Association of Mental Health Factors with COVID-19 Test Status

Annalise Cramer

**Authors**

* Annalise Cramer (ORCID: 0009-0002-8718-3593)

**Author affiliations**

1. College of Public Health, University of Georgia, Athens, GA, USA.
2. Center for Ecology of Infectious Diseases, Athens, GA, USA.

Corresponding author: atc10260@uga.edu

Disclaimer: The opinions expressed in this article are the author’s own and don’t reflect those of the University of Georgia.

# 1. Summary/Abstract

Isolation and quarantine strategies employed for containment of SARS-Cov-2 infection led to unexpected mental health struggles among the population of the United States. Infected individuals and healthy individuals were affected, but little is known about the potential relationships between mental health and physical health. In this study, we utilized Nataional Health Information Surveys from 2020-2022 to examine the relationships between receieving a positive COVID-19 test and self reporting mental health factors such as meeting recommended guidelines for exercise, meeting the recommended number of hours of sleep per night, reporting feeling depressed recently, reporting having trouble sleeping recently, and reporting receiving less social support.

Analysis is completed using complex survey design to find prevalence of mental health factors. Single models are used to determine associations, and logistic regression, LASSO regression, and a random forest model are explored to determine predictive power.

We find touble sleeping, meeting exercise requirements, and getting enough sleep were common mental health struggles in the United States. Social support was found to vary by COVID-19 test status, and found to be protective againt receiveing a postive COVID-19 test. Trouble sleeping and getting enough sleep were found to very slightly increase chances of recieving a positive COVID-19 test.

# 2. Introduction

## 2.1 General Background Information

When the SARS-CoV-2 pandemic began in late 2019, there were very few control measures enacted within the United States. By Spring of 2020, this had drastically changed, and control measures were broadly put into effect across the country. Social distancing, quantining, isolation, and lockdowns were prioritized control of transmission, but societal impacts were not initially considered. Disruptions to daily routine and social interactions quickly became a widespread concerns for their potential impacts on mental health, with widespread increases in anxiety, depression, loneliness, and sleep disturbances.

Pandemic control strategies have become commonly criticized for the impacts on individuals’ psychological well-being. Research documents the mental health consequences of pandemic control strageies, but the potential relationships between mental health factors and vulnerability to SARS-CoV-2 infection is less understood.

This study seeks to add to the limited understanding of mental health and COVID-19 vulnerability by expliorung associations between mental health factors and COVID-19 test positivity status. Self reported bouts of depressed mood, perceived social support, amount of exercise, and sleep quality (via hours of sleep and reporting difficulty falling asleep) are examined.Simple and multiple regression models including Lasso and Random Forest are used to explore the predicitve power of mental health factors to a positive COVID-19 test.

## 2.2 Description of data and data source

Data for this project is survey data collected from the National Health Interview Survey, produced by IPUMS through the University of Minnesota. This data was collected among United States residents of all ages and genders from 2020 - 2022 in the United States.

## 2.3 Questions/Hypotheses to be addressed

1.) What is the prevalence of poor mental health? 2.) Is poor mental health associated with a positive COVID-19 test? 3.) Can mental health facotrs be used to predict if an individual will recieve a positive COVID-19 test?

Factors studied include: 1.) Average self-reported hours of sleep per night, 2.) Self-reported depressed mood, 3.) Self-reported trouble sleeping, 4.) Self-reported average weekly exercise, and 5.) Self-reported changes to social and emotional support.

# 3. Methods

## 3.1 Schematic of workflow

|  |
| --- |
| Figure 1: Flowchart depicting methodology of this study. |

## 3.2 Data aquisition

National Health Information Survey data was obtained from IPUMS, used with permission. Data was subsetted to years 2020-2022, to cover the brunt of pandemic years. Despite low prevelance of COVID-19 cases in 2019 worldwide, the appearance of COVID-19 in the news (and consequential potential impact on mental health factors) led us to include this data. Additionally, United States citizens were not being widely tested for COVID-19 before 2020. The final dataset contains 151406 observations.

## 3.3 Data import and cleaning

Data was used in 2020 and after, given the NHIS didn’t have questions about COVID-19 testing in 2019, these questions began in 2020. Binary variables were created to allow for interpretability.

EXERCISE is a combination of number of minutes of passive or aerboic exercise, descirbing whether weekly amount meets the American Heart Association’s guidelines (based on Office of Disease Prevention and Health Promotion recommendations) for exercise needed to stay healthy. For adults, this includes 150 minutes per week of moderate intensity aerobic activity, 75 minutes per week of vigorous aerobic activity, or a combination of both.Children ages 6-17 should get 60 minutes of moderate to vigorous intenisty physical activity. A value of 1 for this binary variable means that the individual met exercise recommendations, and a 0 means that the individual did not meet exercise recommendations.

SLEEP is based on the NIH guidelines for recommended hours of sleep per day, which differs by age. Individuals under age 1 are suggested to get 12-16 hours of sleep per day, children ages 1 to 2 are suggested to get 11-14 hours of sleep per day, children ages 3 to 5 years are suggested to get 10-13 hours of sleep per day, children ages 6 to 12 years are suggested to get 9-12 hours of sleep per day, teens ages 13-18 are suggested to get 8-10 hours of sleep per day, and adults (over the age of 18) are suggested to get 7-9 hours of sleep per day. Guidelines do not require that sleep takes place at nighttime, and details that naps count towards sleep totals. A value of 1 for this binary variable means that the individual met recommendations for hours of sleep per day, and a 0 means that the individual did not meet recommendations for hours of sleep recommendations.

DEPRESSED includes a 1 if the individual reports feeling depressed on a weekly or daily basis, but a 0 for feeling depressed at a frequency of monthly or less.

TROUBLE\_SLEEPING includes a 1 if the individual reports struggling to fall asleep several days per week, more than half the days, or nearly every day, and 0 if they report otherwise.

SOCIAL includes a value of 1 for an individual describing feeling like they have less social support than their regular amount, and a 0 for feeling like they recieve the same or more support.

For all variables, answers for don’t know or unsure were converted to missing.

## 3.4 Statistical analysis

After creating binary variables, significant amounts of data were missing, seen in table below. The proportions of values for answers 1 and 0 for binary variables are shown out out total answers, and then a column including NAs are shown, which details the proportions of NA answers among total individuals included.

Multiple imputation was used to create 5 sets of imputed data, and used for the remainder of analysis. To calculate later statistics, processes were preformed on each of the 5 sets and then averaged.

|  |
| --- |
| Table 1: Table depicting porportion of responses among all observations, incliding missingness. |

# 4. Results

## 4.1 Exploratory/Descriptive analysis

To examine prevelance of mental health factors among the United States population, compmlex survey design was used to allow for generalizability of results, specifically, variables PSU, STRATA, and SAMPWEIGHT. A summary table was generated to find occurence of each of the factors and their inverses among the data, seen below.

|  |
| --- |
| Table 2: Table depicting prevalence estimates with 95% confidence intervals among the United State population. Results are split by COVID-19 test status, and shown for overall population. |

## 4.2 Basic statistical analysis

Next, single models were ran to find associations between each variable with covid test postivity. Complex survey design was used to allow generalizability of results to the United States population.

|  |
| --- |
| Table 3: Assocation of mental health factors with positive COVID-19 test results among the United State population. |

## 4.3 Full analysis

To better understand the relationship between COVID-19 test status and mental health factors, three new models were constructed: a multivariate logistic regression model, LASSO regression, and a random forest model. Each model used the imputed data and five fold cross validation to affect predictive performance. All models accounted for sample wighting, but not clustering. Notably, survey design could not be fully accounted for due to the complexity of the LASSO and random forest models, making results not interpretable to the country-wide level, rather, predictive results are relevant within the NHIS population, which accounts for 110,216 individuals.

All models included COVID-19 test status as the outcome and all mental helath variables as predictors (hours of sleep, touble sleeping, feeling depressed, meeting exercise guidelines, and amount of social support). The multivariate logistic regression was chosen for it’s ease of interpretability and to serve as a baseline against the more advanced models. The LASSO model was chosen for it’s ability to penalize predictors to indicate which mental health factors were most important. The random forest model was used to capture potential complex relationships between variables.

To evalaute the performance of these models, the area under the curve (AUC) was calculated. This metric was chosen due to it’s discrimination between binary outcome of the COVID-19 variable. Higher AUC indicates better model preformance. Despite the Logistic Regression model and the Lasso Regression model have the same mean AUC, they are not the same, shown by the differing values for the imputations. Code was checked to ensure the models were not the same.

|  |
| --- |
| Figure 2: Plot comparing area under the curve (AUC) metrics among imputations between logistic regression model, LASSO regression model, and random forest model. Higher AUC indicates better preformance. |

Observed vs. expected plots for the three models compare the predictive power of the models from the cross validated data. Training data was used for expectations and observed data was drawn from the testing data. Plots evaluate the models’ modest predictive power on COVID-19 test outcomes from mental health variables.

|  |
| --- |
| Figure 3: Logistic Regression calibration plot. |

|  |
| --- |
| Figure 4: LASSO Regression calibration plot. |

|  |
| --- |
| Figure 5: Random Forest calibration plot. |

# 5. Discussion

## 5.1 Summary and Interpretation

The exploratory analysis revelaed that feeling depressed occurs a prevalence of 0.095 (0.09, 0.10). meeting exercise requirements occurs at a prevalence of 0.526 (0.52, 0.53), getting enough sleep occurs at a prevalence of 0.494 (0.49, 0.50), recieving less social support occurs at a prevalence of 0.097 (0.09, 0.10), and having trouble sleeping occurs at a prevalence of 0.630 (0.62, 0.64) (Tab. 2). Several of these factors occur in the majority of the population, that is, trouble sleeping and exercise, and getting enough sleep occured in almost half the population. Feeling depressed and recieving less socual support occured in a small fraction of the population (Tab. 2).

The exploratory analysis revealed prevalence of mental health factors among the population by COVID-19 test status. Interestingly, individuals who reported receiving less social support occured at a higher rate for those who did not test positive for COVID-19 (0.10 (0.10, 0.10)) than those who did test positive for COVID-19 (0.08 (0.07, 0.09)) (Tab. 2). A difference in prevalence between COVID-19 test status was not noted for individuals who reported no change/more social support.

Single models quantified association of each mental health factor with a positive COVID-19 test. Feeling depressed, and getting exercise were not significantly associated witha positive COVID-19 test (Tab. 3). Getting enough sleep was found to slightly increase the odds of a positive COVID-19 test (OR = 1.076 (1.026–1.13), p = 0.003). Getting less social support was found to slightly decrease the odds of a positive COVID-19 test (OR = 0.777, (0.691, 0.875), p < 0.0001). This may indicate effectiveness of societal control strategies, that is, social distancing, quarantine, and isolation in preventing spread of COVID-19. Reporting trouble sleeping was found to slightly increase the odds of a positive COVID-19 test (OR = 1.134 (1.067, 1.206), p < 0.0001), which slightly disagrees with results of the SLEEP variable described earlier (Tab. 3).

Individuals who reported recent trouble sleeping were occured at a higher prevalence for those who tested postive for COVID-19 (0.374 (0.36, 0.39)) than those who did not recieve a positive COVID-19 test (0.345 (0.34, 0.35)). Individuals who did not report recent trouble sleeping showed a difference by COVID-19 test status, with less testing positive for COVID-19 (0.626 (0.61, 0.64)) and more not testing positive for COVID-19 (0.655 (0.65, 0.66)) (Tab. 2).

Model preformance was evaluated via the AUC metric. The logistic regression model had a pooled AUC of 0.5031. The LASSO regression model showed a pooled AUC of 0.5301. The random forest model showed the highest pooled AUC of 0.5474. AUC values for all models are fairly close to 0.50, with only small improvement with the randome forest model (Fig. 2). This indicates the models have little ability to discriminate between a postive or negative COVID-19 test based on mental health factors, with the random forest model being the best.

Plots for observed vs. expected values provide a secondary assessment of model preformance, which show how closesly the training data matched the testing data. Plots for well calibrated models should show the points for imputations falling along the 45 degree dotted line, which is not present in the logistic regression (Fig. 3) or LASSO regression models (Fig. 4). The random forest model shows some closeness to the 45 degree line, confirming it to be the best model (Fig. 5). However, no model shows good predicitive power, given there is minimal closeness to the line.

## 5.2 Strengths and Limitations

This study used data collected from the National Health Information Surveys, a well established and thorough survey program run by the CDC. NHIS uses sampling methods and covers many health-related topics, allowing it to be generalizable to the United States population. Additionally, this study uses multiple imputation to fill missing data points, which reduces the introduciton of potential bias from discarding missings or from single imputation. Applying cross validation to the three full models allows for analysis of predictive power, and the use of the LASSO regression allows for penalization of unecessary predictors, and the use of the random forest models allows for potential non-linear relationships.

Limitations of this study include the nature of survey data. Because individuals self-reported this data, it is possible that recall bias may influence accuracy. Associations between predictors and COVID test status cannot be established as causal. Without inclusions of full survey design variables, the three full models are not generalizable to the United States population. Finally, all three models showed moderate to low predicitive power.

## 5.3 Conclusions

We found that indications of poor mental health were common in the population United States population, with trouble sleeping, exercise, and getting enough sleep, while depressed and social isolation occured in at a lower prevalence (Tab. 2). Sleep difficulties were commonly reported during the SARS-CoV-2 pandemic, which supports these results.

Social support was the only mental health variable to vary among the population by COVID-19 test status (Tab. 2). Interestingly, recieving less social support was found to be protective against SARS-CoV-2 (Tab. 3). This is likely due to isolation and quarantining practices that occured during thee years 2020-2022.

Predictive models showed weak preformancce and poor predictive power (Figs. 2, 3). The relationship between SARS-CoV-2 and mental health continues to be challenging to understand and interpret.

# 6. References

1. Kumar A, Nayar KR. COVID-19 and its mental health consequences. J Ment Health. 2020;30(1):1–2. https://doi.org/10.1080/09638237.2020.1757052
2. Blewett LA, Rivera Drew JA, King ML, Williams KCW, Backman D, Chen A, et al. IPUMS Health Surveys: National Health Interview Survey, Version 7.4 [dataset]. Minneapolis, MN: IPUMS; 2024. https://doi.org/10.18128/D070.V7.4
3. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services; 2018.
4. U.S. Department of Health and Human Services. 2005 Guide to Healthy Sleep. Maryland: U.S. Department of Health and Human Services; 2011.
5. Gupta R, Pandi-Perumal SR. COVID-Somnia: How the Pandemic Affects Sleep/Wake Regulation and How to Deal with it?. Sleep Vigil. 2020;4:51–53. https://doi.org/10.1007/s41782-020-00118-0
6. Szkody E, Stearns M, Stanhope L, McKinney C. Stress-buffering role of social support during COVID-19. Fam Process. 2021;60(3):1002–1015. https://doi.org/10.1111/famp.12618
7. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr. 2020;52:102066. https://doi.org/10.1016/j.ajp.2020.102066