STA 440 Final Project

Shirley Mathur

4/14/2022

Introduction

Background

Note to self:Already know of difference between LGB vs. Het, now in this study want to examine differences by other demographic characteristics amongst LGB individuals to see how they are differentially impacted even though they are all LGB individuals.

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

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.
- Race - variable that characterizes the self-reported race of a respondent as being one of: White, Hispanic/Latino, Asian, Multiracial, Middle Eastern, American Indian, Black/African American, or Native Hawaiian/Pacific Islander.
- Sex and Gender - variable that combines sex and gender information to classify respondents as one of: women, non-transgender; men, non- transgender; genderqueer/non- binary (GQNB), female; GQNB, male.
- Poverty - variable that characterizes whether a respondent lives in poverty or not, where living in poverty is defined as living below 100% Federal Poverty Line (FPL) according to the FPL calculated in 2017.

For the final dataset, I selected the above variables from the original dataset that contained 894 observations for participants who participated in both Wave 1 and Wave 2 and 929 variables, of which majority were survey questions that were used to compute measures for outcomes of interest (such as the Kessler-6 Score). I removed all observations that had missing data for any of the variables of interest, which resulted in dropping 57 observations and a final dataset containing 837 observations.

Objective

The main objective of this report is to explore the associations between various demographic characteristics of a LGB individual and their mental health outcomes. Specifically, in the context of this study, since participants were measured at Wave 1 and Wave 2, I want to examine how, if at all, a participant's mental health outcome changed over the course of the year and what associations this change might have with the background characteristics and an individual. This would allow us to gain insight into how other societal factors and minority statuses that LGB individuals hold might impact their mental health outcomes.
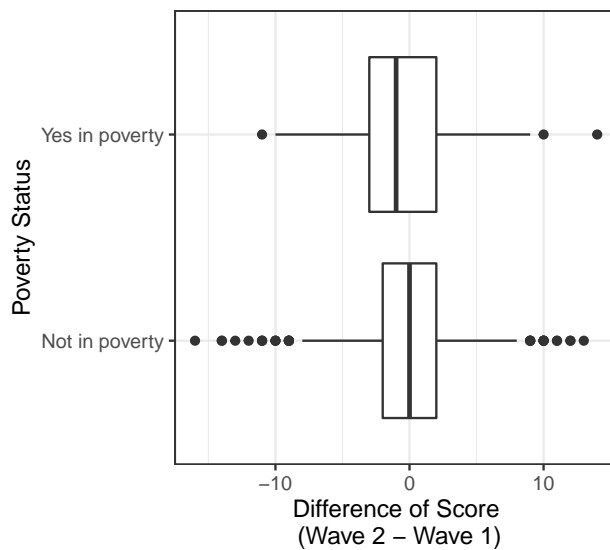
Exploratory Data Analysis

Distribution in Change of Kessler−6 Scores Among Participants
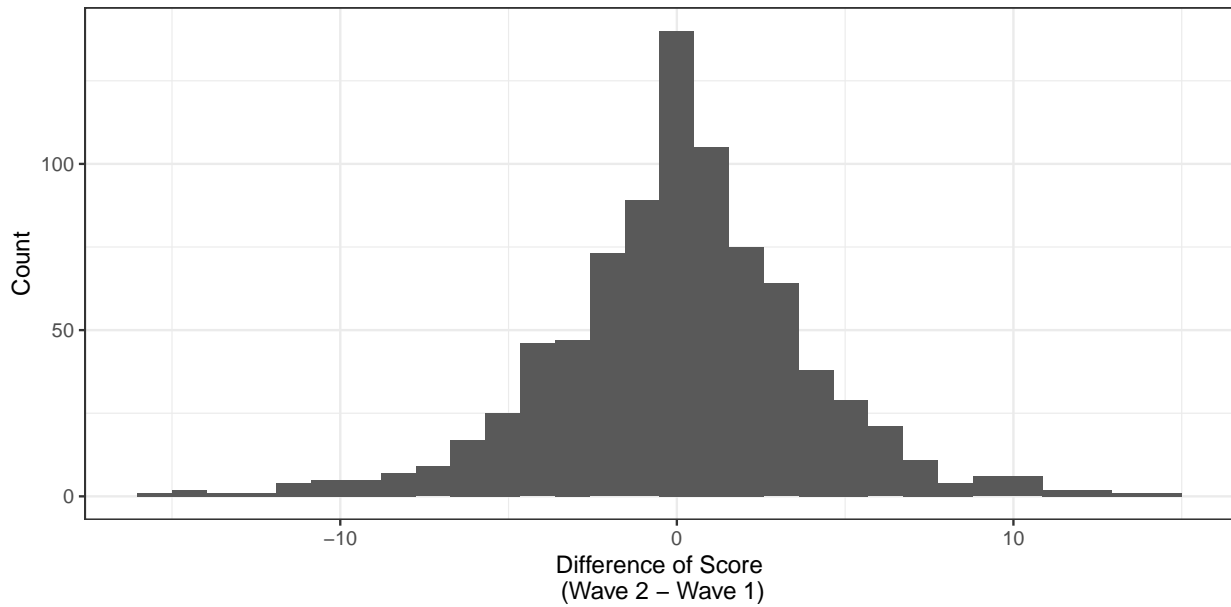
...ler−6 Scores from Wave 1 survey to Wave 2 survey. As seen above, the distribution of difference in score appears to be approximately normal.
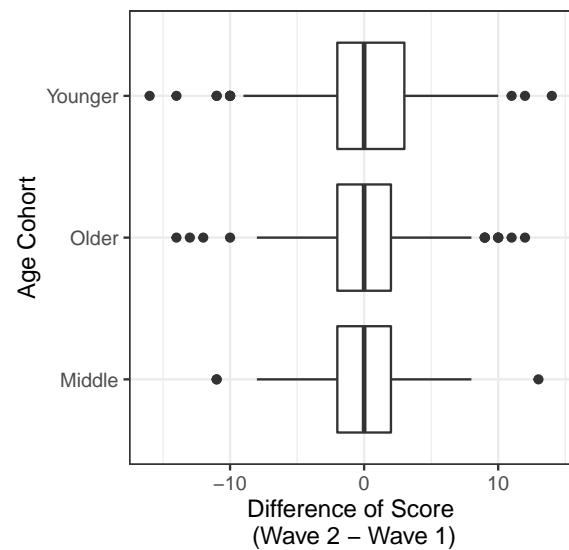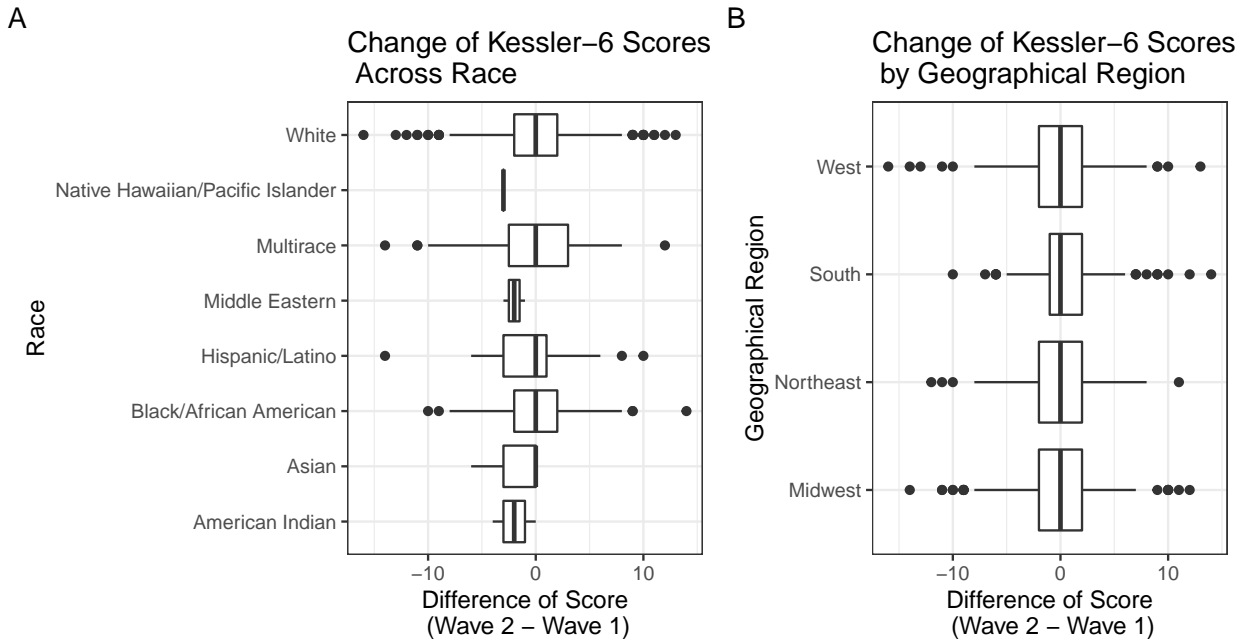
A

Change of Kessler−6 Scores Across Poverty Status

B

Change of Kessler−6 Scores by Age Cohort

...scores from Wave 1 survey to Wave 2 survey amongst participants with varying demographic characeristics of poverty status and age cohort.

A — Change of Kessler–6 Scores Across Race

B — Change of Kessler–6 Scores by Geographical Region

1 survey to Wave 2 survey amongst participants with varying demographic characeristics of racial identity and geographic region of residence.

From Figure 1 above, we can observe that the distribution of difference in Kessler-6 score is approximately normal amongst the sample of participants that participated in both waves of the Generations survey.

From Figure 2A, we can see that the median change in Kessler-6 Score for individuals who live below the FPL appears to be lower than the median change in Kessler-6 Score for individuals who live above the FPL. Thus, we might want to consider poverty status as a covariate in our model to see whether there is a statistically significant association between poverty status and change in Kessler-6 Score.

From Figure 2B, we can see that the median change in Kessler-6 Score for individuals across all of the age cohorts appears to be relatively the same. However, we would need to include age cohort as a covariate in our model to see whether there is a statistically significant association between age cohort and change in Kessler-6 Score.

From Figure 2C, we can see that the median change in Kessler-6 Score seems to be different for different racial subgroups of the survey sample. From the plot, it appears that Multiracial, White, Hispanic/Latino, Black/African American, and Asian individuals have approximately the same median change in Kessler-6 Score. However, it appears that Native Hawaiian/Pacific Islander and Middle Eastern individuals have a lower median Kessler-6 Score than individuals in all of the other racial categories. Due to the apparent differences in the distributions for change in Kessler-6 Score across the different racial groups, we might want to consider race as a covariate in our model to see whether there is a statistically significant association between race and change in Kessler-6 Score.

From Figure 2D, we can see that the median change in Kessler-6 Score seems to be roughly the same across all of the various geographic census regions. However, while all of the regions seem to have fairly symmetric distributions, the South seems to be skewed left, perhaps implying that more individuals in the South region had a positive change in Kessler-6 score. Given that even though the median is roughly the same, but the distributions vary, it might be useful to include geogrphic region as a covariate in our model to assess whether it has a statistically significant association with change in Kessler-6 Score.

Methodology

Model Selection

For modeling the relationship between the covariates of interest and the difference in Kessler-6 outcomes between the Wave 1 survey and Wave 2 survey, I decided to utilize a mixed effects linear regression model. The rationale behind this modeling approach is that while all of the covariates of interest are categorical variables, some have only two factors, such as poverty status, whereas others have many more, such as race. Furthermore, the variables that have many more covariates, such as race, seem to have imbalanced counts of different categories, which is where having random intercepts by race could be useful in modeling the association between racial category and change in Kessler-6 Score.

Variable Selection

Model Specification

$$Y_i = \beta_0 + \beta_1(\text{Sexuality}_i = \text{Bisexual}) + \beta_2(\text{Sexuality}_i = \text{Other Sexual Minority}) + \beta_3(\text{Education}_i = \text{Some college})$$
$$+ \beta_4(\text{Education}_i = \text{College}) + \beta_5(\text{Education}_i = \text{More than college}) + \beta_6(\text{Urbanicity}_i = \text{Urban})$$
$$+ \beta_7(\text{Region}_i = \text{Midwest}) + \beta_8(\text{Region}_i = \text{South}) + \beta_9(\text{Region}_i = \text{West}) + \beta_{10}(\text{Age Cohort}_i = \text{Middle})$$
$$+ \beta_{11}(\text{Age Cohort}_i = \text{Older}) + \beta_{12}(\text{Sex, Gender}_i = \text{Men, non-transgender})$$
$$+ \beta_{13}(\text{Sex, Gender}_i = \text{Genderqueer/Non-Binary, Female}) + \beta_{14}(\text{Sex, Gender}_i = \text{Genderqueer/Non-Binary, Male})$$
$$+ \beta_{15}(\text{Poverty Status}_i = \text{Below FPL}) + b_{0j} + \epsilon_{ij}$$

$$\text{where} \beta_{0i} \sim N(0, \sigma_0)$$

Model Diagnostics

Results

Table 1: Summary of Parameter Posterior Distributions

Variables	Posterior Mean	2.5th Percentile	97.5th Percentile
Intercept	0.270	-1.641	2.143
Sexual Identity Coefficients			
Bisexual	-0.155	-0.815	0.499
Other Sexual Minority	0.142	-0.757	1.058
Education Coefficients			
Some College	-0.472	-1.267	0.351
College	-0.330	-1.077	0.413
More than College	-0.009	-0.664	0.649
Urbanicity Coefficients			
Urban	0.093	-0.691	0.872
Geographical Region Coefficients			
Midwest	0.043	-0.792	0.876
South	0.946	0.179	1.708
West	0.224	-0.550	1.009
Age Cohort Coefficients			
Middle Age Cohort	0.144	-0.539	0.803
Older Age Cohort	0.111	-0.627	0.827
Sex and Gender Coefficients			
Men, Non-transgender	-1.705	-3.769	0.368
Genderqueer Non-binary, Female	-0.742	-2.152	0.615
Genderqueer Non-binary, Male	-0.537	-1.923	0.780
Poverty Status Coefficients			
Below FPL	-0.417	-1.257	0.439
Random Intercepts by Race			
Sigma for Race Intercept	0.410	0.020	1.364
Asian	-0.048	-1.164	0.887
Black/African American	-0.087	-1.326	0.908
Hispanic/Latino	0.111	-0.559	1.014
Middle Eastern	-0.102	-0.919	0.621
Native Hawaiian/Pacific Islander	-0.038	-1.057	0.924
White	-0.038	-0.760	0.701
American Indian	-0.052	-1.264	0.955
Multiracial	0.226	-0.283	0.980

Table 2: Summary of Conditional Means of Comparisons of Race Intercepts (66th Percentile Intervals)

Race Comparison	Lower Quantile	Conditional Mean	Upper Quantile
Hispanic/Latino - Black/African American	-0.6069652	-0.1219018	0.1268060
Native Hawaiian/Pacific Islander - Black/African American	-0.5625016	-0.0379788	0.2749012
Middle Eastern - Black/African American	-0.5758847	-0.0441052	0.2472536
Multiracial - Black/African American	-0.5185077	-0.0789970	0.1655548
Hispanic/Latino - American Indian	-0.4553600	-0.0350438	0.3031430
Asian - American Indian	-0.4607878	-0.0047750	0.4095088
Hispanic/Latino - Asian	-0.4440438	-0.0185441	0.3812692
Native Hawaiian/Pacific Islander - Multiracial	-0.3768565	0.0096023	0.4268930
Native Hawaiian/Pacific Islander - Middle Eastern	-0.4517237	-0.0004304	0.4278856
Native Hawaiian/Pacific Islander - American Indian	-0.4248480	0.0050036	0.4634834
Multiracial - Middle Eastern	-0.4002284	-0.0059728	0.3738069
Middle Eastern - American Indian	-0.4179411	0.0012498	0.4473560
Multiracial - American Indian	-0.3958191	-0.0050149	0.3709469
Native Hawaiian/Pacific Islander - Asian	-0.4011761	0.0126477	0.4854862
Middle Eastern - Asian	-0.3863392	0.0112772	0.4841614
Multiracial - Asian	-0.3647146	0.0017183	0.4325589
Native Hawaiian/Pacific Islander - Hispanic/Latino	-0.3038000	0.0391255	0.4844674
Middle Eastern - Hispanic/Latino	-0.3217003	0.0334919	0.4882795
Multiracial - Hispanic/Latino	-0.2362806	0.0314052	0.3836777
White - Black/African American	-0.1541605	0.0746608	0.4091669
Black/African American - American Indian	-0.2508776	0.0518294	0.5766827
Black/African American - Asian	-0.2191250	0.0644682	0.6437496
White - Multiracial	-0.0348269	0.1983659	0.6012106
White - Middle Eastern	-0.1130099	0.1396055	0.6961362
White - American Indian	-0.1032948	0.1456271	0.6835856
White - Native Hawaiian/Pacific Islander	-0.1255469	0.1385871	0.6899715
White - Asian	-0.0951765	0.1597954	0.7665180
White - Hispanic/Latino	-0.0099557	0.2488704	0.7112958

Table 3: Summary of Conditional Means of Comparisons of Race Intercepts (95th Percentile Intervals)

Race Comparison	Lower Quantile	Conditional Mean	Upper Quantile
Hispanic/Latino - Black/African American	-1.2698270	-0.1219018	0.5314615
Native Hawaiian/Pacific Islander - Black/African American	-1.7255636	-0.0379788	0.9722469
Middle Eastern - Black/African American	-1.6394566	-0.0441052	0.9270387
Multiracial - Black/African American	-1.1155690	-0.0789970	0.6251284
Hispanic/Latino - American Indian	-1.3420128	-0.0350438	1.3328738
Asian - American Indian	-1.6990791	-0.0047750	1.3794118
Hispanic/Latino - Asian	-1.2490293	-0.0185441	1.3821421
Native Hawaiian/Pacific Islander - Multiracial	-1.4398761	0.0096023	1.2525232
Native Hawaiian/Pacific Islander - Middle Eastern	-1.5837765	-0.0004304	1.4445310
Native Hawaiian/Pacific Islander - American Indian	-1.6990586	0.0050036	1.6058973
Multiracial - Middle Eastern	-1.2020672	-0.0059728	1.3135137
Middle Eastern - American Indian	-1.4566054	0.0012498	1.4608530
Multiracial - American Indian	-1.1531018	-0.0050149	1.4182255
Native Hawaiian/Pacific Islander - Asian	-1.6343722	0.0126477	1.6781250
Middle Eastern - Asian	-1.4785204	0.0112772	1.6525923
Multiracial - Asian	-1.1405999	0.0017183	1.5083461
Native Hawaiian/Pacific Islander - Hispanic/Latino	-1.2560010	0.0391255	1.3838883
Middle Eastern - Hispanic/Latino	-1.2028227	0.0334919	1.3531045
Multiracial - Hispanic/Latino	-0.7378790	0.0314052	0.9456370
White - Black/African American	-0.5829397	0.0746608	0.8836359
Black/African American - American Indian	-0.9390413	0.0518294	1.6840134
Black/African American - Asian	-0.9402577	0.0644682	1.9121986
White - Multiracial	-0.3090280	0.1983659	1.0669075
White - Middle Eastern	-0.6658782	0.1396055	1.6691092
White - American Indian	-0.6769844	0.1456271	1.7736021
White - Native Hawaiian/Pacific Islander	-0.7884975	0.1385871	1.8221440
White - Asian	-0.6642948	0.1597954	1.8191185
White - Hispanic/Latino	-0.2463898	0.2488704	1.2454000

Discussion

Conclusion

Limitations and Future Work

Summary

Appendix

Model Diagnostics

Plot of Conditional Mean of Posterior Comparisons for Race Intercepts