Albany GA Risk of COVID-19 Result As of May 17th 2023

# Cleaning file for the statistical analysis investigating the association between Covid-19 exposure and symptom severity with vaccination status or intent

## Run Date: May 17th 2023

## Qualtrics download date: May 17th 2023 with 1089 observations

## Pen & Paper last entry date, download date and observations: 4/08/203, 05/17/2023, 459 observations

## Library

here() starts at /Users/jacobmatta/Documents/R/ASU-Project-

── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
✔ ggplot2 3.4.0 ✔ purrr 0.3.5  
✔ tibble 3.1.8 ✔ dplyr 1.1.0  
✔ tidyr 1.2.1 ✔ stringr 1.5.0  
✔ readr 2.1.3 ✔ forcats 0.5.2  
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()

## Loading the clean data

## Cleaning the exposed questions: changing all values that follow question 1(Q2.1) if this question was completed as either: no, not sure or do not know then mark Question 2 - Question 14 as missing

## Subsetting the survey in order to just contain symptom questions

## Checking the varibale types

tibble [1,548 × 16] (S3: tbl\_df/tbl/data.frame)  
 $ Question\_1 : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Fever : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Cough : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Shortness\_of\_breath: chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Fatigue : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Muscle\_Body\_Aches : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Headache : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Loss\_of\_Taste : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Sore\_Throat : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Congestion : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Nausea : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Diarrhea : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Asymptomatic : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Other : chr [1:1548] NA NA NA NA ...  
 $ Question\_15 : chr [1:1548] NA NA NA NA ...  
 $ Question\_27 : chr [1:1548] "1" "2" "2" NA ...

## All Variables are saved as character variables so lets convert them to numeric variables

Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion

## Removing all participants who answered No, not sure, do not know or left question 1 blank

## Now we have a data set that is 496 observations which is all of the participants who had Covid-19 or had a family member who had Covid-19. These are exposed individuals

## Removing those in the exposed group who have answered missing to question 15 since less than 5% of the exposure group answered missing

## After removing missing there are 467 observations

## Calculating the scale mean for Question 15

[1] 1443

[1] 3.089936

[1] 1.341179

## Taking a look at the symptom severity of those exposed to covid-19 after participants who answered missing to question 15 were removed

0 1   
197 270

0 1   
173 294

0 1   
303 164

0 1   
237 230

0 1   
245 222

0 1   
245 222

0 1   
304 163

0 1   
277 190

0 1   
249 218

0 1   
370 97

0 1   
369 98

0 1   
425 42

0 1 5   
452 14 1

## Creating a new dataframe with just asymptomatic individuals

## Asymptomatic dataframe has 42 observations

## Calculating the scale mean for Question 15 for those assymptomatic individuals

[1] 121

[1] 2.880952

[1] 1.347127

## Creating a new data set from the exposed individuals who are symptomatic

## 449 indivdiuals were symptomatic out of the exposed grouo

## Determining the risk score of symptomatic individuals

[1] 1404

[1] 3.126949

[1] 1.331088

## Unexposed group: Subsetting the survey in order to just contain symptom questions

## Checking the varibale types

tibble [1,548 × 16] (S3: tbl\_df/tbl/data.frame)  
 $ Question\_1 : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Fever : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Cough : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Shortness\_of\_breath: chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Fatigue : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Muscle\_Body\_Aches : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Headache : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Loss\_of\_Taste : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Sore\_Throat : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Congestion : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Nausea : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Diarrhea : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Asymptomatic : chr [1:1548] NA NA NA NA ...  
 $ Question\_13\_Other : chr [1:1548] NA NA NA NA ...  
 $ Question\_15 : chr [1:1548] NA NA NA NA ...  
 $ Question\_27 : chr [1:1548] "1" "2" "2" NA ...

## All Variables are saved as character variables so lets convert them to numeric variables

Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion

## Removing all participants who answered Yes, not sure, do not know or left question 1 blank

Unexposed group has 863 observations

0 1 2 3 4 5   
 25 193 203 163 171 108

0 1 2 3 4 5   
0.02896871 0.22363847 0.23522596 0.18887601 0.19814600 0.12514484

## Removing those in the unexposed group who have answered missing to question 15 since less than 5% of the unexposure group answered missing

Now have 838 observations

## Calculating the scale mean for Question 15

[1] 2312

[1] 2.75895

[1] 1.351714

# Creating a data set with just question 1 and question 15 to test the association betweeen exposed and unexposed and percieved risk

## Loading the clean data

## Subsetting the survey to contain question 1 and question 15

## Changing data to numeric data

Warning: NAs introduced by coercion  
  
Warning: NAs introduced by coercion

0 1 2 3 4   
125 496 863 43 21

0 1 2 3 4 5   
164 290 285 308 307 194

0 1 2   
125 496 863

0 1 2 3 4 5   
162 279 277 282 294 190

## running the chi square test with missing

Pearson's Chi-squared test  
  
data: table(exposed\_unexposed$Question\_1, exposed\_unexposed$Question\_15)  
X-squared = 832.22, df = 10, p-value < 2.2e-16

indicates there is a strong association between the two questions

## running the chi square test without missing

1 2   
467 838

1 2 3 4 5   
277 272 277 292 187

Pearson's Chi-squared test  
  
data: table(exposed\_unexposed$Question\_1, exposed\_unexposed$Question\_15)  
X-squared = 27.374, df = 4, p-value = 1.67e-05

## Creating a new variable for question 15

## making a contigency table between question 1 and question 15

at risk neutral not at risk  
 1 200 114 153  
 2 279 163 396

## Calculating the ORs

[1] 1.500812

[1] 0.1804226

[1] 1.053774

[1] 2.137495

Warning in matrix(c(200, 114, 110, 153, 279, 163, 396), nrow = 2, byrow = TRUE):  
data length [7] is not a sub-multiple or multiple of the number of rows [2]

Fisher's Exact Test for Count Data  
  
data: contingency\_table  
p-value = 2.776e-14  
alternative hypothesis: two.sided

[1] 1.337354

[1] 0.939004

[1] 1.904694

[1] 0.5438622

[1] 0.3818651

[1] 0.7745826