What factors impact maternal mental health outcomes?

Final Report  
  
Adebisi Akinyemi

# Introduction

Between 1987 and 2018, pregnancy-related mortality (maternal mortality during pregnancy, delivery, and up to a year postpartum) has increased in the United States. In 2018, the pregnancy-related mortality rate was 17.3 per 100,000 live births compared to 7.2 in 1987 (Centers for Disease Control and Prevention, 2022). Therefore, the pregnancy-related mortality rate in the United States has become a significant public health issue. However, the reasons for the overall increase in pregnancy-related mortality are still unclear. One factor that has emerged is mental health among mothers (Chin et al., 2022). In fact, mental health conditions, including substance use and suicide, were leading causes of pregnancy-related deaths from 2017 to 2019 (Trost et al., 2022). This makes sense considering that U.S. women of reproductive age had the highest rate of mental health needs (Gunja et al., 2022). To understand causes for the increasing pregnancy-related mortality rates in the U.S., it is vital to examine factors related to maternal mental health. In particular, I was interested in examining the most important predictors of maternal mental health outcomes.

# Business Impact

Elucidating the factors that predict maternal mental health outcomes will help identify mothers at most risk for pregnancy-related mortality so that they can be targeted for mental health interventions. It will also help identify variables on which to intervene and allocate resources to prevent or reduce mental health conditions among mothers. Identifying mothers most at risk as well as variables on which to intervene will help save lives, promote public health, reduce healthcare costs, and ensure proper healthcare resource allocation. These impacts will be beneficial for any businesses that provide maternal physical or mental healthcare as well as governments and our overall society. In addition, saved lives and reduced healthcare costs may promote the economy through an increased number of workers and improved productivity. By determining predictors of maternal mental health and highlighting mothers to target for aid and variables on which to intervene, future generations may be supported through living mothers and more effective parenting.

# Data

Dataset: National Health Interview Survey

Source: <https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>

Description:

* 30,000+ interviews were conducted per year since 1957 by staff at the U.S. Census Bureau.
* The NHIS questionnaire was redesigned in 2019 to better meet the needs of data users.
* I used data from 2019 to 2021.
* Variables in this dataset included:
  + Chronic conditions (Hypertension, High cholesterol, Cardiovascular conditions, Asthma, Cancer, Diabetes, Other chronic conditions, Height and weight)
  + Functioning and Disability (Vision, Hearing, Mobility, Communication, Cognition, Self-care and upper-body limitations, Anxiety, Depression, Social functioning)
  + Health Insurance (Coverage status, Sources of coverage, Characteristics of coverage, Continuity of coverage, Reasons for no health insurance)
  + Health Care Access and Use (Primary and urgent care, Financial barriers to care, Prescription medication, Flu and pneumonia immunization, Mental health care)
  + Health-related Behaviors (Cigarettes and E-cigarettes, Physical activity, Walking, Sleep, Fatigue, Smoking history and cessation, Alcohol use)
  + Demographics (Race and Hispanic origin, Marital status, Sexual orientation, Veterans status, Nativity, Schooling, Employment, Family income, Food-related program participation, Housing, Telephone use)

A major strength of this dataset was that it included many participants and many variables from which to choose for my model. Particularly, it contained many maternal health variables and variables about pregnancy, birth, and health insurance. This was also a disadvantage because I had to determine which variables to keep or leave out to limit the number of variables in my model. Another disadvantage of the dataset was that it did not have a direct measure of prenatal care which was likely an important predictor of my outcomes of interest. However, the dataset included multiple variables that were used as proxies for this measure (e.g., health insurance, doctor visits).

# Data Analysis & Computation

Data Wrangling & Cleaning

The data were sourced from the Centers for Disease Control and Prevention website

(<https://www.cdc.gov/nchs/nhis/2020nhis.htm>). They were freely available to be downloaded as CSV files and I downloaded data from 2019, 2020, and 2021. This was because the NHIS questionnaire was redesigned in 2019 and the 2022 data were unavailable. First, I found the relevant variables I wanted to use in the dataset codebook. Next, I loaded the three datasets (one CSV file per year) into Python and selected the relevant variables. These variables included survey year, random id, region, age, sex, education, race, sexual orientation, marital status, income-poverty ratio, nativity, food security, WIC benefits (last 12 months), physical health, body mass index (BMI), health insurance (last 12 months), medical care (last 12 months), live birth (last 12 months), currently pregnant, social participation, anxiety frequency, depression frequency, and mental health care (last 12 months). I then merged the three datasets.

Once I had one final dataset, I changed the names of the variables to be more intuitive and lowercase. Next, I dropped any individuals who had not had a live birth in the past 12 months, because this was a key criterion for my project aims. This left us with only 1,098 individuals remaining in my dataset. I then recoded data for missingness as well as regrouping. That is, I coded “skipped” or “refused” on variables as missing data and regrouped some data so that the variable included only two groups. For example, I recoded the sexual orientation variable which had more than three groups to only “straight” or “lgbt.” I checked the amount of missing data there were in each variable once I had accurate numbers for this due to recoding. Race had the most missing data with 94, while multiple variables had no missing data.

I also checked the amount of missing data for each individual. The maximum number of variables on which an individual was missing data was eight, while many individuals were not missing data on any variables. Next, I checked the data types of all the variables, but only changed the data type for the variable “random id.” Following that change, I checked that the data made intuitive sense (e.g., min, max, average) using “dataframe.describe().” Then, I dropped irrelevant and/or useless variables such as live birth (last 12 months), sex, and random id for the sake of parsimony.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| name\_of\_column | The column’s data format | Brief description of the field. If the field follows a specific format (e.g., a specific date format) include that here too. |
| survey\_year | Int64 | year |
| region | Int64 | census region |
| age | Int64 | age |
| education | Float64 | education |
| race | Float64 | race/ethnicity |
| sex\_orient | Float64 | sexual orientation |
| inc\_pov\_ratio | Int64 | income to poverty ratio |
| marital\_stat | Float64 | marital status |
| us\_born | Float64 | nativity |
| food\_security | Float64 | food security |
| wic\_benefits\_12m | Float64 | received WIC benefits, past 12 months |
| phys\_health | Int64 | general health status |
| bmi | Float64 | body mass index |
| health\_insur | Int64 | have health insurance |
| med\_care\_12m | Int64 | needed medical care but did not get it due to cost, past 12 months |
| pregnant\_now | Float64 | currently pregnant |
| soc\_participation | Int64 | difficulty participating in social activities |
| anx\_frequency | Int64 | frequency of feeling anxious |
| dep\_frequency | Int64 | frequency of feeling depressed |
| mh\_care\_12m | Int64 | received mental health care, past 12 months |

Exploratory Data Analysis

To explore the data, I began by examining descriptive statistics for my potential outcome variables by race. The outcome variables included social participation, anxiety, and depression.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Here, race was coded as 1 White, 2 Black, 3 Asian, 4 American Indian/Alaska Native, and 5 Multiracial. Average anxiety, depression, and social participation did not seem greatly different across these races. Therefore, I wanted to understand the frequency distribution of these potential outcome variables.

Chart

Description automatically generated with medium confidence

Chart, bar chart

Description automatically generated

Chart, histogram

Description automatically generated

These distributions demonstrated that social participation was severely positively skewed with little to no variation in the data. Conversely, anxiety and depression were negatively skewed but included sufficient variation. Thus, I later decided to drop social participation as an outcome variable.

Next, I decided to examine variation in data for all the other variables, many of which were categorical variables with only two groups. There were an approximately equal number of people in each year (2019, 2020, 2021). However, most participants (over 700 of the 1,098 participants) identified as heterosexual, were married, were US-born, had health insurance (last 12 months), did not receive required medical care (last 12 months), were not currently pregnant, did not receive WIC benefits (last 12 months), and did not receive mental health care (last 12 months). Mental health care (last 12 months) was a potential outcome variable for a short time, but the lack of variation in the data and the fact that most people (over 800 of the 1,098 participants) had not received mental health care in the last 12 months caused us to drop the variable. For numerical variables, I produced summary statistics and histograms, the results of which are below.

Graphical user interface, application, table

Description automatically generated

Chart, histogram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Chart, histogram

Description automatically generated

Graphical user interface, chart

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Chart

Description automatically generated with medium confidence

Graphical user interface, chart, application

Description automatically generated

Chart, histogram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Chart, histogram

Description automatically generated

These histograms showed that age was approximately normally distributed, income-poverty ratio was negatively skewed but included sufficient variation, and physical health and BMI were positively skewed but there was adequate variance in the data. Conversely, food security was severely negatively skewed and contained little variation. The histograms also showed that there was sufficient variation in how many participants resided in each of the four U.S. census regions, education was negatively skewed but contained adequate variance, and most participants (over 700 of the 1,098 participants) self-identified their race as White. Race was later dichotomized to White vs. non-White due to this lack of variation.

After examining descriptive statistics, I explored correlations between my potential outcome variables and all the other variables. Below are the variables with the highest correlations with social participation, anxiety, and depression.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

These correlation coefficients demonstrated that all the potential outcome variables (social participation, anxiety, depression) were highly correlated with one another. Potential numeric predictor variables most highly correlated with social participation were food security, physical health, education, and income-poverty ratio. Anxiety was most highly correlated with physical health and food security, while numeric variables most highly correlated with depression included income-poverty ratio, education, age, physical health, food security, and BMI.

Statistical Analysis

Before I began building my model, I dichotomized my race variable to 0 White, 1 non-White due to the lack of variation in the data – there more than 700 participants who identified as White.

Graphical user interface, text, application

Description automatically generated

Next, I conducted t-tests to examine whether there were racial differences (White vs. non-White) in average anxiety and average depression. At this point, I had dropped social participation as a potential outcome variable due to lack of variance. The variable was severely positively skewed, meaning that a very large majority people did not have any difficulty with their social participation.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

The t-tests showed that average anxiety was greater among non-White mothers, while average depression was not different among non-White and White mothers. I next built ordinary least squares multivariate regression models to identify important predictors of anxiety and depression, respectively. Both models included only theoretically relevant predictors (e.g., race) and variables with sufficient variation.

Graphical user interface

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

The multivariate regression models conducted listwise deletion, meaning that individuals with any missing data were dropped. This is why there were only 945 observations in the models. Both models fit the data well, based on the p-values that correspond to their F statistics. Statistically significant predictors of maternal anxiety were year, age, physical health, and race. The model’s R2 was .08. Similarly, WIC benefits, physical health, and race were the statistically significant predictors of depression. The model’s R2 was .11. These R2 values seemed surprisingly low, so I decided to run additional models for each outcome variable including all relevant variables, regardless of how much variation there was in the variable (except for region). I did not include region as a variable in the second set of models because it was a categorical variable with four groups, which would cause its regression coefficient to not be interpretable.

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

The second set of multivariate regression models also conducted listwise deletion, so there were only 927 observations in them. Like the first set of models, both models fit the data well, based on the p-values that correspond to their F statistics. However, these models explain more of the variation in the outcome variables – based on their R2 values. The R2 of the model predicting maternal anxiety was .12 – .04 units more than the first model predicting maternal anxiety. There was a .05 unit increase in the R2 of the second model predicting maternal depression (R2 = .16) compared to the first model (R2 = .11). Because of these differences in R2 values and the fact that the second set of models also fit the data well, the second set of models were my final models.

The final model predicting maternal anxiety showed that surprisingly, average maternal anxiety decreased over time (year) and as physical health and food security declined. Both physical health and food security were measured such that a higher number indicated worse physical health and lower food security. However, average maternal anxiety was greater on average for non-White and foreign-born mothers. The final model predicting maternal depression also surprisingly showed that average maternal depression decreased as physical health and food security declined. But maternal depression was greater on average for mothers who identified as heterosexual and mothers who did not receive WIC benefits in the last 12 months. Mothers who did not receive required medical care or have health insurance in the last 12 months also experienced greater depression. All other variables were not statistically significant predictors of either maternal anxiety or depression.

# Dashboard

I created an explanatory dashboard using Tableau to visually describe my data and the results of my exploratory data analysis.

Chart

Description automatically generated

The first page of my dashboard includes a line graph of maternal mental health from 2019 to 2021. It shows that average maternal depression increased from 2019 to 2020, which is likely associated with the COVID-19 pandemic. In 2021, average maternal depression dropped back to where it was in 2019. Next, the dashboard demonstrates differences in average maternal depression and anxiety across races and regions. Finally, it includes the total number of participants per year, race, and region in pie charts. Hovering over the graphs on this page will show corresponding numeric values and categories (if applicable). A use case for this page of the dashboard is getting an overall feel for the data, particularly maternal anxiety and depression by interesting variables. This page may also be used to identify a few descriptive statistics.

Graphical user interface, application, table, Excel

Description automatically generated

The second page of my dashboard includes a scatter plot examining the correlation between age and average maternal anxiety and depression. It includes a second scatter plot looking at how income-poverty ratio is associated with maternal anxiety and depression. Next, I use bar graphs to examine how food security, BMI, and physical health are related to maternal anxiety and depression. Surprisingly, we see that as physical health and food security improve, average maternal anxiety and depression increase. Finally, this page of the dashboard includes a bar graph exploring maternal anxiety and depression by education. Like the first page of the dashboard, hovering over the graphs on this page will show exact numeric values. A use case for this page of the dashboard is exploring correlations between maternal anxiety and depression and other key variables. Thus, this page may help a person decide which variables to include in a more complex statistical model.

[DASHBOARD LINK HERE](https://public.tableau.com/app/profile/adebisi.akinyemi/viz/DS4ADashboard1_16743454750780/Overview)

# Conclusion & Future Work

Based on my findings, women at most risk for pregnancy-related adverse mental health (and pregnancy-related mortality) are women of color, women who identify as heterosexual, foreign-born women, and women who did not receive WIC benefits, required medical care, or have health insurance in the last 12 months. Therefore, these are important factors that practitioners should account for as part of holistic care for pregnant women and mothers. For example, pregnant women and recent mothers could be given a pre-care screening which asks about these variables. Then, interventions, more intensive care, and resources could be directed at women who meet the criteria that my findings identified. In addition, health professionals and other relevant stakeholders could advocate on behalf of these women so that they qualify for resources like WIC benefits. These stakeholders could also help these women acquire required medical care. Finally, health professionals could provide special short-term (during pregnancy and delivery) and long-term (one year or more after birth) attention to these women to ensure they receive proper care as soon as any signs or risk factors of mental illness are present.

My final models included only 927 participants and future work should include more participants to give the model more power to detect statistically significant associations (if they are present). This relatively small sample is even more concerning given that my original dataset included approximately 90,000 participants. The massive reduction in the sample makes it highly likely that my sample was no longer representative of the population to which it was recruited to generalize. It is also highly unlikely that my smaller sample (due to my stringent criterion – live birth in the past 12 months) was not representative of the women who had given birth in the last 12 months. Future work should use a sample specifically recruited because they are pregnant or gave birth recently via probability sampling. This will increase the likelihood that the sample is representative of the relevant population. Future work should also include theoretically important predictors such as prenatal care and income. Not having these predictors in my models means that it was likely not correctly specified. The model I used for this project could also be improved by considering variables’ level of measurement. I treated ordinal variables as continuous variables, which could have biased my parameter estimates. Finally, I would have loved to successfully match another dataset with the one I used to investigate region-level mental health outcomes.

# References

Centers for Disease Control and Prevention. (2022, June 22). *Pregnancy mortality surveillance system*. <https://www.cdc.gov/reproductivehealth/maternal-mortality/pregnancy-mortality-surveillance-system.htm#:~:text=16.3%20deaths%20per%20100%2C000%20live,persons%20living%20in%20noncore%20counties>.

Chin, K., Wendt, A., Bennett, I. M., & Bhat, A. (2022). Suicide and maternal mortality. *Current Psychiatry Reports, 24*(4), 239-275. <https://doi.org/10.1007/s11920-022-01334-3>

Gunja, M. Z., Seervai, S., Zephyrin, L., & Williams, R. D. (2022, April 5). *Health and Health Care for Women of Reproductive Age*. The Commonwealth Fund. <https://www.commonwealthfund.org/publications/issue-briefs/2022/apr/health-and-health-care-women-reproductive-age>

Trost, S., Beauregard, J., Chandra, G., Njie, F., Berry, J., Harvey, A., & Goodman, D. A. (2022). *Pregnancy-related deaths: Data from mortality review committees in 36 States, 2017-2019*. <https://reviewtoaction.org/sites/default/files/2022-10/Pregnancy-Related-Deaths-Data-MMRCs-2017-2019-H.pdf>