. Clusters are created within the four stages of NHANES: PSUs, segments within PSUs (census blocks), dwelling units (households), and individuals inside households. Additionally, NHANES uses probability-based sampling techniques, meaning that each member of the target population has a known chance of being chosen for participation. NHANES employs a multistage, stratified, clustered probability-based sampling design to ensure that the data that is collected adequately represents the features of the noninstitutionalized civilian U.S. population.

2. **Study Population**

        The target population in the NHANES data is the resident civilian noninstitutionalized population of the United States. The exclusion criteria of the NHANES data is all people in supervised care or custody in institutional settings, all active-duty military personnel, active-duty family members living overseas, and any other people living outside the 50 states and the District of Columbia (Akinbami, et.al, 2022). The sample size for the sample of individuals that completed the interview was 15,560, while the sample size for those who additionally had a medical examination was 14,300. Furthermore, the target population for the analysis of depression and education level as derived from the NHANES sample consists of adults aged 20 years or older. After removing individuals under 20 years old, the final analytic sample for the study includes n=9,232 adults, including those with missing values (Refused and Don't Know). This exclusion criterion makes sure that the study's results and conclusions apply

only to adults in the specified age range and aren't confounded by the inclusion of those under 20 years old.

3. **Main Outcome of Interest:**

The main outcome of interest—depression—is measured in the NHANES dataset using the Patient Health Questionnaire (PHQ-9), a standardized tool for detecting depressive disorders that are commonly experienced. The PHQ-9 is a self-administered questionnaire in which participants assess their frequency of their symptoms over the past two weeks, with response criteria as "0" (not at all) to "3" (nearly every day) (Kroenke, et.al, 2001). The total score is calculated by summing up the scores for all nine items, yielding a range of 0 to 27, with higher scores signifying more severe symptoms. In this study, depression will be operationalized as a dichotomous variable (depression yes/no), with total scores between 0 and 9 indicating having no depression and scores between 10 and 27 indicating having depression.

4. **Main Exposure of Interest:**

The main exposure of interest in the NHANES dataset—education level—is represented by the variable DMDEDUC2, which is collected during the household interview. This categorical variable indicates the participant's highest level of education. Education level is divided into five categories: 1. Less than 9th grade, 2. 9-11th grade (Includes 12th grade with no diploma), 3. High school graduate/GED or equivalent, 4. Some college or AA degree, and 5. College graduate or above. Additionally, participants were able to answer 7. Refused or 9. Don't Know. These categories will be used to operationalize education level and examine its association with depression in adults 20 years or older.

5. **Covariates Of Interest**

In the NHANES dataset, the covariates of interest include potential confounders, gender and race/hispanic origin with non-hispanic asian, and the effect modifier, ratio of family income to poverty. The variable RIDRETH3 represents the participant's race and Hispanic origin, with an additional category for Non-Hispanic Asians. Race/Hispanic Origin with Non-Hispanic Asian is divided into six categories: Mexican American as 1, Other Hispanic as 2, Non-Hispanic White as 3, Non-Hispanic Black as 4, Non-Hispanic Asian as 6, and Other Race - Including Multi-Racial as 7. Additionally, the variable RIAGENDR is used to represent the participant's gender. In this study, a participant's gender is collected as either male, coded as 1, or female, coded as 2.

The variable INDFMPIR represents the effect measure modifier, ratio of family income to poverty. This ratio is an indicator of the participant's socioeconomic status that is based on family size and household income. The ratio is calculated by dividing the family income by the poverty threshold that varies depending on the size of the household. This categorical variable is divided into three groups: income below the poverty level are represented by values ranging from 0 to 4.98, while incomes at or above the poverty level are shown by values of 5 or above. Moreover, missing values are represented by a period. However, because the categorization of the family income to poverty variable is a ratio, it needs to be recategorized to be a discrete

variable. A new ratio of family income to poverty variable was generated, 'indfmpir1', and replaced the values of the original ratio of family income to poverty variable 'indfmpir' that are greater than or equal to 0 and less than or equal to 1.99 with 1, that could represent low income ratios. Moreover, 'indfmpir1' replaced the values of 'indfmpir' that are greater than or equal to 2 and less than or equal to 3.99 with 2, that could represent mid level income ratios. Lastly, values of 'indfmpir' that are greater than or equal to 4 with 3, that could represent high income ratios. Therefore, the values of the original family income to poverty variable are categorized into 1, low income, 2 mid level income, and 3 high income categories in the new ratio of family income to poverty variable.

6. Statistical Analysis Plan

a) Descriptive Analysis

The study's variables will be organized using descriptive statistics. Categorical variables such as education level, gender, race/Hispanic origin, and family income-to-poverty ratio will be computed as frequencies and percentages. Means, standard deviations, medians, and interquartile ranges will be computed for continuous/ordinal variables such as age and family income to poverty ratio. These statistics offer useful details on the variables' distribution and may help the comprehension of the characteristics of the study population.

The bivariable analysis will include the association between interest exposure, education level, and the outcome of interest, depression, and other covariates (gender, race/ethnicity, family income to poverty ratio) being evaluated using chi-square tests and t-tests. For categorical data, chi-square tests will be employed, while t-tests will be employed for continuous variables. Stratified measures of association between education and depression will be produced to assess potential confounders (gender, race/Hispanic origin) and the effect measure modifier (ratio of family income to poverty). This method makes it possible to determine if the association between depression and education level varies depending on the degree of potential confounders and effect measure modifiers. Moreover, results will be considered statistically significant at alpha <0.05. Moreover, model building procedures will be used for both logistic and linear regression analyses. The depression score (PHQ-9) will be considered a continuous outcome variable in a linear regression analysis, while education level, gender, race/Hispanic origin with non-Hispanic Asian, and family income to poverty ratio will be accounted for as independent factors. To make sure the predictor variables are independent, tests for collinearity and goodness of fit will be performed. The same independent variables used in the linear regression analyses will be utilized in the logistic regression analyses, with depression being the outcome variable. After accounting for gender, race/Hispanic origin with non-Hispanic Asian, and family income to poverty ratio, the associations between education level and depression will be measured using regression coefficients, along with the corresponding confidence intervals and p-values.

Results

Univariable Analysis:

Table 1. Univariable analysis of sociodemographic characteristics including education level, gender, race/Hispanic origin, and family income-to-poverty ratio among adults 20 years or older in the resident civilian noninstitutionalized population of the NHANES dataset, United States, 2017- 2020, n=9232

Key characteristics	Study Sample N (%) Mean (SD) Median (IQR)
Education Level	
Less than 9th grade	719 (7.79)
9-11th grade (Includes 12th grade with no diploma)	1041 (11.28)
High school graduate/GED or equivalent	2225 (24.10)
Some college or AA degree	2975 (32.22)
College graduate or above	2257 (24.45)
Refused	2 (0.02)
Don't Know	13 (0.14)
Race/Ethnicity	
Mexican American	1057 (11.45)
Other Hispanic	940 (10.18)
Non-Hispanic White	3217 (34.85)
Non-Hispanic Black	2459 (26.64)
Non-Hispanic Asian	1120 (12.13)
Other Race - Including Multi-Racial	439 (4.76)
Gender	
Male	4479 (48.52)
Female	4753 (51.48)
Ratio of family income to poverty	
Low	3579 (38.77)
Mid	2118 (22.94)
High	3535 (38.29)
Depression	
No	7163 (77.59)
Yes	2069 (22.41)
Depression Score	
Mean (SD)	2.88 (4.39)
Median (IQR)	1 (4)

Univariable analysis and summary:

The table presents the results of a univariable analysis of sociodemographic characteristics, including education level, gender, race/Hispanic origin, and family income-to-poverty ratio among adults aged 20 years or older in the resident civilian noninstitutionalized population of the NHANES dataset from 2017 to 2020. The analytic sample consisted of 9,232 individuals. Regarding education level, the majority of

Figure 2: Histogram of Depression Score Distribution

participants had completed at least a high school education or equivalent, with 32.22% having some college or an associate's degree, and 24.45% being college graduates or above. However, a notable proportion (7.79%) had less than a 9th-grade education. In terms of race/ethnicity, non-Hispanic whites comprised the largest group (34.85%), followed by non-Hispanic blacks (26.64%) and Hispanics (21.63%). Furthermore, females slightly outnumbered males in the sample (51.48% vs. 48.52%). The distribution of family income-to-poverty ratio showed that a considerable proportion of participants (38.77%) fell into the low-income category, while 38.29% were classified as high income, and the remaining 22.94% belonged to the mid-level income group.

Moreover, the distribution of depression scores is further illustrated by figure 2, which shows a right-skewed distribution. The majority of participants had low depression scores, as evidenced by the peak at a score of 0. However, there is a long tail extending to the right, indicating that a considerable proportion of participants had higher depression scores. Specifically, the histogram reveals that while a large number of participants had a depression score of 0, there was a small number of participants who had depression scores as high as 15. This visual representation reinforces the wide variation in depression scores observed in the sample, with a mean score of 2.88 (SD = 4.39) and a median of 1.

Overall, the results indicate that the analytic sample has a wide range of sociodemographic characteristics. A significant percentage of the population had lower incomes and education levels, which is noteworthy and offers a suitable context for examining the association between depression and education level while considering potential moderators like income and adjusting for confounding variables like gender and race/ethnicity. Additionally, given that 22.41% of participants were classified as depressed, the relatively high prevalence of depression further emphasizes the need for additional research into potential risk factors and modifiers, such as family income and education level, in order to inform targeted interventions meant to lessen depressive symptoms in vulnerable populations. Furthermore, there were skip patterns in the distribution of variables.
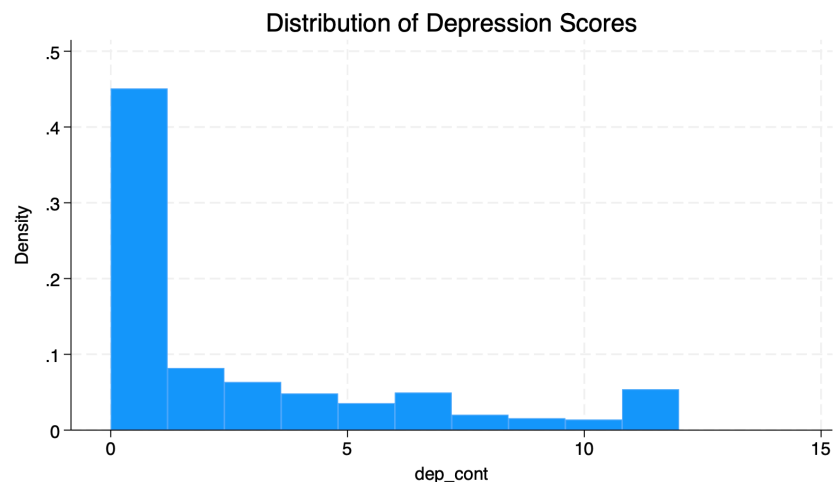
BIVARIABLE ANALYSES:

Table 2. Bivariable associations between education level and depression and key sociodemographic characteristics and covariates, among adults 20 years or older in the resident civilian noninstitutionalized population of the NHANES dataset, United States, 2017- 2020, n=9,232 .

Key Characteristics	Education Level						p-value
	Total N*	Less than 9th grade N (%)	9-11th grade (Includes 12th grade with no diploma) N (%)	High school graduate/ GED or equivalent N (%)	Some college or AA degree N (%)	College graduate or above N (%)	
Race/Ethnicity							<0.001²
Mexican American	1055	279 (26.45)	178 (16.87)	244 (23.13)	261 (24.74)	93 (8.82)	
Other Hispanic	935	195 (20.86)	141 (15.08)	192 (20.53)	256 (27.38)	151 (16.15)	
Non-Hispanic White	3215	81 (2.52)	287 (8.93)	835 (25.97)	1155 (35.93)	857 (26.66)	
Non-Hispanic Black	2455	62 (2.53)	321 (13.08)	704 (28.68)	909 (37.03)	459 (18.70)	
Non-Hispanic Asian	1119	87 (7.77)	75 (6.70)	150 (13.40)	202 (18.05)	605 (54.07)	
Other Race-Including Multi-Racial	438	15 (3.42)	39 (8.90)	100 (22.83)	192 (43.84)	92 (21.00)	
Gender							0.001²
Male	4472	365 (8.16)	534 (11.94)	1131 (25.29)	1361 (30.43)	1081 (24.17)	
Female	4745	354 (7.46)	507 (10.68)	1094 (23.06)	1614 (34.01)	1176 (24.78)	
Ratio of family income to poverty							<0.001²
Low	3576	408 (11.41)	612 (17.11)	1111 (31.07)	1099 (30.73)	346 (9.68)	
Mid	2117	101 (4.77)	175 (8.27)	496 (23.43)	850 (40.15)	495 (23.38)	
High	3524	210 (5.96)	254 (7.21)	618 (17.54)	1026 (29.11)	1416 (40.18)	
Covered by Health Insurance							<0.001²
Yes	7733	512 (6.62)	795 (10.28)	1790 (23.15)	2517 (32.55)	2119 (27.40)	
No	1464	206 (14.07)	241 (16.46)	432 (29.51)	449 (30.67)	136 (9.29)	
Age							<0.001⁴
Mean (SD)	9217	57.85(15.38)	52.86 (17.9)	51.26 (18.46)	49.18 (17.82)	50.59(16.67)	
Depression Score							<0.001⁴
Mean (SD)	8805	2.43 (3.82)	2.82 (3.8)	2.63 (3.58)	2.63 (3.44)	1.76 (2.69)	
Depression							<0.001²
No	7159	478 (6.68)	735 (10.27)	1711 (23.90)	2360 (32.97)	1875 (26.19)	
Yes	2058	241 (11.71)	306 (14.87)	514 (24.98)	615 (29.88)	382 (18.56)	

*Frequencies may not add up to total analytic sample size due to missing values.

1 p-value for Pearson's r

2 p-value for chi-square test

3 *p*-value for t-test

4 *p*-value for ANOVA

Table 3a. Bivariable associations between depression (categorical) and education level, key sociodemographic characteristics and other covariates, among adults 20 years or older in the resident civilian noninstitutionalized population of the NHANES dataset, United States, 2017- 2020, , n=9,232.

Key Characteristics	Depression			*p*-value
	Total N (%)*	NO N (%)	YES N (%)	
Race/Ethnicity				0.080²
Mexican American	1057	835 (79.00)	222 (21.00)	
Other Hispanic	940	719 (76.49)	221 (23.51)	
Non-Hispanic White	3217	2529 (78.61)	688 (21.39)	
Non-Hispanic Black	2459	1906 (77.51)	553 (22.49)	
Non-Hispanic Asian	1120	852 (76.07)	268 (23.93)	
Other Race - Including Multi-Racial	439	322 (73.35)	117 (26.65)	
Gender				<0.001²
Male	4479	3589 (80.13)	890 (19.87)	
Female	4753	3574 (75.19)	1179 (24.81)	
Ratio of family income to poverty				<0.001²
Low	3579	2680 (74.88)	899 (25.12)	
Mid	2118	1742 (82.25)	376 (17.75)	
High	3535	2741 (77.54)	794 (22.46)	
Education Level				<0.001²
Less than 9th grade	719	478 (66.48)	241 (33.52)	
9-11th grade (Includes 12th grade…)	1041	735 (70.61)	306 (29.39)	
High school graduate/GED or equivalent	2225	1711 (76.90)	514 (23.10)	
Some college or AA degree	2975	2360 (79.33)	615 (20.67)	
College graduate or above	2257	1875 (83.07)	382 (16.93)	
Covered by Health Insurance				0.870²
Yes	7,743	6011 (77.63)	1732 (22.37)	
No	1,467	1136 (77.44)	331 (22.56)	
Age				0.0001⁴
Mean (SD)	9232	50.76 (17.46)	52.48 (18.38)	

*Frequencies may not add up to total analytic sample size due to missing values.
1 p-value for Pearson's r
2 p-value for chi-square test
3 *p*-value for t-test
4 *p*-value for ANOVA

Table 3b. Bivariable associations between depression (continuous) and education level, key sociodemographic characteristics and other covariates, among adults 20 years or older in the resident civilian noninstitutionalized population of the NHANES dataset, United States, 2017- 2020, n=9,232.

Key Characteristics	Depression Score		p-value
	Total N (%)*	Mean (SD)	
Race/Ethnicity			<0.001[4]
Mexican American	1006	2.36 (3.36)	
Other Hispanic	892	2.74 (3.69)	
Non-Hispanic White	3077	2.6 (3.45)	
Non-Hispanic Black	2344	2.38 (3.39)	
Non-Hispanic Asian	1096	1.45 (2.55)	
Other Race - Including Multi-Racial	405	3.31 (4.01)	
Gender			<0.001[3]
Male	4302	2.08 (3.16)	
Female	4518	2.74 (3.59)	
Ratio of family income to poverty			<0.001[4]
Low	3370	2.95 (3.8)	
Mid	2038	2.33 (3.19)	
High	3412	1.94 (3.02)	
Education Level			<0.001[4]
Less than 9th grade	677	2.42 (3.82)	
9-11th grade (Includes 12th grade…)	986	2.82 (3.8)	
High school graduate/GED or equivalent	2121	2.63 (3.58)	
Some college or AA degree	2827	2.63 (3.44)	
College graduate or above	2194	1.76 (2.69)	
Covered by Health Insurance			0.0721[3]
Yes	7412	2.39 (3.39)	
No	1387	2.57 (3.49)	
Age			0.0013[1]
Mean (SD)	8,820	2.42 (3.4)	

*Frequencies may not add up to total analytic sample size due to missing values.
1 p-value for Pearson's r
2 p-value for chi-square test
3 p-value for t-test
4 p-value for ANOVA

Table 2 displays a bivariable association between education level and depression and key sociodemographic characteristics, race/ethnicity, gender, and ratio of family income to poverty, among adults 20 years or older in the resident civilian noninstitutionalized population of the NHANES dataset. The table shows the distribution of education levels across different socio demographic groups, along with the proportions within each education level category. Chi-square and anova tests were conducted to examine the association between education level and each sociodemographic characteristic and depression variables. It's important to note that the values representing "Refused" and "Don't Know" in education level, depression scores, and depression categories were recorded as missing values and not included in the bivariable associations. While causing the difference in total proportions, this ensures that the analyses focus only on valid responses and avoids potential biases introduced by uncertain or incomplete data. The results indicate significant associations between education level and all sociodemographic characteristics, with p-values of 0.001 and less than 0.001 at the significance level of 0.05. There are clear differences in the distributions of race/ethnicity, gender, ratio of family income to poverty, covered by health insurance, age, and depression across education levels. For example, a higher proportion of Non-Hispanic Whites (26.66%) and Asians (54.07%) had college or higher education compared to other racial/ethnic groups. Similarly, those with higher incomes were more likely to have higher educational attainment, with 40.18% of individuals in the high-income group being college graduates or above, compared to 9.68% in the low-income group (p<0.001). Among college graduates, 48.1% were males and 51.9% were females. Additionally, a higher proportion of the college educated (27.4%) had health insurance coverage compared to those with less than 9th grade education (6.62%). Moreover, the mean age increased with higher education levels, from 57.85 years for <9th grade to 50.59 years for college graduates. Notably, the mean depression scores decreased with higher education, from 2.43 for <9th grade to 1.76 for college graduates. This explains the results of depression prevalence being higher at lower education levels - 33.52% for <9th grade compared to 16.93% for college graduates.

Table 3a shows the bivariable associations between depression (categorical) and education level, and sociodemographic characteristics. The prevalence of depression (yes) was higher among individuals with lower education levels, with 33.52% of those with less than 9th-grade education having depression, compared to 16.93% of college graduates or above (p<0.001). However, the prevalence of depression did not significantly differ across racial/ethnic groups, with (p= 0.080) for the association between race/ethnicity and depression (yes/no). Females had a higher prevalence of depression (24.81%) compared to males (19.87%) (p<0.001). Additionally, individuals from low-income households had a higher prevalence of depression (25.12%) than those from mid-income (17.75%) and high-income (22.46%) households (p<0.001). The prevalence of depression (yes) was higher among older individuals, with a mean age of 52.48 (SD=18.38) years for those with depression compared to 50.76 (SD=17.46) years for

those without depression (p<0.001). However, the prevalence of depression did not significantly differ between those covered by health insurance (22.37%) and those not covered (22.56%), with a p-value of 0.870. The results indicate significant associations between depression (categorical) and education level, gender, ratio of family income to poverty, and age, with p-values of less than 0.001 while race/ethnicity and covered by health insurance are not.

In table 3b, mean depression scores were compared across sociodemographic characteristics using ttests, ANOVA, and Pearson's r tests. The mean depression score is lowest among Non-Hispanic Asians (1.45, SD=2.55) and highest among Other Race - Including Multi-Racial individuals (3.31, SD=4.01)(p<0.001). Females had significantly higher mean depression scores (2.74, SD=3.59) than males (2.08, SD=3.16) (p<0.001). The ratio of family income to poverty level showed a significant association with depression scores, with individuals from low-income households having higher mean scores (2.95, SD=3.8) compared to those from mid (2.33, SD=3.19) and high-income (1.94, SD=3.02) households (p<0.001). Furthermore, lower education levels were associated with higher mean depression scores, ranging from 2.42 (SD=3.82) for those with less than 9th-grade education to 1.76 (SD=2.69) for college graduates or above (p<0.001). Moreover, not being covered by health insurance had higher mean depression scores 2.57 (SD=3.49) than being covered by health insurance 2.39 (SD=3.39) (<0.001). Additionally, the mean depression score of age is 2.42, with a standard deviation of 3.4 and a p value of 0.0013, indicating variability in depression scores across different age groups. The results indicate significant associations between depression scores and education level, gender, ratio of family income to poverty, race/ethnicity, and age while covered by health insurance is not.

Table 4a. Multivariable logistic regression examining the association between education level and depression symptom severity, key sociodemographic characteristics and other covariates, among adults 20 years or older, NHANES, 2017-pre pandemic 2020

	Crude OR (95% *CI*)	Full model 1 Adjusted *OR* (95% *CI*)	Full model 2 Adjusted *OR* (95% *CI*)
Education Level			
Less than 9th grade	ref	ref	ref
9-11th grade (Includes 12th grade with no diploma)	0.83 (0.67-1.01)	0.79 (0.64-0.98)	0.78 (0.59-1.03)
High school graduate/GED or equivalent	0.60 (0.50-0.72)	0.56 (0.46-0.69)	0.65 (0.5-0.84)
Some college or AA degree	0.52 (0.43-0.62)	0.49 (0.4-0.59)	0.58 (0.44-0.75)
College graduate or above	0.40 (0.33-0.49)	0.35 (0.29-0.44)	0.56 (0.4-0.78)
Race/ethnicity			
Mexican American	0.98 (0.82-1.16)	0.75 (0.62-0.9)	0.74 (0.61-0.89)
Other Hispanic	1.13 (0.95-1.34)	0.9 (0.75-1.08)	0.88 (0.73-1.06)
Non-Hispanic White	ref	ref	ref
Non-Hispanic Black	1.07 (0.94-1.21)	1.02 (0.89-1.16)	1.003 (0.88-1.14)
Non-Hispanic Asian	1.16 (0.98-1.36)	1.24 (1.05-1.47)	1.24 (1.05-1.47)
Other Race - Including Multi-Racial	1.34 (1.06-1.68)	1.36 (1.08-1.71)	1.35 (1.07-1.7)
Gender			
Male	ref	ref	ref
Female	1.33 (1.21-1.47)	1.36 (1.23-1.51)	1.36 (1.23-1.5)
Ratio of family income to poverty			
Low	ref	ref	ref
Mid	0.64 (0.56-0.74)	0.74 (0.64-0.85)	1.03 (0.64-1.65)
High	0.86 (0.77-0.96)	1.02 (0.91-1.15)	1.35 (0.95-1.92)
Age	1.006 (1.003-1.008)	1.004 (1.0007-1.007)	1.004 (1.0007-1.007)
Education Level x Ratio of family income to poverty			
Less than 9th grade, Low	NA	NA	ref
9-11th grade (Includes 12th grade with n..), Mid	NA	NA	0.71 (0.38-1.32)
9-11th grade (Includes 12th grade with n..), High	NA	NA	1.28 (0.8-2.05)
High school graduate/GED or equivalent, Mid	NA	NA	0.67 (0.39-1.15)
High school graduate/GED or equivalent, High	NA	NA	0.76 (0.5-1.15)
Some college or AA degree, Mid	NA	NA	0.7 (0.41-1.18)
Some college or AA degree, High	NA	NA	0.68 (0.45-1.03)
College graduate or above, Mid	NA	NA	0.52 (0.29-0.94)
College graduate or above, High	NA	NA	0.49 (0.31-0.77)

OR = Odds ratio

CI = Confidence interval

ref = Reference Group

Table 4b. Multivariable linear regression examining the association between education level and depression symptom severity, key sociodemographic characteristics and other covariates, among adults 20 years or older, NHANES, 2017-pre pandemic 2020

	Unadjusted model		Full model 1		Full model 1	
	beta (*SE*)	p-value	beta (*SE*)	p-value	beta (*SE*)	p-value
Education Level						
Less than 9th grade	ref		ref	ref	ref	
9-11th grade (Includes 12th grade with no diploma)	0.39 (0.17)	0.020	0.25 (0.17)	0.147	0 .5 (0.22)	0.027
High school graduate/GED or equivalent	0.2 (0.15)	0.180	0.05 (0.16)	0.725	0.27 (0.21)	0.186
Some college or AA degree	0.2 (0.14)	0.168	0.06 (0.15)	0.714	0.36 (0.21)	0.085
College graduate or above	-0.67 (0.15)	<0.001	-0.46 (0.16)	0.004	-0.27 (0.26)	0.284
Race/ethnicity						
Mexican American	-0.24 (0.12)	0.048	-0.49 (0.13)	<0.001	-0.49 (0.13)	<0.001
Other Hispanic	0.14 (0.13)	0.263	-0.67 (0.13)	0.610	-0.07 (0 .13)	0.607
Non-Hispanic White	ref	ref	ref	ref	ref	ref
Non-Hispanic Black	-0.22 (0.09)	0.017	-0.4 (0.09)	<0.001	-0.41 (0.09)	<0.001
Non-Hispanic Asian	-1.14 (0.12)	<0.001	-1.003 (0.12)	<0.001	-1.004 (0.12)	<0.001
Other Race - Including Multi-Racial	0.71 (0.18)	<0.001	0.54 (0.18)	0.002	0.53 (0.18)	0.003
Gender						
Male	ref		ref		ref	
Female	0.66 (0.07)	<0.001	0.65 (0.07)	<0.001	0.64(0.07)	<0.001
Ratio of family income to poverty						
Low	ref	ref	ref		ref	ref
Mid	-0.63 (0.09)	<0.001	-0.5 (0.1)	<0.001	-0.27 (0.38)	0.477
High	-1.01 (0.08)	<0.001	-0.74 (0.09)	<0.001	-0.1 (0.29)	0.731
Age	-0.01 (0.002)	0.001	-0.01 (0.002)	<0.001	-0.01 (0.002)	<0.001
Education Level x Ratio of family income to poverty						
Less than 9th grade, Low	NA	NA	NA	NA	ref	
9-11th grade (Includes 12th grade with n..), Mid	NA	NA	NA	NA	-0.25 (0.48)	0.595
9-11th grade (Includes 12th grade with n..), High	NA	NA	NA	NA	-0.73 (0.39)	0.058
High school graduate/GED or equivalent, Mid	NA	NA	NA	NA	-0.23 (0.42)	0.589
High school graduate/GED or equivalent, High	NA	NA	NA	NA	-0.63 (0.34)	0.062
Some college or AA degree, Mid	NA	NA	NA	NA	-0.33 (0.41)	0.422
Some college or AA degree, High	NA	NA	NA	NA	-0.79 (0.33)	0.016
College graduate or above, Mid	NA	NA	NA	NA	-0.18 (0.45)	0.686
College graduate or above, High	NA	NA	NA	NA	-0.6 (0.36)	0.092

SE=standard error

ref = Reference Group

Table 4a shows the results of a multivariable logistic regression analysis that highlights several significant associations between sociodemographic characteristics and depression severity. A notable observation in the unadjusted model is the odds of depression among individuals with a 9-11th grade (Includes 12th grade with no diploma) education are 0.83 times the odds of depression among individuals with an education of less than 9th grade (referent group) in this study sample (OR=0.83, 95% CI 0.67-1.01). Additionally, in both the unadjusted and adjusted models, the odds of depression severity among Other Race - Including Multi-Racial are around 1.34 times the odds of depression severity among Non-Hispanic Whites in this study sample (OR = 1.34, 95% CI 1.06-1.68) (OR = 1.36, 95% CI 1.08-1.71). Moreover, in both the unadjusted and adjusted models, the odds of depression among females are around 1.33 times the odds of depression among males (referent group) in this study sample (OR=1.33, 95% CI 1.21-1.47) (OR=1.36, 95% CI 1.23-1.51). Additionally, those in the high income group had higher odds of 0.86 of depression severity compared to the low income reference (OR=0.86, 95% CI 0.77-0.96). The ORs for higher education levels generally showed an increasing trend for depression symptom severity in the fully adjusted models. For example, the OR for some college or AA degree was 0.35 (95% CI: 0.29-0.44) in model 1 and 0.56 (95% CI: 0.4-0.78) in model 2. Even after adjusting for education level, gender, age, race/ethnicity, and income-to-poverty ratio, higher education levels, males, Mexican American ethnicity, and attaining high level income are associated with lower odds of depression in this study sample. The interaction terms between education level and the ratio of family income to poverty in model 2 were not statistically significant, as the confidence intervals crossed 1. Thus, there was no evidence of effect measure modification by ratio of family income to poverty on the relationship between education level and depression in this study sample.

Table 4b presents the results of a multivariable linear regression analysis. On average individuals who are college graduates or above had 0.67 point lower depression score compared with individuals with less than a 9th grade education in this study sample (= -0.67[SE=0.15], p < 0.001) Additionally, non hispanic asians had lower mean depression scores in all three models where they had around a 1 point lower depression score compared to non hispanic whites (= -1.14[SE=0.12], p < 0.001) (=-1.004[SE= 0.12], p < 0.001). Additionally, females scored 0.66 point higher depression scores compared with men across all three models (=0.66[SE=0.07], p<0.001) (=0.64[SE=0.07], p<0.001). Moreover, individuals who attained higher income levels had around a 1 point lower depression score compared with individuals with lower income levels across the unadjusted and first adjusted model but only had a 0.27 point lower depression score (0.38) (=-1.01[SE=0.08], p<0.001) (=-0.27 [SE=0.38] , p<0.477). Most of the interaction term beta coefficients have large standard errors and non-significant p-values (e.g. less than 9th grade*mid income: = -0.25, SE = 0.48, p = 0.595). Thus, there was no evidence of effect modification by ratio of family income to poverty on the relationship between education level and depression score in this study sample. So lower education, Mexican American/Other Hispanic ethnicity, males, lower ratios of family income to poverty (though the high ratio was not statistically significant) were associated with lower depression symptom severity after controlling for education level, gender, race/ethnicity, age, and ratio of family income. For both the logistic and linear regressions, most of the associations between race/ethnicity, gender, and ratio of family income to poverty with depression symptom severity remained significant even after adjusting for education level and other covariates in the fully

adjusted models.

Discussion

The study investigated the association between education level and depression in adults aged 20 years or older, while considering potential influence of gender, race/Hispanic origin with non-Hispanic Asian, and the ratio of family income to poverty. The main findings indicate a significant inverse relationship between education level and depression risk. Specifically, individuals with higher education levels exhibited lower odds and severity of depression compared to those with lower education levels. These findings are consistent with previous research demonstrating the protective effect of education against depression (Bauldry et al., 2015). However, the study extends this understanding by examining the role of additional sociodemographic factors such as gender, race/ethnicity, and income-to-poverty ratio. Strengths of the study include the use of data from the NHANES, which employs a rigorous multistage sampling design to ensure representation of the U.S. population. The large sample size and comprehensive assessment of sociodemographic variables allowed for robust analyses. Additionally, the utilization of standardized measures, such as the Patient Health Questionnaire (PHQ-9), enhances the reliability and validity of the findings. However, several limitations should be acknowledged. First, the cross-sectional nature of the study design precludes causal inference. Additionally, the reliance on self-reported measures, including education level and depressive symptoms, introduces potential information bias leading to misclassification of exposure or outcome variables. Furthermore, the exclusion of certain population groups, such as individuals in institutional settings, may limit the generalizability of our findings to these populations and potentially result in a non-representative sample leading to selection bias. The presence of biases, both selection and information, limits both the internal and external validity of the study. Moreover, to minimize bias, several strategies were employed. These included rigorous quality control measures during data collection and analysis, as well as sensitivity analyses to assess the robustness of our findings to potential biases. While the study provides valuable insights into the association between education level and depression, the presence of biases must be considered when interpreting the findings. Future research utilizing longitudinal designs and objective measures of both exposure and outcome variables may help to overcome these limitations and provide further clarity on this important public health issue.

References:

Abate, K. H. H. *Gender disparity in prevalence of depression among patient population: A systematic review*. Ethiopian journal of health sciences. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3847538/. Published November, 2013. Accessed February 19, 2024.

Akinbami, L. J., Chen, T. C., Davy, O., Ogden, C. L., Fink, S., Clark, J., Riddles, M. K., & Mohadjer, L. K. (2022). *National Health and Nutrition Examination Survey, 2017-March 2020 Prepandemic File: Sample Design, Estimation, and Analytic Guidelines.* Vital and health statistics. Ser. 1, Programs and collection procedures. https://pubmed.ncbi.nlm.nih.gov/35593699/. Published May, 2022. Accessed February 29, 2024.

Bauldry, S. *Variation in the Protective Effect of Higher Education against Depression*. Sage Journals. https://journals.sagepub.com/doi/full/10.1177/00081256231179993. Published January 12, 2015. Accessed February 19, 2024.

González, H. M., Tarraf, W., Whitfield, K. E., & Vega, W. A. *The epidemiology of Major Depression and ethnicity in the United States*. Journal of Psychiatric Research. https://www.sciencedirect.com/science/article/pii/S0022395610000932?casa_token=Bq AQSVLkCOwAAAAA%3AK-XicndRmZKtl8njsiRXRAr_Pm0XC-WJbAKvm06J55nIrIjjimUgVuP 5Rg4L7uxalqRYkg#bib20. Published May 27, 2010. Accessed February 19, 2024.

Hargrove, T. W., Halpern, C. T., Gaydosh, L., Hussey, J. M., Whitsel, E. A., Dole, N., Hummer, R. A., & Harris, K. M. *Race/ethnicity, gender, and trajectories of depressive symptoms across early- and mid-life among the add health cohort*. Journal of racial and ethnic health disparities. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7338256/#:~:text=Racial%2Fethnic%20 minorities%20reported%20more,increases%20in%20their%20late%2030s. Published August, 2020. Accessed February 19, 2024.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001, September). The PHQ-9: Validity of A brief depression severity measure. Journal of general internal medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/. Published September, 2001. Accessed February 24, 2024

Pratt, L. A., & Brody, D. J. *Depression in the U.S. Household Population, 2009-2012*. Centers

for Disease Control and Prevention.
https://www.cdc.gov/nchs/products/databriefs/db172.htm. Published November
6, 2014. Accessed February 19, 2024.