The Association Between Depression and Marital Status Among Adults in the United States (50 states + District of Columbia)

Misato Nasu

National University

ANALYTICS 625: Categorical Data Methods

Dr. Kevin Duffy-Deno

12/18/2023

Objective

Depression is a problem that affects not only the US but countries all around the world. Though there are no specific causes for depression, it is necessary to investigate this problem to understand and find cures and prevention methods. The objective of this study is to investigate the association between depression and marital status, while controlling for sex, age, region, and if the individual exercises or not.

Introduction

Major Depressive Disorder (MDD) has been ranked as the third largest contributor to the burden of disease worldwide by the World Health Organization in 2008. Prevalence of MDD is projected only to grow, with estimates putting it as the number one contributor by as soon as 2030 (Bains & Abdijadid, 2023). According to the DSM 5 which is a publication by the American Psychiatric Association that is used as the standard for diagnostic criteria in the US, an individual is classified as having MDD when they experience 5 or more depressive symptoms in the same 2-week period nearly every day. Depressive symptoms include depressed mood, loss of pleasure, weight gain or loss, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue, feeling worthless, decreased concentration, and suicidal ideation (American Psychiatric Association, 2013). In 2017, it is estimated that 17.3 million adults in the US, or 7.1% of the population 18 years or older experienced depression in a given year. This additionally doesn't include the 1.9 million children aged 2-17 who have received a MDD diagnosis nor the countless who go undiagnosed for reasons such as stigma, bias, lack of resources or other complex social and personal factors (Depression and Bipolar Support Alliance, 2019).

Despite the widespread and prevalent nature of MDD, its etiology is still not entirely known. Research shows that the cause is likely multifactorial, stemming from genetic, environmental,

and psychosocial factors. As both the limbic system and medial prefrontal cortex play significant roles in mood regulation, it is believed that disruptions to these areas may result in altered mood. While it is not known what mechanisms lead to disruption, it is thought that neurotransmitters and neuronal receptivity are likely contributing factors. This theory, known as the neurotransmitter hypothesis, proposes that the neurotransmitters serotonin and norepinephrine are significantly influential in mood disorders as decreased levels have been associated with episodes of depression. Other theories hypothesize that life events and personality traits are significant factors in the development of MDD as well. One such idea is the learned helplessness theory, which associates the occurrence of depression with the experience of uncontrollable events (Bains & Abdijadid, 2023). With this theory in mind, it could help to inform us why traits such as low self-esteem, physical/sexual abuse, death of loved ones, difficult relationships, financial problems, comorbid illnesses, substance use, and being a member of the LGBTQ community are risk factors for MDD (Mayo Clinic, 2022).

As divorce, loss of a spouse, or lack of a support system such as that provided through a marriage can be a stressful life events that sometimes lead to significant and uncontrollable changes in one's life, as per the learned helplessness theory, this paper seeks to find if there is data supporting a link between depression and marital status in the US. Currently, there is still much to be learned about Major Depressive Disorder as well as what can be done to treat or prevent it. The complex social factors and circumstances that lead to the development of MDD in an individual can vary greatly, but research into various causative factors can help us to gain a better understanding of this disease. Hopefully, through research such as this, we will be able to gain more insight about what factors could be predictors for the development of MDD and in

turn take measures to circumvent its development or catch it early by offering the proper resources to at-risk individuals.

Methods

Data

This study is held to observe the relationship between depression and marital status among adults ages 18 - 99 in 50 states and the District of Columbia of the United States. The sample of this population is taken from CDC - 2014 BRFSS Survey Data.

BRFSS collects data from all 50 states in the US as well as three US territories by taking surveys about the individual's health-related risk behaviors, chronic health conditions, and use of preventive services (*CDC - BRFSS - BRFSS Frequently asked questions (faqs)* 2018). Every year over 400 thousand individuals are surveyed. The BRFSS surveys are commonly used in health-related studies. One example is a NIH report on healthcare and CPS usage in the population (Okoror et al., 2017).

For this study, 443,724 (95.49%) of the 464,664 participants in the 2014 BRFSS survey met the target population: they were aged between 18 – 99 years, reside in one of the 50 states or in the District of Columbia, and had complete data for all variables used in the model.

Model

The research objective of this study is to investigate the association between depression and marital status, and the model can be summarized as follows:

DEPRESS =f(MARITAL1, AGE1, SEX1, REGION, EXERCISE)

where DEPRESS represents whether a BRFSS survey participant reported they have been diagnosed with depression, including major depression, dysthymia, or minor depression(0 = no; 1 = yes); MARITAL1 represents whether the individual is married or not (0 = not married; 1 = married); AGE1 represents age and is coded as 0,1, or 2 depending on whether the respondent is aged 18-44 years, 46-69 years, or 70-99 years old; SEX1 represents sex and is the reported biological sex of the respondent (0 = male; 1 == female); REGION is coded as 0,1,2 or 3 depending on whether the respondent is from the Northeast region, Midwest region, South region, or the West region; and EXERCISE indicates whether the respondent participated in any physical activities of exercise such as running, calisthenics, golf, gardening, or walking in the past month (0 = no; 1 = yes).

Methods: Statistical Analysis

To perform the statistical analysis in this study, both tests of association and logistic regression were utilized. As shown in Table 1, Pearson X^2 tests of association were conducted between the control variables and the exposure variable, along with univariate statistics. Similar methods were applied to the data in Table 2, where the relationships between the control, exposure, and outcome variables, as well as the univariate statistics, were analyzed.

To estimate the adjusted odds ratio with 95% confidence intervals for the outcome variable (DEPRESS) in relation to the exposure (MARITAL1) and control variables (AGE1, SEX1, REGION, EXERCISE), logistic regression was employed, as shown in Table 3. Additionally, regression analysis included tests for confounding between the exposure and control variables,

reporting of goodness-of-fit statistics, and investigation of interactions between the exposure and control variables. All analyses were conducted using SAS.

Results

Of the 464,664 BRFSS participants, 443,724 (95.49%) had complete data for the objective. The demographic characteristics of this population are compared in Table 1 with respect to the exposure variable, whether the participant is married (MARITAL1).

Table 1. Characteristics of 443,724 BRFSS 2014 participants by Marital Status

	Population		Not Married		Married		
Variable	N	%	n	%	n	%	p value *
	443,724	100	205,031	46.21	238,693	53.79	
Sex							
Male	184,859	41.66	75,503	36.83	109,356	45.81	
Female	258,865	58.34	129,528	63.17	129,337	54.19	<.0001
Region							
Northeast	80,112	18.05	39,133	19.09	40,979	17.17	
Midwest	123,470	27.83	54,108	26.39	69,362	29.06	
South	132,903	29.95	64,224	31.32	68,679	28.77	
West	107,239	24.17	47,566	23.2	59,673	25.00	<.0001
Age							
18 – 44 years	116,621	26.28	59,581	29.06	57,040	23.90	
45 – 69 years	221,874	50.00	86,032	41.96	135,842	56.91	
70 – 99 years	105,229	23.71	59,418	28.98	45,811	19.19	<.0001

Exercise							
No	107,140	24.15	58,419	28.49	48,721	20.41	
Yes	336,584	75.85	146,612	71.51	189,972	79.59	<.0001

* p values based on Pearson chi-square test of association

Of the entire population, 58.34% were female, 29.95% were from the South, 27.8% were from the Midwest, 24.17% were from the West, 50% were aged 45-69 years, 75.85% exercised, and 53.79% were married.

There were proportionately fewer females than expected in the married group (54.19% vs 58.34%; $p < .0001$). With respect to region, there were proportionately more individuals living in the Midwest in the married group than expected (29.06% vs 27.83%; $p<.0001$), proportionately fewer individuals living in the South region in the married group than expected (28.77% vs 29.95; $p<.0001$), and proportionately more individuals living in the West region in the married group than expected (25% vs 24.17%; $p<.0001$).

With respect to age, there were proportionately more individuals aged 45 - 69 years in the married group than expected (56.91% vs 50%; $p<.0001$) but proportionately fewer individuals aged 70 – 99 years in the married group than expected (19.19% vs 23.71%; $p<.0001$). With respect to exercise, there were proportionately more exercisers than expected in the married group (79.59% vs 75.85%; $p<.0001$).

The demographic characteristics of this population are compared in Table 2 with respect to the outcome variable, diagnosis of depression (DEPRESS).

Table 1. Characteristics of 443,724 BRFSS 2014 participants by presence of Depression

Variable	Population N	%	Depression - No n	%	Depression - Yes n	%	p value *
	443,724	100	358,126	80.71	85,598	19.29	
Marital Status							
Not Married	205,031	46.21	156,829	43.79	48,202	56.31	
Married	238,693	53.79	201,297	56.21	37,396	43.69	<.0001
Sex							
Male	184,859	41.66	159,030	44.41	25,829	30.17	
Female	258,865	58.34	199,096	55.59	59,769	69.83	<.0001
Region							
Northeast	80,112	18.05	64,149	17.91	15,963	18.65	
Midwest	123,470	27.83	100,722	28.12	22,748	26.58	
South	132,903	29.95	106,554	29.75	26,349	30.78	
West	107,239	24.17	86,701	24.21	20,538	23.99	<.0001
Age							
18 – 44 years	116,621	26.28	94,457	26.38	22,164	25.89	
45 – 69 years	221,874	50.00	172,880	48.27	48,994	57.24	
70 – 99 years	105,229	23.71	90,789	25.35	14,440	16.87	<.0001
Exercise							
No Exercise	107,140	24.15	79,188	22.11	27,952	32.65	
Exercised	336,584	75.85	278,938	77.89	57,646	67.35	<.0001

* p values based on Pearson chi-square test of association

Overall, 19.29% of the entire population had a diagnosis of depression. There were proportionately fewer individuals married than expected who had depression (43.69% vs

53.79%; p<.0001) and proportionately more females who were depressed than expected (69.83 vs 58.34%; p<.0001). With respect region, there were proportionately fewer depressed individuals who lived in the Midwest than expected (26.58% vs 27.83%; p<.0001), proportionately more depressed individuals who lived in the South than expected (30.78% vs 29.95%; p<.0001), and proportionately fewer depressed individuals who lived in the West than expected (23.99% vs 24.17%; p<.0001).

With respect to age, there were proportionately more individuals aged 45 – 69 years who were depressed than expected (57.24% vs 50%; p<.0001) but proportionately fewer individuals aged 70 - 99 years who were depressed than expected (16.87% vs 23.71%; p<.0001). With respect to exercise, there were proportionately fewer exercisers who were depressed than expected (67.35% vs 75.85%; p<.0001).

Adjusted odds ratio for a diagnosis of depression with respect to the exposure and control variables obtained from the logistic regression are presented in Table 3.

Table 3. Logistic regression analysis comparing the adjusted odds ratio of depression in 443,724 BRFSS 2014 participants

	Depression - No		Depression - Yes		OR*	95% CI
Variable	n	%	n	%		
	358,126	80.71	85,598	19.29	---	---
Marital Status						
Not Married	156,829	43.79	48,202	56.31	---	---
Married	201,297	56.21	37,396	43.69	0.600	0.590 - 0.609
Sex						
Male	159,030	44.41	25,829	30.17	---	---
Female	199,096	55.59	59,769	69.83	1.819	1.790 - 1.849
Region						
Northeast	64,149	17.91	15,963	18.65	---	---

Midwest	100,722	28.12	22,748	26.58	0.937	0.916 - 0.959
South	106,554	29.75	26,349	30.78	0.966	0.944 - 0.988
West	86,701	24.21	20,538	23.99	1.008	0.985 - 1.032
Age						
18 - 44	94,457	26.38	22,164	25.89	---	---
45 - 69	172,880	48.27	48,994	57.24	1.216	1.194 - 1.239
70 - 99	90,789	25.35	14,440	16.87	0.556	0.543 - 0.569
Exercise						
No Exercise	79,188	22.11	27,952	32.65	---	---
Exercised	278,938	77.89	57,646	67.35	0.591	0.581 - 0.601

* 95% confidence intervals are for reported odds ratio.

Those who reported being married had 40% fewer odds of reporting depression compared to individuals not married after controlling for sex, region, age, and exercise (OR = 0.6; 95% CI = 0.590 – 0.609). Females had almost double the odds of reporting depression compared to the males after controlling for marital status, age, region, and exercise (OR = 1.819; 95% CI = 1.790 – 1.849).

With respect to regions, those who lived in the Midwest had 6.3% fewer odds of reporting being depressed compared to the Northeast after controlling for marital status, sex, age, and exercise (OR = 0.937; 95% CI = 0.916 – 0.959). Those who lived in the South had 3.4% fewer odds of reporting being depressed compared to the Northeast after controlling for marital status, sex, age, and exercise (OR = 0.966; 95% CI = 0.944 – 0.988). Those who lived in the West had slightly higher odds of reporting being depressed compared to the Northeast after controlling for marital status, sex, age, and exercise but the OR was not statistically significant (OR = 1.008; 95% CI = 0.985 – 1.032). But the difference is not statistically significant.

With respect to age, those aged 45 – 69 years had greater odds of reporting being depressed compared to those aged 18 – 44 years after controlling for marital status, sex, region, and

exercise (OR = 1.216; 95% CI = 1.194 – 1.239). Those aged 70 – 99 years had lower odds (44.4% lower) of reporting depression compared to those aged 18 – 44 years after controlling for marital status, sex, region, and exercise (OR = 0.556; 95% CI = 0.543 – 0.569).

With respect to exercise, individuals who exercised had lower odds of reporting depression compared to those who did not exercise after controlling for marital status, sex, region, and age (OR = 0.591; 95% CI = 0.581 – 0.601).

The AUC statistic for the logistic regression model was 0.6275 which is categorized as poor discrimination according to the Hosmer and Lemeshow rule. The rescaled R^2 value for the model was 0.0514. The deviance test statistic had a p-value of <.0001. This points out that the current model did not fit well with the data presented. So, a more saturated model was made below using backward elimination method:

$$\text{DEPRESS} = \langle + ®_1\text{MARITAL1} + ®_2\text{SEX1} + ®_3\text{REGION}_1 + ®_4\ \text{REGION}_2 + ®_5\text{REGION}_3 + ®_6\text{AGE1}_1 + ®_7\text{AGE}_2 + ®_8\text{EXERCISE} + ®_9\text{MARITAL1*SEX1} + ®_{10}\text{MARITAL1*AGE1} + ®_{11}\text{MARITAL1*AGE1} + ®_{12}\text{MARITAL1*EXERCISE} + ®_{13}\text{SEX1*REGION}_1 + ®_{14}\text{SEX1*REGION}_2 + ®_{15}\text{SEX1*REGION}_3 + ®_{16}\text{SEX1*AGE1}_1 + ®_{17}\text{SEX1*AGE}_2 + ®_{18}\text{SEX1*EXERCISE}$$

The backward elimination method used 0.05 as the benchmark for the p-values. If the p-values were greater than 0.05, the variable would be deemed non-significant. Only variables with a p-value less than 0.05 were left in the new model. Three-way interactions were not tested for this model to keep the models less complicated. The deviance test statistic for this new model had a p-value of <0.0001, indicating the model is statistically significant at an alpha value of 0.05; however, further testing for additional interactions may be required to fully assess the model's

adequacy. The AUC statistic for the new model is 0.6461 and a rescaled R^2 value of 0.0677 indicating a slight improvement from the original model with no two-way interactions. For future analysis, three-way interactions should be used. The original model had an AIC value of 416,893.15 while the new two-way interaction model yielded an AIC value of 416,076.45 indicating the new model is a better fit model. Following the new AUC, rescaled R^2 value, and AIC value of the new model, the new model including the two-way interaction variables will be used as the main model for analysis.

As seen in the model, MARITAL1(exposure variable) interacts with all the control variables except for REGION. These interactions are statistically significant at the 0.05 level. This indicates that the relationship between MARITAL1 and the outcome variable (DEPRESS) is modified by these control variables. So, it is necessary to examine the OR values within for all eight variables that is categorized with AGE1, SEX1, and EXERCISE with respect to DEPRESS. For simplicity, only the range of these OR values are listed below in Table 4.

Table 4 OR for DEPRESS with respect to MARITAL1 accounting for effect modification with SEX1, AGE1, and EXERCISE

MARITAL1	OR Range
No	----
Yes	0.460 – 0.908

According to Table 4, the odds of married individuals experiencing depression are 9.2% - 54% lower than non-married individuals, depending on their AGE1, SEX1, and EXERCISE category.

Finally, the control variables (AGE1, EXERCISE, REGION, SEX) were tested to see if any of these were confounding with MARITAL1, the exposure variable. The 10% rule was used to investigate the confounding between the exposure variable and the control variables. Each control variable was removed from the model to observe if the OR of the exposure variable would change by 10% with each removal. EXERCISE changed the OR value by around 20% when removed, indicating that it is a confounding variable and should be controlled for in the final model to avoid biased estimates of the effect of MARITAL1 on the outcome.

Strengths and Limitations

Strengths

The data for this study was obtained from the Centers for Disease Control and Prevention (CDC). As CDC works with the government, it is certain the data is trustworthy and the analysis using the data should also represent the target population close to accurate. Also, the number of samples that were within the study's target population was an enormous number (443,724 samples) which helped the study's credibility statistically. This is because there was enough representation for each subgroup of this study to insure the representation of the target population.

Limitations

One limitation for this study was the year the BRFFS survey was taken by the CDC. The dataset used for this study originated from 2014 which is relatively old data. The population distribution might have shifted in the US compared to recent years which might affect the conclusion's validity in 2023. Another limitation this study held is that Puerto Rico and Guam was not included. Because the REGION represented four regions (Northeast, Midwest, South, and West),

Puerto Rico and Guam were not included. By including the two territories, it is possible to analyze the US completely. MARITAL1 was also limited into two categories, yes, whether an individual is married or not married. If MARITAL1 was allowed to be divided into further categories, the analysis would have given a different and more in-depth analysis of the depression in US. For example, the "no" category could have been divided into "divorced" and "widowed" and other categories. Lastly, DEPRESS variable only asked if the individual was diagnosed with depression before. There is no set time frame for when they were diagnosed so it is difficult to observe if the individual has an ongoing depression.

Conclusion

The objective of this study was to investigate the association between depression and marital status, while controlling for sex, age, region, and if the individual exercises or not. The study had a target population for adults aged 18-99 years and lived in the 50 states or District of Columbia in the US. The data was taken from CDC's BRFSS survey from 2014.

The study found that the individuals married had 9.2% - 54% fewer odds of reporting depression compared to individuals not married depending on the individual's sex, exercise habit, and age.

This study would further improve the comprehension of the cause of depression. By understanding how marital status and depression were connected, scientists can further research what is different between married and unmarried individuals, how this difference could lead to depression, and possibly why emotional support aides softening the blow of depression. These studies would further improve the prevention and cure of depression, which is a mystery that still needs to be solved.

Though this current study has dived into correlation between marital status and depression, there are alternate ways to dig into the problem of depression. For example, distinct control variables can be used to comprehend the relationship between marital status and depression such as substance abuse. The exposure variable can also be altered to observe depression from a different angle. Substance abuse can also be used as an exposure variable to understand depression. The variable selections should be thought out thoroughly to improve the statistical analysis.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Bains, N., & Abdijadid, S. (2023, April 10). *Major depressive disorder - statpearls - NCBI bookshelf*. National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK559078/

Centers for Disease Control and Prevention. (2018, January 2). *CDC - BRFSS - BRFSS Frequently asked questions (faqs)*. Centers for Disease Control and Prevention. https://www.cdc.gov/brfss/about/brfss_faq.htm

Depression and Bipolar Support Alliance. (2019, July 12). *Depression statistics*. Depression and Bipolar Support Alliance. https://www.dbsalliance.org/education/depression/statistics/

Mayo Foundation for Medical Education and Research. (2022, October 14). *Depression (major depressive disorder)*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007

Okoro, C. A., Zhao, G., Fox, J. B., Eke, P. I., Greenlund, K. J., & Town, M. (2017). Surveillance for Health Care Access and Health Services Use, Adults Aged 18-64 Years - Behavioral Risk Factor Surveillance System, United States, 2014. Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002), 66(7), 1–42. https://doi.org/10.15585/mmwr.ss6607a1

Appendix: SAS Code for Data Preparation and Analysis

```
/* Step 1: Import Dataset and Initial Setup */
libname mylib xport '/home/u63557237/my_courses/ANA625/data/LLCP2014.XPT';
data sasdataset;
   set mylib.LLCP2014;
run;

data one;
   set sasdataset;
run;

proc contents data=one;
run;

/* Frequency Distribution for Key Variables */
proc freq data=one;
   tables ADDEPEV2 MARITAL SEX _STATE _AGEG5YR;
run;

/* Step 2: Data Cleaning and Preparation */
data two;
   set one;
   proc sql;
      delete from two where ADDEPEV2 in (7, 9);
```

```
    quit;
run;

data two;
    set two;
    if not missing(ADDEPEV2) and not missing(MARITAL);
run;

data three;
    set two;
    proc sql;
        delete from three where MARITAL = 9;
    quit;
run;

data four;
    set three;
    proc sql;
        delete from four where _STATE in (66, 72);
    quit;
run;

/* Step 3: Recoding Variables */
proc freq data=four;
    tables ADDEPEV2 MARITAL SEX _STATE _AGEG5YR;
```

```
run;

data five;
   set four;
   if ADDEPEV2 = 1 then DEPRESS = 1;
   else if ADDEPEV2 = 2 THEN DEPRESS = 0;
run;

data six;
   set five;
   if MARITAL = 1 then marital1 = 1;
   else marital1 = 0;
run;

data seven;
   set six;
   if SEX = 1 then SEX1 = 0;
   else if SEX = 2 then SEX1 = 1;
run;

data eight;
   set seven;
   if _STATE in (9, 23, 25, 33, 44, 50, 34, 36, 42) then Region = 0;
   else if _STATE in (17, 18, 26, 39, 55, 19, 20, 27, 29, 31, 38, 46) then Region = 1;
```

```
   else if _STATE in (10, 12, 13, 24, 37, 45, 51, 54, 1, 21, 28, 47, 5, 22, 40, 48, 11) then
Region = 2;

   else if _STATE in (4, 8, 16, 30, 32, 35, 49, 56, 2, 6, 15, 41, 53) then Region = 3;

run;


proc contents data=eight;

run;


proc freq data=eight;

   tables _AGEG5YR;

run;


data eight;

   set eight;

   proc sql;

      delete from eight where _AGEG5YR = 14;

   quit;

run;


proc freq data=eight;

   tables EXERANY2;

run;


data nine;

   set eight;
```

```
    proc sql;
        delete from nine where EXERANY2 in (7, 9);
    quit;
run;

proc freq data=nine;
    tables EXERANY2;
run;

proc sql;
    create table ten as
    select * from nine where not missing(EXERANY2);
quit;

proc freq data=ten;
    tables EXERANY2;
run;

data eleven;
    set ten;
    if EXERANY2 = 2 then EXERCISE = 0;
    else if EXERANY2 = 1 then EXERCISE = 1;
run;

proc freq data=eleven;
```

```
    tables EXERCISE;
run;

/* Step 4: Age Recoding */
data final;
    set eleven;
    if _AGEG5YR in (1, 2, 3, 4, 5) then AGE1 = 0;
    else if _AGEG5YR in (6, 7, 8, 9, 10) then AGE1 = 1;
    else if _AGEG5YR in (11, 12, 13) then AGE1 = 2;
run;

proc freq data=final;
    tables marital1;
run;

proc contents data=final;
run;

/* Step 5: Analysis */
/* Frequency of Variables */
proc freq data=final;
    tables DEPRESS marital1 SEX1 Region AGE1 EXERCISE;
```