Mental Health Factors and Covid

Annalise Cramer

This uses MS Word as output format. [See here](https://quarto.org/docs/output-formats/ms-word.html) for more information. You can switch to other formats, like html or pdf. See [the Quarto documentation](https://quarto.org/) for other formats.

[1] "/Users/annalisecramer/MADA/CRAMER-MADA-Project"

**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

*Write a summary of your project.*

# 2. Introduction

## 2.1 General Background Information

During the SARS-CoV pandemic, methods implemented to control the transmission of the disease had broader societal impacts. A common complaint surrounding social distancing and quarantine policies is the effects on mental health 1. This study aims to better understand the associations between mental health factors and contracting covid. Using data from the National Health Interview Surveys collected during the SARS-CoV-2 pandemic, simple and multiple regression models are created.

Cite this properly later: 1.) https://www.tandfonline.com/doi/full/10.1080/09638237.2020.1757052

## 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 2019 - 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 test?

Factors studied will include hours of sleep, reported depressed mood, reported trouble sleeping, recent exercise, and reported social/emotional support.

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above and have the right bibtex key. Then you can include like this:

# 3. Methods

*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

There’s 151406 observations and 47 variables.

Survey design variables were used and new binary variables were created from other variables. Using complex survey design variables, a descriptive table was created.

I will create models (basic and adjusted) to study association (generating prevelance ratios with CIs) between mental health factors (hours of sleep, feeling depressed mood, amount of social and emotional support, trouble sleeping, recent exercise benchmarks) with a postiive covid test result. I will generate tables and plots for all results.

## 3.1 Schematic of workflow

Sometimes you might want to show a schematic diagram/figure that was not created with code (if you can do it with code, do it). fig-schematic is an example of some - completely random/unrelated - schematic that was generated with Biorender. We store those figures in the assets folder.

## 3.2 Data aquisition

This data was obtained from IPUMS, used with permission.

## 3.3 Data import and cleaning

The following code details the data cleaning process. Labels are removed from the data to make it easier to work with to create models later in the process. Variables needed are selected, then a series of binary varibales are created. Additional details are available in the processing folder.

here() #set working directory

[1] "/Users/annalisecramer/MADA/CRAMER-MADA-Project"

data\_original <- readRDS(here("data","raw-data", "NHIS\_COVID.rds"))  
data1 <- labelled::remove\_val\_labels(data\_original) #removes labels from haven package which cause problems  
  
data1 <- data\_original %>%  
 select(STRATA, PSU, PERNUM, SAMPWEIGHT, AGE, PA18AERSTR, HRSLEEP, DEPFREQ, PHQSLEEP, CMPSUPPORT, CVDTESTRSLT)  
  
data2 <- data1 %>%  
 mutate(  
 PA18AERSTR = ifelse(PA18AERSTR %in% c(0, 7, 8, 9), NA, PA18AERSTR), #put NA, 0, 6, 7, 8, 9 to NA  
 EXERCISE = if\_else(PA18AERSTR %in% c(2,3,4), 1, 0, missing = NA\_real\_) #create EXERCISE variable, 1 if guidelines by Am Heart Assoc met for aerobics, strength, or both, 0 if not met  
 ) %>%  
 mutate(  
 AGE = ifelse(AGE %in% c(997, 999), NA, AGE), #invalid ages to NA  
 SLEEP = case\_when( #Based on age, amount of sleep required from https://www.nhlbi.nih.gov/health/sleep/how-much-sleep#:~:text=Experts%20recommend%20that%20adults%20sleep,or%20more%20hours%20a%20night.  
 AGE < 1 ~ if\_else(HRSLEEP >= 12 & HRSLEEP <= 16, 1, 0, missing = NA\_real\_),  
 AGE >= 1 & AGE < 3 ~ if\_else(HRSLEEP >= 11 & HRSLEEP <= 14, 1, 0, missing = NA\_real\_),  
 AGE >= 3 & AGE < 6 ~ if\_else(HRSLEEP >= 10 & HRSLEEP <= 13, 1, 0, missing = NA\_real\_),  
 AGE >= 6 & AGE < 13 ~ if\_else(HRSLEEP >= 9 & HRSLEEP <= 12, 1, 0, missing = NA\_real\_),  
 AGE >= 13 & AGE <= 18 ~ if\_else(HRSLEEP >= 8 & HRSLEEP <= 10, 1, 0, missing = NA\_real\_),  
 TRUE ~ NA\_real\_ #default NA for unspecified ages  
 )) %>%  
 mutate(  
 DEPRESSED = case\_when(  
 DEPFREQ %in% c(1, 2) ~ 1, #daily and weekly is considered depressed  
 DEPFREQ %in% c(3, 4, 5) ~ 0, #monthly and a few times a year and never are considered not depressed  
 DEPFREQ %in% c(0, 7, 8, 9) ~ NA\_real\_,#NIU & Unknowns are NA  
 TRUE ~ NA\_real\_ #default NA for unexpected values  
 )) %>%  
 mutate(  
 TROUBLE\_SLEEPING = case\_when(  
 PHQSLEEP %in% c(1, 2, 3) ~ 1, #several days, more than half days, nearly every day are trouble sleeping  
 PHQSLEEP == 0 ~ 0, #not at all is no trouble sleeping  
 PHQSLEEP %in% c(6, 7, 8, 9) ~ NA\_real\_, #NIU & Unknowns are NA  
 TRUE ~ NA\_real\_ #default NA for unexpected values  
 )) %>%  
 mutate(  
 SOCIAL = case\_when(  
 CMPSUPPORT == 1 ~ 0, #more social and emotional support  
 CMPSUPPORT == 2 ~ 1, #less social and emotional support  
 CMPSUPPORT == 3 ~ 0, #about the same  
 CMPSUPPORT %in% c(0, 7, 8, 9) ~ NA\_real\_, # NIU & Unknowns → NA  
 TRUE ~ NA\_real\_ #default NA for unexpected values  
 )) %>%  
 mutate(  
 COVID = case\_when(  
 CVDTESTRSLT == 2 ~ 1, #yes, had COVID  
 CVDTESTRSLT == 1 ~ 0, #no, did not have COVID  
 CVDTESTRSLT %in% c(0, 3, 7, 8, 9) ~ NA\_real\_, #NIU, did not receive results, unknowns, and NA are NA  
 TRUE ~ NA\_real\_ #default NA for unexpected values  
 ))

## 3.4 Statistical analysis

*Explain anything related to your statistical analyses.*

We will begin with exploratory tables to explore proportions of NA answers, and generate early estimates.

Our first figure with be a discriptive table stratified by COVID status, with estimates and 95% CIs.

Next, we will construct simple models before moving onto adjusted models.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

*Use a combination of text/tables/figures to explore and describe your data. Show the most important descriptive results here. Additional ones should go in the supplement. Even more can be in the R and Quarto files that are part of your project.*

The first table below details the proportion of NAs for each mental health variable. We can see this is a little bit high, which may affect analysis later on. Histograms of each variable are available in the eda folder.

The second table shows proportions calculated using complex survey design variables.

data <- data2  
#summarize EXERCISE  
summary\_exercise <- data %>%  
 summarise(  
 Variable = "EXERCISE",  
 Prop\_1 = mean(EXERCISE == 1, na.rm = TRUE),  
 Prop\_0 = mean(EXERCISE == 0, na.rm = TRUE),  
 Prop\_NA = mean(is.na(EXERCISE))  
 )  
  
#summarize SLEEP  
summary\_sleep <- data %>%  
 summarise(  
 Variable = "SLEEP",  
 Prop\_1 = mean(SLEEP == 1, na.rm = TRUE),  
 Prop\_0 = mean(SLEEP == 0, na.rm = TRUE),  
 Prop\_NA = mean(is.na(SLEEP))  
 )  
  
#summarize DEPRESSED  
summary\_depressed <- data %>%  
 summarise(  
 Variable = "DEPRESSED",  
 Prop\_1 = mean(DEPRESSED == 1, na.rm = TRUE),  
 Prop\_0 = mean(DEPRESSED == 0, na.rm = TRUE),  
 Prop\_NA = mean(is.na(DEPRESSED))  
 )  
  
#summarize TROUBLE SLEEPING  
summary\_trouble\_sleeping <- data %>%  
 summarise(  
 Variable = "TROUBLE SLEEPING",  
 Prop\_1 = mean(TROUBLE\_SLEEPING == 1, na.rm = TRUE),  
 Prop\_0 = mean(TROUBLE\_SLEEPING == 0, na.rm = TRUE),  
 Prop\_NA = mean(is.na(TROUBLE\_SLEEPING))  
 )  
  
#summarize SOCIAL  
summary\_social <- data %>%  
 summarise(  
 Variable = "SOCIAL",  
 Prop\_1 = mean(SOCIAL == 1, na.rm = TRUE),  
 Prop\_0 = mean(SOCIAL == 0, na.rm = TRUE),  
 Prop\_NA = mean(is.na(SOCIAL))  
 )  
  
#summarize COVID  
summary\_covid <- data %>%  
 summarise(  
 Variable = "COVID",  
 Prop\_1 = mean(COVID == 1, na.rm = TRUE),  
 Prop\_0 = mean(COVID == 0, na.rm = TRUE),  
 Prop\_NA = mean(is.na(COVID))  
 )  
  
#combine all summaries and convert to gt table  
summary\_table <- bind\_rows(  
 summary\_exercise,   
 summary\_sleep,   
 summary\_depressed,   
 summary\_trouble\_sleeping,   
 summary\_social,   
 summary\_covid  
) %>%  
 gt() %>%  
 tab\_header(  
 title = "Proportions of 1, 0, and NA for Various Variables"  
 )  
  
#print the combined table  
summary\_table

Table 1: Proportions of 1, 0, and NA for Various Variables

| Variable | Prop\_1 | Prop\_0 | Prop\_NA |
| --- | --- | --- | --- |
| EXERCISE | 0.20186783 | 0.7981322 | 0.00000000 |
| SLEEP | 0.01411453 | 0.9858855 | 0.91015548 |
| DEPRESSED | 0.09519991 | 0.9048001 | 0.06929712 |
| TROUBLE SLEEPING | 0.34300841 | 0.6569916 | 0.61452650 |
| SOCIAL | 0.11486353 | 0.8851365 | 0.69921271 |
| COVID | 0.26836904 | 0.7316310 | 0.59495661 |

#list variables  
varstable1 = c("EXERCISE", "SLEEP", "DEPRESSED", "TROUBLE\_SLEEPING",   
 "SOCIAL", "COVID")  
  
#list categories  
cat\_vartab1 = c("EXERCISE", "SLEEP", "DEPRESSED", "TROUBLE\_SLEEPING",   
 "SOCIAL", "COVID")  
  
#use function to make table  
Tab1 = CreateTableOne(vars = varstable1, factorVars = cat\_vartab1, data = data)  
  
# Convert the TableOne object to a data frame  
tab1\_df <- as.data.frame(print(Tab1, showAllLevels = TRUE, quote = FALSE, noSpaces = TRUE))

level Overall   
 n 151406   
 EXERCISE (%) 0 120842 (79.8)  
 1 30564 (20.2)   
 SLEEP (%) 0 13411 (98.6)   
 1 192 (1.4)   
 DEPRESSED (%) 0 127499 (90.5)  
 1 13415 (9.5)   
 TROUBLE\_SLEEPING (%) 0 38344 (65.7)   
 1 20019 (34.3)   
 SOCIAL (%) 0 40310 (88.5)   
 1 5231 (11.5)   
 COVID (%) 0 44868 (73.2)   
 1 16458 (26.8)

# Clean up row names  
tab1\_df <- tibble::rownames\_to\_column(tab1\_df, "Variable")  
  
# More advanced formatting  
tab1\_gt <- gt(tab1\_df) %>%  
 # Add a title  
 tab\_header(  
 title = md("\*\*COVID test status and Mental Health Factors\*\*"),  
 subtitle = "Survey Results Summary"  
 ) %>%  
 # Format the columns  
 fmt\_percent(  
 columns = contains("%"),  
 decimals = 1  
 ) %>%  
 # Group the variables  
 tab\_row\_group(  
 label = "COVID Status",  
 rows = contains("COVID")  
 ) %>%  
 tab\_row\_group(  
 label = "Mental Health Indicators",  
 rows = contains(c("DEPRESSED", "TROUBLE\_SLEEPING"))  
 ) %>%  
 tab\_row\_group(  
 label = "Lifestyle Factors",  
 rows = contains(c("EXERCISE", "SLEEP", "SOCIAL"))  
 ) %>%  
 # Style the table  
 tab\_style(  
 style = cell\_fill(color = "#e6f3ff"),  
 locations = cells\_row\_groups()  
 ) %>%  
 tab\_style(  
 style = list(  
 cell\_text(weight = "bold")  
 ),  
 locations = cells\_row\_groups()  
 ) %>%  
 # Add footnotes if needed  
 tab\_footnote(  
 footnote = "Percentages may not sum to 100% due to rounding",  
 locations = cells\_column\_labels(columns = contains("%"))  
 ) %>%  
 # Adjust column widths  
 cols\_width(  
 Variable ~ px(250),  
 everything() ~ px(120)  
 )  
  
tab1\_gt

Table 1: **COVID test status and Mental Health Factors**

Survey Results Summary

| Variable | level | Overall |
| --- | --- | --- |
| n |  | 151406 |
| EXERCISE.... | 0 | 120842 (79.8) |
| X | 1 | 30564 (20.2) |
| SLEEP.... | 0 | 13411 (98.6) |
| X.1 | 1 | 192 (1.4) |
| DEPRESSED.... | 0 | 127499 (90.5) |
| X.2 | 1 | 13415 (9.5) |
| TROUBLE\_SLEEPING.... | 0 | 38344 (65.7) |
| X.3 | 1 | 20019 (34.3) |
| SOCIAL.... | 0 | 40310 (88.5) |
| X.4 | 1 | 5231 (11.5) |
| COVID.... | 0 | 44868 (73.2) |
| X.5 | 1 | 16458 (26.8) |

[Table 1](#tbl-summarytable) shows a summary of the data.

Note the loading of the data providing a **relative** path using the ../../ notation. (Two dots means a folder up). You never want to specify an **absolute** path like C:\ahandel\myproject\results\ because if you share this with someone, it won’t work for them since they don’t have that path. You can also use the here R package to create paths. See examples of that below. I generally recommend the here package.

[1] "/Users/annalisecramer/MADA/CRAMER-MADA-Project"

|  |
| --- |
| Table 1: Table. 1. Mental health characteristics among Americans from 2019-2022, by SARS-CoV-2 test result status, NHIS |

## 4.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

fig-result shows a scatterplot figure produced by one of the R scripts.

## 4.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example tbl-resulttable2 shows a summary of a linear model fit.

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper [@leek2015] discusses types of analyses.

These papers [@mckay2020; @mckay2020a] are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like.

# 6. References