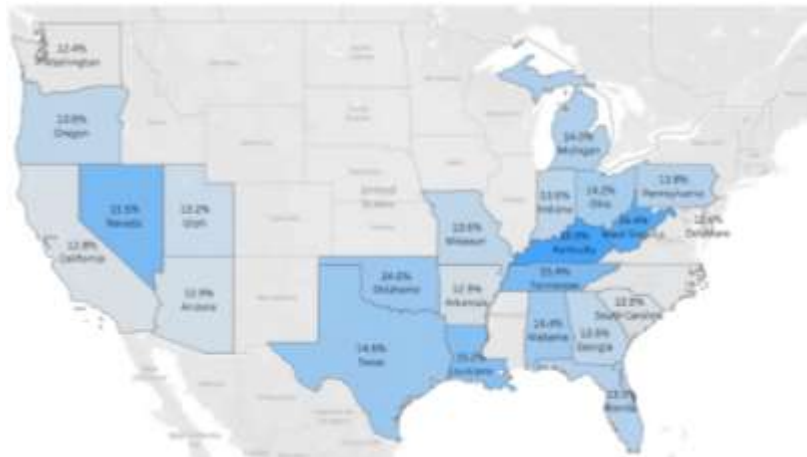


# USA Mental Health Summary 2020 Project

Lindsey Fisher - February 2022

Using the Behavioral Risk Factor Surveillance System Survey this analysis attempts to identify critical drivers of poor mental health among American citizens. A Gradient Boosting Model was chosen as the best classification model with an Area Under the ROC Curve (AUC) of 83%. From this model, the 15 most important predictors were extracted and visualized. With these visualizations, this project attempts to identify population segments that may benefit from additional mental health assistance. The recommendations are detailed below:

- 1) Target mental health support for the following populations
  - a. Under 40, especially the 18-24 segment of young adults. 1 in 5 persons in the 18-24 segment report 14+ days of poor mental health in the past 30 days.
  - b. States experiencing above-average proportion of residents experiencing poor mental health. Average = 12.2%



- c. Citizens who are unable to afford medical care or those who are uninsured
    - d. Citizens who are unable to work or unemployed
    - e. Current smokers
    - f. Persons reporting their race as Native American or Other
    - g. Females
    - h. Separated, divorced, never-married, or long-term unmarried couples
    - i. Persons engaging in high-risk behaviors related to HIV
  - 2) Examine factors that may be both symptoms and drivers of poor mental health
    - a. difficulty concentrating, remembering, or making decisions
    - b. poor physical health
    - c. average sleep time
    - d. difficulty doing errand/other tasks alone

## Overview:

Using the Behavioral Risk Factor Surveillance System from the Centers for Disease Control and Prevention, this project will investigate predictors of poor mental health. The following question from the survey represents the target variable for this analysis: "Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?" To address this issue, we need to identify the population struggling with mental health by defining common characteristics among the at-risk population. By identifying the most important variables driving the reports of more than 14 days of poor mental health, advocates can design effective assistance programs to target struggling populations.

### (1) Data Source:

- a. The Behavioral Risk Factor Surveillance System (BRFSS) from the Centers for Disease Control and Prevention. <https://www.cdc.gov/brfss/>

### (2) Data Description:

- b. The BRFSS is a system of ongoing health-related telephone surveys designed to collect data on health-related risk behaviors, chronic health conditions, and use of preventive services from the noninstitutionalized adult population ( $\geq 18$  years) residing in the United States. The dataset for the BRFSS from 2020 contains 279 variables and 401,958 observations.

### (3) Data Cleaning:

- c. Filtered out partially completed interviews
- d. Remove variables with greater than 75% non-response (NA or Blank)
- e. Removed variables which are not related to the interviewee are redundant or nonsensical
- f. Created a "Don't know or Refused" category for each variable
- g. Factorize all variables

### (4) Define Target Variable:

SAS Variable Name: _MENT14D				
Question Prologue:				
Question: 3 level not good mental health status: 0 days, 1-13 days, 14-30 days				
Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Zero days when mental health not good	255,707	63.62	61.15
2	1-13 days when mental health not good	89,979	22.39	23.66
3	14+ days when mental health not good	49,343	12.03	13.21
9	Don't know/Refused/Missing	7,929	1.97	1.99

### Transforming Target Variable:

- i. I created two levels: the event of interest, which is a report of 14+ days of poor mental health in the past 30 days (1), and the non-event, which was 13 days or less of poor mental health (0). The rationale for this choice is the criteria for diagnosis for Major Depressive Disorder, which entails a report of 2-weeks or more of persistent depressive mood.
- ii. I removed the "don't know/refused/missing" level since it represents only 1.97% of the respondents. The distribution of responses for the target variable is detailed in table 2.

Poor Mental Health: 0-13 days	Poor Mental Health: 14+ Days
87.8%	12.2%

*Table 2: Distribution of the event*

## (5) Machine Learning for Classification Task:

For this task, I tested five machine learning models to predict the target variable: Distributed Random Forest (DRF), Gradient Boosting Machine (GBM), Neural Network (NN), Generalized Linear Model (GLM) with the logistic option selected, and Naïve Bayes (Bayes). The model's performance is summarized below in Figure 3. I used the H2O package in R to utilize parallel computing and speed up the modeling process.

Model	AUC
GBM	0.83
NN	0.825
GLM	0.82
DRF	0.787
Bayes	0.777

*Table 3: AUC performance of each of the ML models*

## Variable Importance

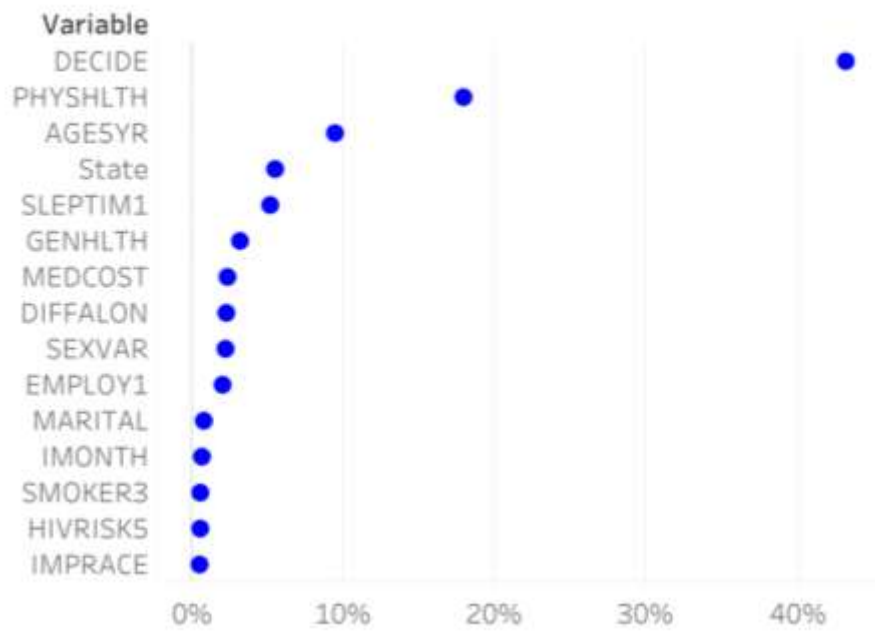


Figure 1: GBM Most Important Variables (Importance above 1%)

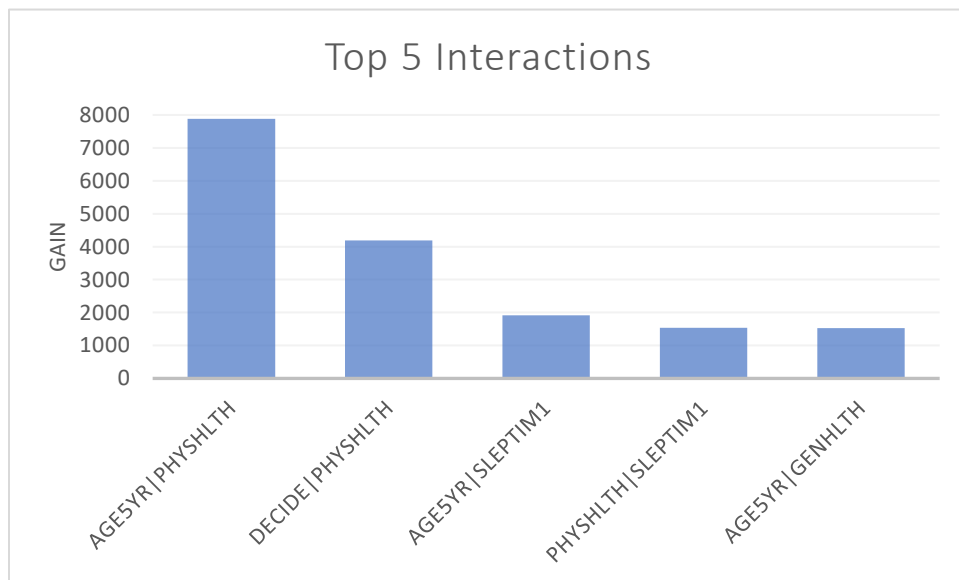


Figure 2: Top 5 Interactions (Ranked by Gain)

## (6) Important Predictors

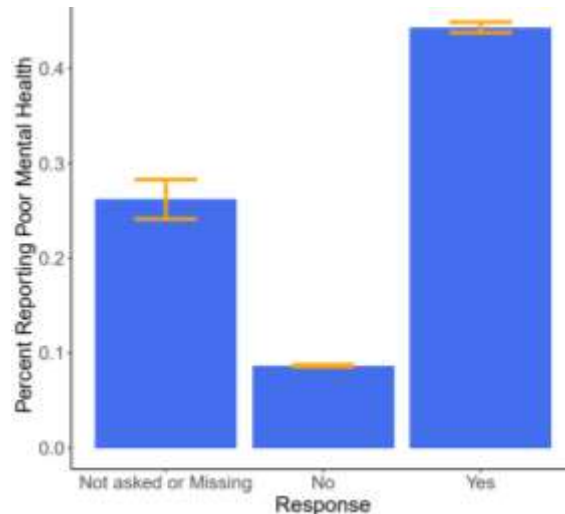
In table 4, we can see the survey question, which corresponds to each important variable identified by the Gradient Boosting Model.

Importance	Variable Code	Survey Question
1	Decide	Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions?
2	Pyshlth	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?
3	Age5year	Reported age in five-year age categories calculated variable
4	State	State of Residence
5	Sleptim1	On average, how many hours of sleep do you get in a 24-hour period?
6	Genhlth	Would you say that in general, your health is: 1=excellent to 5 =poor
7	Medcost	Was there a time in the past 12 months when you needed to see a doctor but could not because of cost?
8	Difffalon	Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone, such as visiting a doctor's office or shopping?
9	SexVar	Sex of Respondent
10	Employ1	Employment Status
11	Marital	Marital Status
12	Imonth	Month of Interview
13	Smoker3	Four-level smoker status: Everyday smoker, Some days smoker, Former smoker, Non-smoker
14	HIVrisk5	I am going to read you a list. When I am done, please tell me if any of the situations apply to you. You do not need to tell me which one. You have injected any drug other than those prescribed for you in the past year. You have been treated for a sexually transmitted disease or STD in the past year. You have given or received money or drugs in exchange for sex in the past year.
15	Imprace	Imputed race/ethnicity value

*Table 4: Survey Questions corresponding to the most important variables*

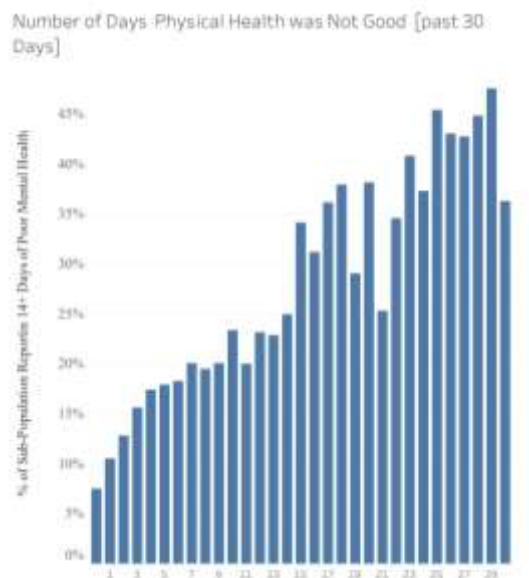
## (7) Visualizing Top Predictors

**Decide:** Because of a physical, mental or emotional condition do you have trouble remembering or making decisions?



Individuals who answered "Yes" to having severe difficulty with memory and making decisions are approximately 5x more likely to report poor mental health for 14 or more days over the past month. According to the Mayo Clinic: "Trouble thinking, concentrating, making decisions and remembering things is a symptom of Major Depressive Disorder."

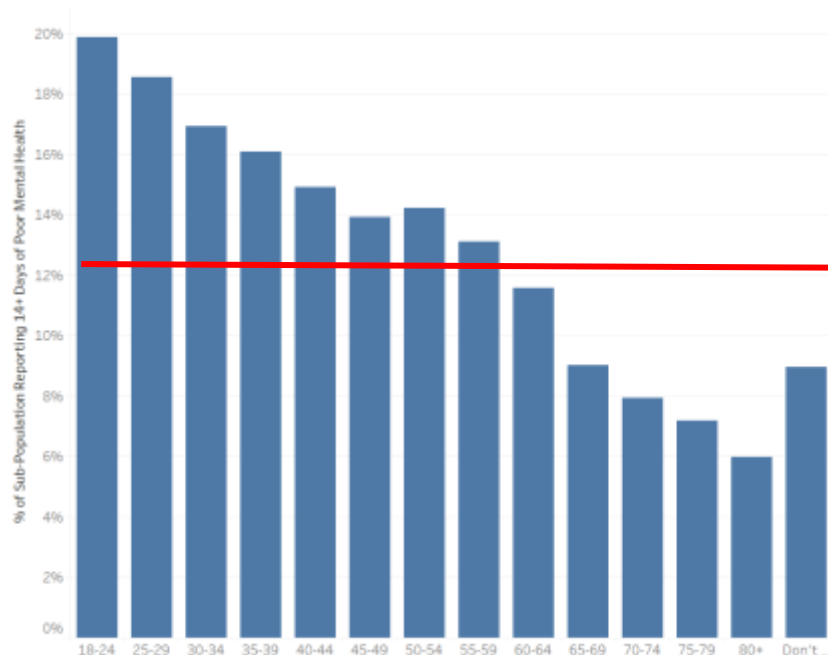
**Physhlth:** During the past 30 days, for how many days was your physical health not good?



As the number of days where physical health was reported to be poor increases, the proportion of individuals reporting poor mental health for 14 or more days over the past month is trending

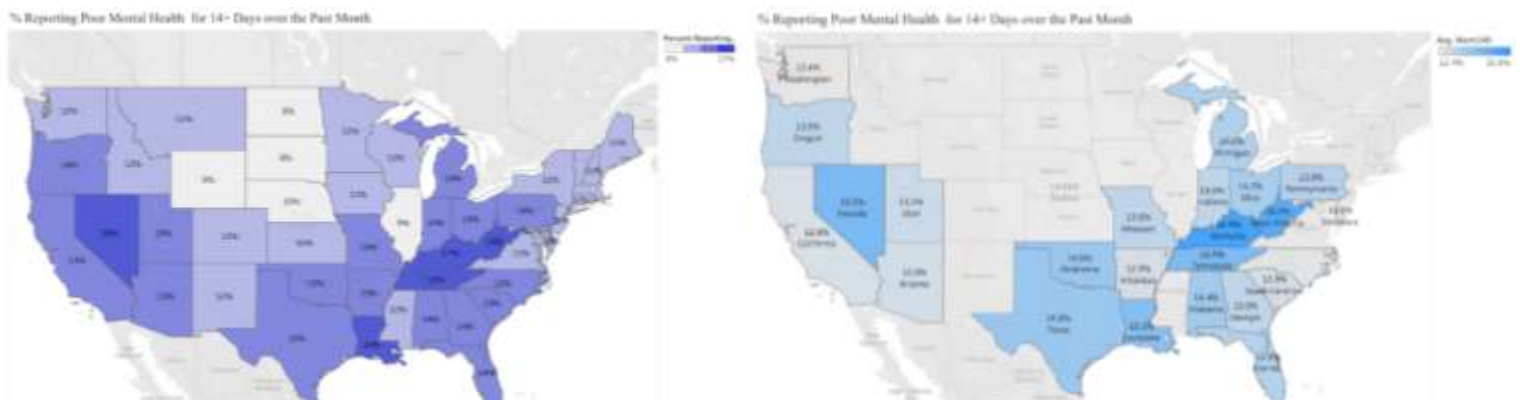
upward. This Graph indicates that there may be a strong relationship between physical pain and mental wellness.

### Age: Age of Respondent



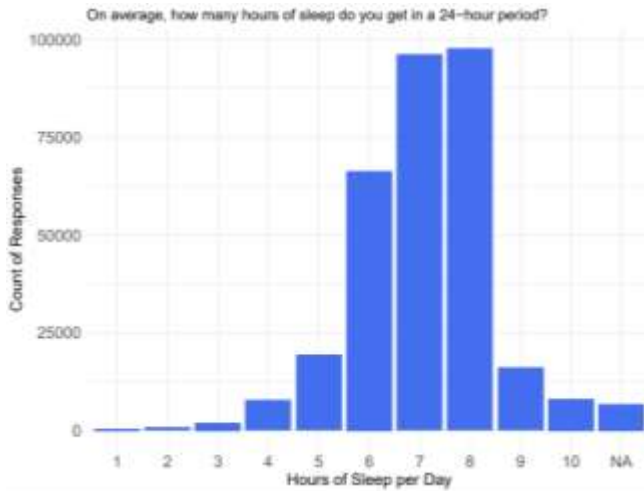
As reported age increases, the proportion of individuals reporting poor mental health for 14 or more days over the past month decreases. Those over the age of 65 are experiencing poor mental health at half the rate of 18-24-year-olds. All age groups less than 60 are above average.

### State: Location of the Respondent

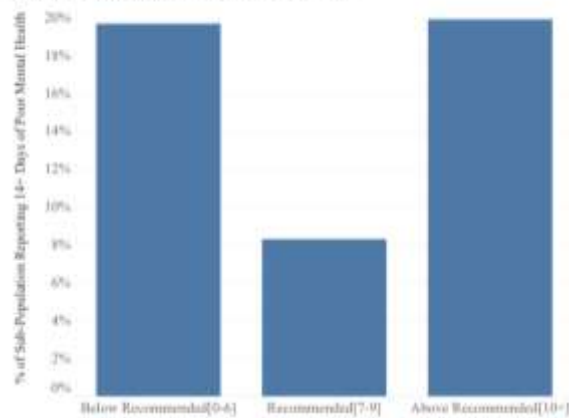


For the map on the left, States with the darkest blue coloring are areas of concern where the highest proportion of residents are reporting two weeks or more of poor mental health. The map on the right represents those states above average (12.2%) for the percent of residents reporting 14+ days of poor mental health.

## Sleptim1: Daily Sleep on Average



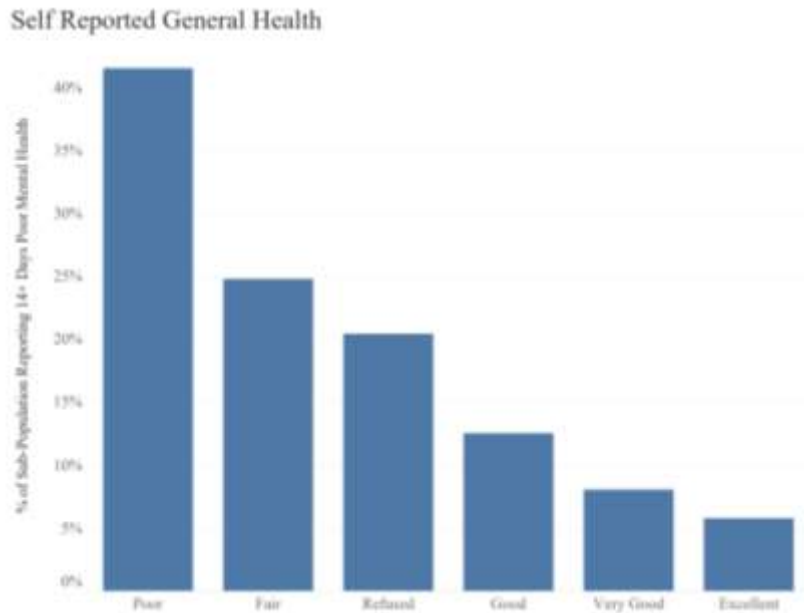
3 Level Calculated Sleep Time Scale



Most of the respondents report getting between 6-8 hours of sleep per night. However, only the respondents getting between 7-9 hours of sleep per night are below the overall average (12.2%). The persons of concern receive less than five or greater than 10 hours per sleep on average. This concern aligns with the mayo clinic guideline that 'sleep disturbances, including insomnia or sleeping too much, are symptoms of major depressive disorder. According to the National Sleep Foundation, "Between the ages of 18 and 64, adults should aim for seven to nine hours of nightly sleep." When the sleep graph is broken down into three buckets: below recommended, recommended, and above recommended, those receiving less than or more than the optimal 7-9 hours are experiencing poor mental health at twice the rate as those who get the recommended amount of sleep.



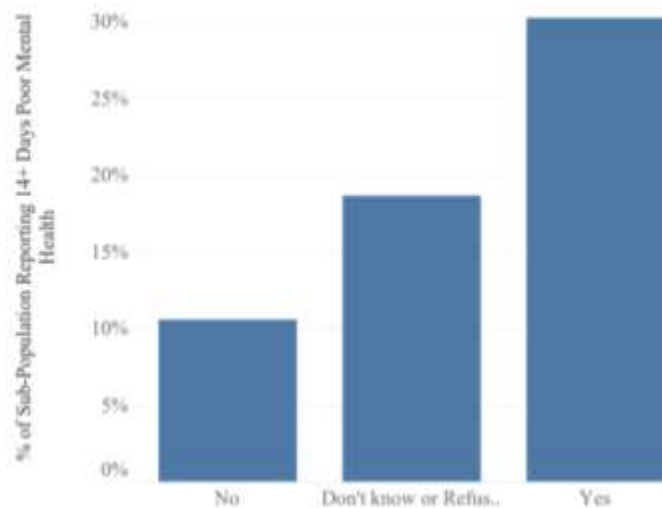
## Genhlth: Self-Reported General Health



Persons experiencing poor general health also experience the highest rate of poor mental health.

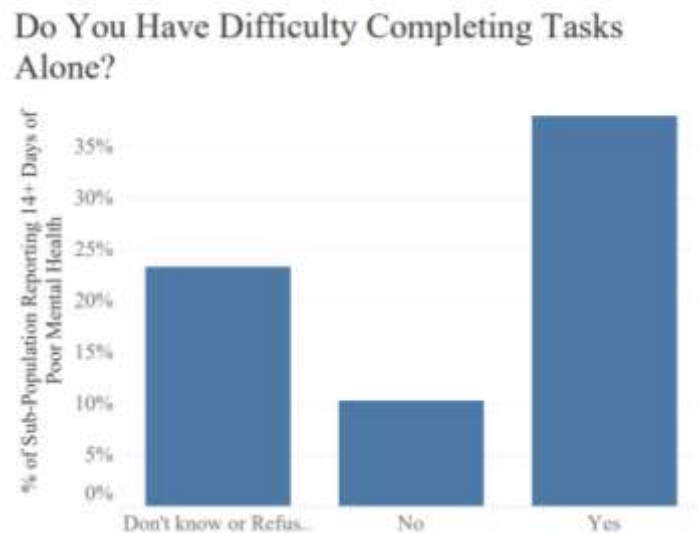
## Medcost: Inability to Afford the Cost of Medical Care

Was there a Time in the Past 12 Months when you needed to see a doctor but couldn't because of cost?



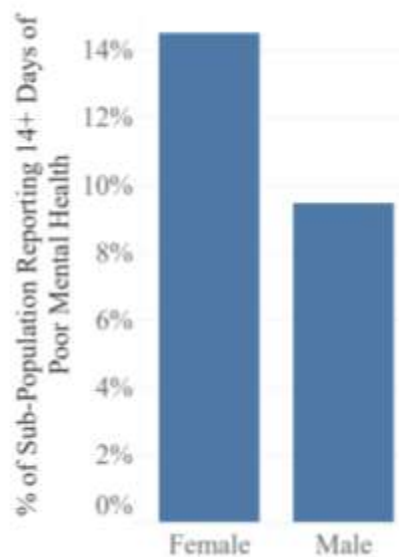
Persons who cannot afford their medical care are experiencing poor mental health at 3x the rate as those who have not had an issue affording care.

### **Diffalon:** Difficulty Doing Tasks Alone



Those who cannot complete daily tasks alone are experiencing poor mental health at more than 3x the rate as those who can complete tasks alone.

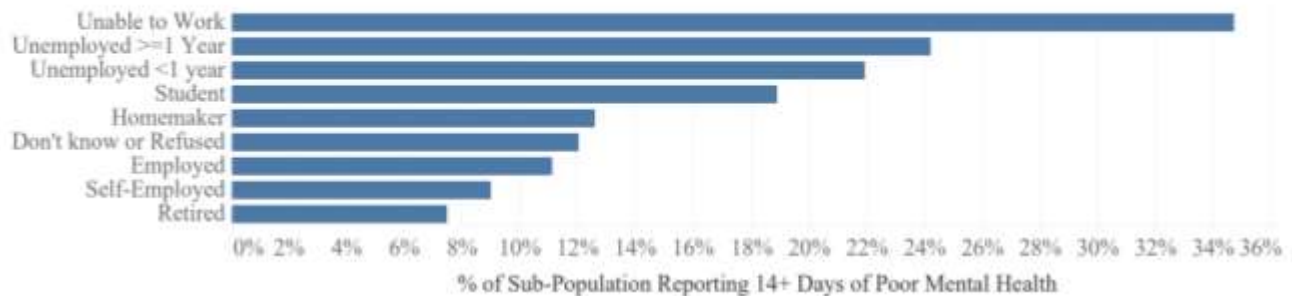
### **Sexvar:** Sex of the Respondent



Females are reporting higher rates of poor mental health than their male counterparts. The American Psychological Association finds that the sexes are different when it comes to mental illness: Women are more likely to be diagnosed with anxiety or depression, while men tend toward substance abuse or antisocial disorders.

## Employ1: Employment Status

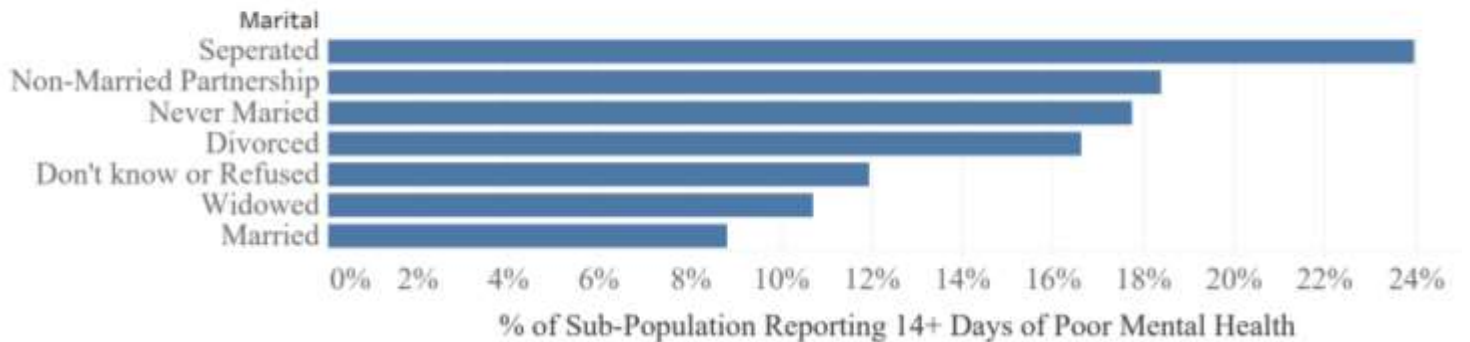
### Employment Status



Those who are unable to work are experiencing poor mental health at 3x the rate compared to those who are employed. Unemployed persons report poor mental health at about 2x the employed rate.

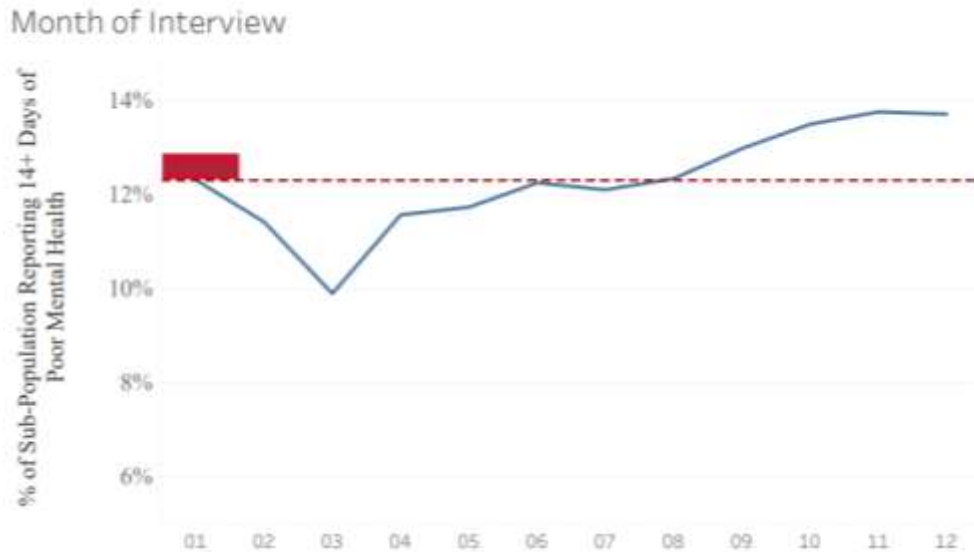
## Marital: Marital Status

### Marital Status



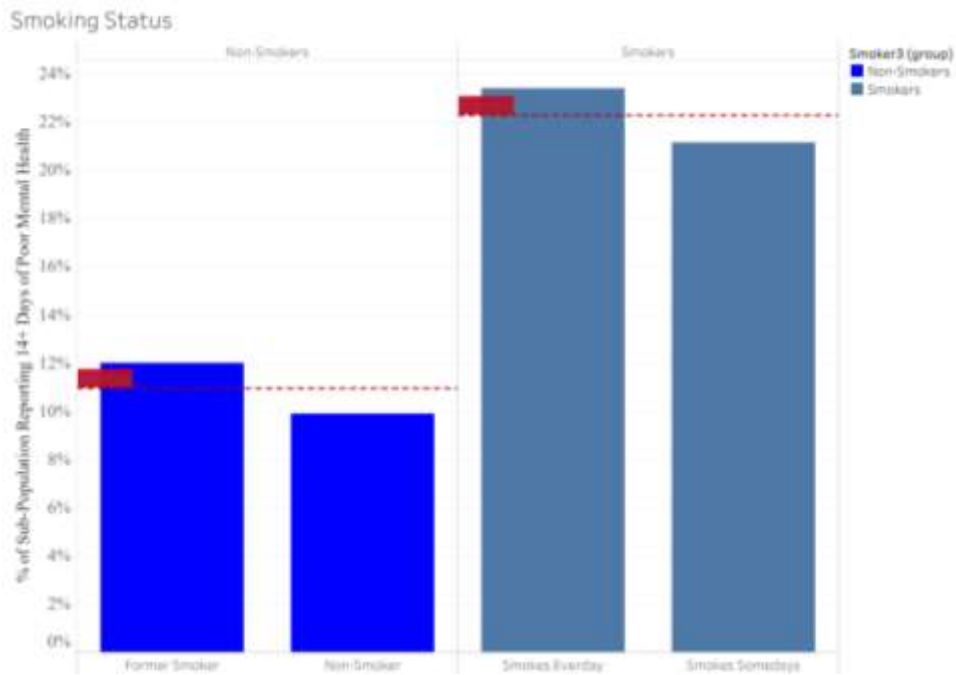
The groups reporting above average (12.2%) poor mental health prevalence are separated partners, non-married partners, persons who were never married, and divorced partners.

## Imonth: Month of Interview



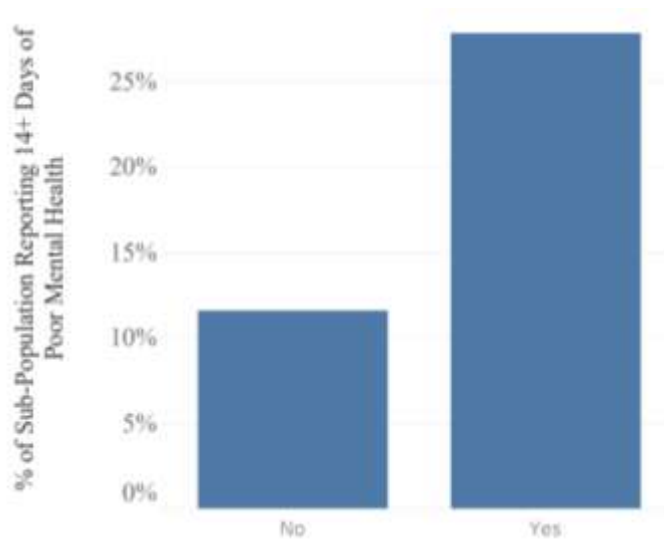
The rate of poor mental health is above average for September-December of 2020. This trend may be due to seasonality or pandemic-related factors or possibly a mix of the two.

## Smoker3: Smoking Status



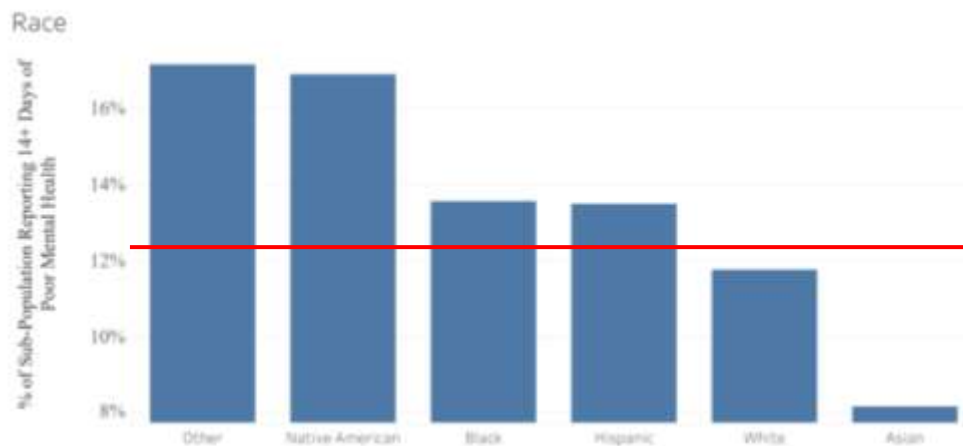
Persons in the smoking group are experiencing poor mental health at 2x the rate versus non-smokers on average. Support for those trying to quit smoking should be an essential part of mental health care.

## HIVrisk5: Engage in Risky HIV Related Behaviors



Persons engaging in risky behavior related to HIV are experiencing poor mental health at 2x the rate as those who are not on average.

## Imprace: Imputed Race Categories



All the minority groups except the Asian group report higher than average poor mental health.