

STA 440 Final Project

Shirley Mathur

4/14/2022

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.0.0      v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.0.5
## Warning: package 'dplyr' was built under R version 4.0.5

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(lme4)

## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

Introduction

Background

Even in modern American society, lesbian, gay, and bisexual (LGB) individuals are still targets of discrimination at a systemic level, as evidenced by the legislation that seeks to target these individuals at all levels of government, ranging from the federal to municipal (@hrc). As posited by Meyer, LGB individuals that have to face such discrimination and prejudice in society are subject to an increased risk of developing mental health problems due to the *minority stress* that they encounter in their daily lives (@minority_stress). However, while this original meta-analysis specifically explored the disparities between the mental health outcomes of LGB individuals against heterosexual individuals, it did not necessarily account for compounding factors that could also influence stress, such as a LGB individuals who are also part of racial minority groups or comparing LGB individuals who are women as opposed to LGB individuals who are men. More recent research exploring these same phenomenon has been done in recent years to see how over time the mental health outcomes of LGB individuals compare to those of heterosexual individuals, while also trying to adjust for these other factors that account for how LGB individuals might also be part of other minority groups. One such study, utilizing data from the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC) that was collected in 2005, found that LGB individuals were associated with a higher odds of

having mental health disorders only if they were subject to other forms of discrimination as due to their other marginalized identities outside of just being a LGB individual (@NESARC).

In this report, I seek to further expand upon the line of exploration of how different marginalized identities could together have an association with higher odds of having mental health disorders. I also seek to control for other factors that might influence differential treatment for individuals based on their societal contexts, such as accounting for the region in which people live.

Dataset Description and Variables of Interest

```
#setwd("/Users/shirley/Downloads/ICPSR_37166/DS0001")
load("ICPSR_37166/DS0001/37166-0001-Data.rda")
load("ICPSR_37166/DS0003/37166-0003-Data.rda")
load("ICPSR_37166/DS0005/37166-0005-Data.rda")
load("ICPSR_37166/DS0007//37166-0007-Data.rda")

# use weights
# longitudinal data
# take difference and regress on that

# CHARACTERISTICS VARS
# w1sexminid, w2sexminid - broad categorization of sexuality (gay/lesbian, bi/pan, other (ace, anti-lab
# educ1 or educ2 (GEDUC1 has 4 options for education, GEDUC2 has 2 options for education)
# gurban, gurban_i (urbanicity, either imputed)
# gcenreg (census division into NE, Midwest, South, West)
# cohort variable looks at age (young, middle, and old)
# utilize survey 2 weights (w2) for longitudinal analysis
# w1_race for racial categories
# w1sex_gender for gender and sex combined variable
# participant did not re
# gmilesaway2 (whether or not there is LGBT health center within 60 miles of respondent)
# w1poverty (whether or not individual is living in poverty (using 2017 thresholds))
# w1conversion (whether or not individual underwent therapy for sexual orientation change)
# w1conversionhc (whether or not individual's sex converesion was from healthcare provider)
# w1conversionrel (whether or not individual's sex converesion was from religious leader)

##### OUTCOMES
# w1socialwb, w1socialwb_i, w2socialwb, w2socialwb_i (social well being, use imputed values or not)
# w1kessler6, w1kessler6_i, w2kessler6, w2kessler6_i (mental disability, use imputed values or not)

# can just use the combined "socialwb" type scales for response variables

# set df equal to full data across both waves
df <- da37166.0007

# filter for only those participants who participated in both waves
full_df <- df %>%
  filter(WAVEPARTICIPATED == "(2) Waves 1 and 2")

# look at counts across race
unique(full_df$W1RACE)
```

[1] (6) White

(3) Hispanic/Latino

```
## [3] (1) Asian (8) Multirace
## [5] (4) Middle Eastern (7) American Indian
## [7] (2) Black/African American (5) Native Hawaiian/Pacific Islander
## 8 Levels: (1) Asian (2) Black/African American ... (8) Multirace
```

```
full_df %>%
  group_by(W1RACE) %>%
  summarise(count = n())
```

```
## # A tibble: 8 x 2
##   W1RACE count
##   <fct>   <int>
## 1 (1) Asian      11
## 2 (2) Black/African American    67
## 3 (3) Hispanic/Latino          71
## 4 (4) Middle Eastern           2
## 5 (5) Native Hawaiian/Pacific Islander 1
## 6 (6) White                639
## 7 (7) American Indian          3
## 8 (8) Multirace             100
```

```
# look at counts across census division
```

```
full_df %>%
  group_by(GCENDIV) %>%
  summarise(count = n())
```

```
## # A tibble: 9 x 2
##   GCENDIV count
##   <fct>   <int>
## 1 (1) New England      59
## 2 (2) Middle Atlantic 130
## 3 (3) East North Central 107
## 4 (4) West North Central  67
## 5 (5) South Atlantic   163
## 6 (6) East South Central  32
## 7 (7) West South Central  74
## 8 (8) Mountain        90
## 9 (9) Pacific        172
```

```
# look at counts for gender
```

```
full_df %>%
  group_by(W1SEX_GENDER) %>%
  summarise(count = n())
```

```
## # A tibble: 4 x 2
##   W1SEX_GENDER count
##   <fct>   <int>
## 1 (1) Women, non-transgender    427
## 2 (2) Men, non-transgender     409
## 3 (3) genderqueer non-binary, female  38
## 4 (4) genderqueer non-binary, male    20
```

```
# look at counts for outcome of interest: mental disability at w1
```

```
full_df %>%
  group_by(W1KESSLER6) %>%
  summarise(count = n())
```

```
## # A tibble: 25 x 2
```

```
##      W1KESSLER6 count
##      <dbl> <int>
## 1          0    56
## 2          1    59
## 3          2    72
## 4          3    70
## 5          4    54
## 6          5    75
## 7          6    65
## 8          7    44
## 9          8    55
## 10         9    42
## # ... with 15 more rows
```

```
# look at counts for outcome of interest: mental disability at w2
full_df %>%
  group_by(W2KESSLER6) %>%
  summarise(count = n())
```

```
## # A tibble: 26 x 2
##      W2KESSLER6 count
##      <dbl> <int>
## 1          0    51
## 2          1    48
## 3          2    74
## 4          3    68
## 5          4    70
## 6          5    74
## 7          6    58
## 8          7    54
## 9          8    54
## 10         9    40
## # ... with 16 more rows
```

```
# look at counts for potential outcome of interest: social well being at w1
full_df %>%
  group_by(W1SOCIALWB) %>%
  summarise(count = n())
```

```
## # A tibble: 73 x 2
##      W1SOCIALWB count
##      <dbl> <int>
## 1      1.73     2
## 2      1.80     1
## 3      2.13     1
## 4      2.20     1
## 5      2.27     1
## 6      2.40     3
## 7      2.47     2
## 8      2.53     3
## 9      2.60     1
## 10     2.67     4
## # ... with 63 more rows
```

```
# look at counts for potential outcome of interest: social well being at w2
full_df %>%
```

```

group_by(W2SOCIALWB) %>%
  summarise(count = n())

## # A tibble: 72 x 2
##   W2SOCIALWB count
##   <dbl> <int>
## 1      1.47     1
## 2      1.87     1
## 3      1.93     1
## 4      2      1
## 5      2.13     1
## 6      2.20     2
## 7      2.27     1
## 8      2.40     1
## 9      2.47     4
## 10     2.53     4
## # ... with 62 more rows

# look at counts for broad sexuality grouping
full_df %>%
  group_by(W2SEXMINID) %>%
  summarise(count = n())

## # A tibble: 4 x 2
##   W2SEXMINID          count
##   <fct>          <int>
## 1 (1) Lesbian/gay          505
## 2 (2) Bisexual            272
## 3 (3) Other sexual minority identity 109
## 4 <NA>                     8

# look at counts for specific sexuality grouping
full_df %>%
  group_by(W2SEXUALID) %>%
  summarise(count = n())

## # A tibble: 9 x 2
##   W2SEXUALID          count
##   <fct>          <int>
## 1 (01) Straight/heterosexual      8
## 2 (02) Lesbian            175
## 3 (03) Gay                330
## 4 (04) Bisexual            272
## 5 (05) Queer               56
## 6 (06) Same-gender loving      15
## 7 (07) Other                6
## 8 (08) Asexual spectrum       13
## 9 (09) Pansexual            19

# look at how many individuals are missing either w1 or w2 kessler data
full_df %>%
  filter(is.na(W2KESSLER6) | is.na(W1KESSLER6)) %>%
  count()

##   n
## 1 29

```

```
# 29 observations missing value from both times for kessler rating

# filter to only get observations that have values at both time periods, and select only variables and
complete_df <- full_df %>%
  filter(!(is.na(W2KESSLER6) | is.na(W1KESSLER6))) %>%
  select(W2SEXMINID, GEDUC1, GURBAN, GCENREG, COHORT, W2WEIGHT, W1RACE, W1SEX_GENDER, W1KESSLER6, W2KES
  drop_na()

missing_num <- count(full_df) - count(complete_df) %>% pull()
```

The dataset that was used for this report consisted of survey data from Wave 1 and Wave 2 of the Generations study that was conducted from 2017 through 2019. This study sought to explore a variety of outcomes, ranging from personal and social wellbeing to health outcomes, for members of the LGB population in the US. With context to the aims of my investigation, I was mainly interested in demographic variables that described the background characteristics of the surveyed individuals and how those were associated with mental health outcomes. The variables of interest are delineated below.

- Kessler-6 Score - the Kessler-6 is a questionnaire that asks participants how often they have felt “nervous,” “hopeless,” “restless or fidgety,” “so depressed that nothing could cheer you up,” “that everything was an effort,” and “worthless” in the past 30 days, and is used as a clinical screening tool for assessing if an individual is suffering from Major Depressive Disorder (MDD) (@kessler). This score was the outcome of interest in assessing the mental health outcomes of respondents.
- Sexuality - this variable is a categorical variables that categorizes respondents as either “gay/lesbian”, “bisexual”, or “other”, depending on the write-in responses they provided for how they label their sexuality.
- Education - this variable is a categorical variables that provides information on a respondent’s level of education as “high school or less”, “some college”, “college completed”, or “more than college completed.”
- Urbanicity - this variable encodes whether a respondent lives in a locale that is considered urban, as defined in the Generations study technical notes, or if a respondent lives in a locale that is not considered urban.
- Geographical Census Region - encodes which geographical census region a respondent is from, and the census regions are defined as
 - Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
 - Midwest: Indiana, Illinois, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
 - South: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, Texas
 - West: Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada, Wyoming, Alaska, California, Hawaii, Oregon, Washington
- Cohort - variable that characterizes whether a respondent belongs to the “young” (16-27), “middle” (32-43), and “old” (50-61) age range.

Objectives

Exploratory Data Analysis

Methodology

Model Selection

Variable Selection

Model Specification

Model Diagnostics

Results

Discussion

Conclusion

Limitations and Future Work

Summary