Data Analysis Project

Chris Okitondo and Irene Cavros

2/3/23

# 1. Summary/Abstract

*Write a summary of your project.*

PIs: Chris Okitondo and Irene Cavros

Abstract to be completed in future

# 2. Introduction

## 2.1 General Background Information

*Provide enough background on your topic that others can understand the why and how of your analysis*

To be completed in future

## 2.2 Description of data and data source

*Describe what the data is, what it contains, where it is from, etc. Eventually this might be part of a methods section.*

The data source used for this project is the National Health Interview Survey (NHIS.) It is a harmonized set of data beginning in 1963 with information on general health status, acute and chronic illness, functional limitations, access to care, insurance coverage, and health behaviors for the U.S. population. On average, the survey covers 100,000 persons in 45,000 households each year. PUMS NHIS currently contains over 17,000 integrated variables from 1963 to present.

As the outcome of interest, researchers will focus on those who, in this household survey between the years of 2012 and 2021, answered no when they were asked whether they had ANY flu vaccine in the past 12 months.

Possible variables to be assessed are as follows: 1. Demographics: age, sex, race, ethnicity, veteran status 2. Socio-economic status: education level, total combined family income, employment status, paid sick leave as part of job benefits, ratio of family income to poverty threshold 3. General health: perceived health status, categorical BMI Conditions: ever told had asthma, ever told had cancer, ever told had coronary heart disease 4. COVID-19: had COVID-19 according to test 5. Health behaviors: alcohol drinking status, cigarette smoking status Access to care: has usual place for medical care, medical care delayed due to cost (past 12 months) 6. Health insurance: health insurance coverage status 7. Mental health: anxiety level, depression level

## 2.3 Questions/Hypotheses to be addressed

*State the research questions you plan to answer with this analysis.*

Using the aforementioned database, our research question of interest is as follows: which groups are at highest risk for low uptake of flu vaccine?

# 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.*

To be completed in future

## 3.1 Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

To be completed in future

## 3.2 Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what kind of cleaning/processing you do, and provide more details and well documented code somewhere (e.g. as supplement in a paper). All materials, including files that contain code, should be commented well so everyone can follow along.*

To be completed in future

## 3.3 Statistical analysis

*Explain anything related to your statistical analyses.*

To be completed in future

# 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.*

To be completed in future

## 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.* To be completed in future

## 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.*

To be completed in future

# 5. Discussion

## 5.1 Summary and Interpretation

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

To be completed in future

## 5.2 Strengths and Limitations

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

To be completed in future

## 5.3 Conclusions

*What are the main take-home messages?*

To be completed in future

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

To be completed in future

# 6. References

To be completed in future