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

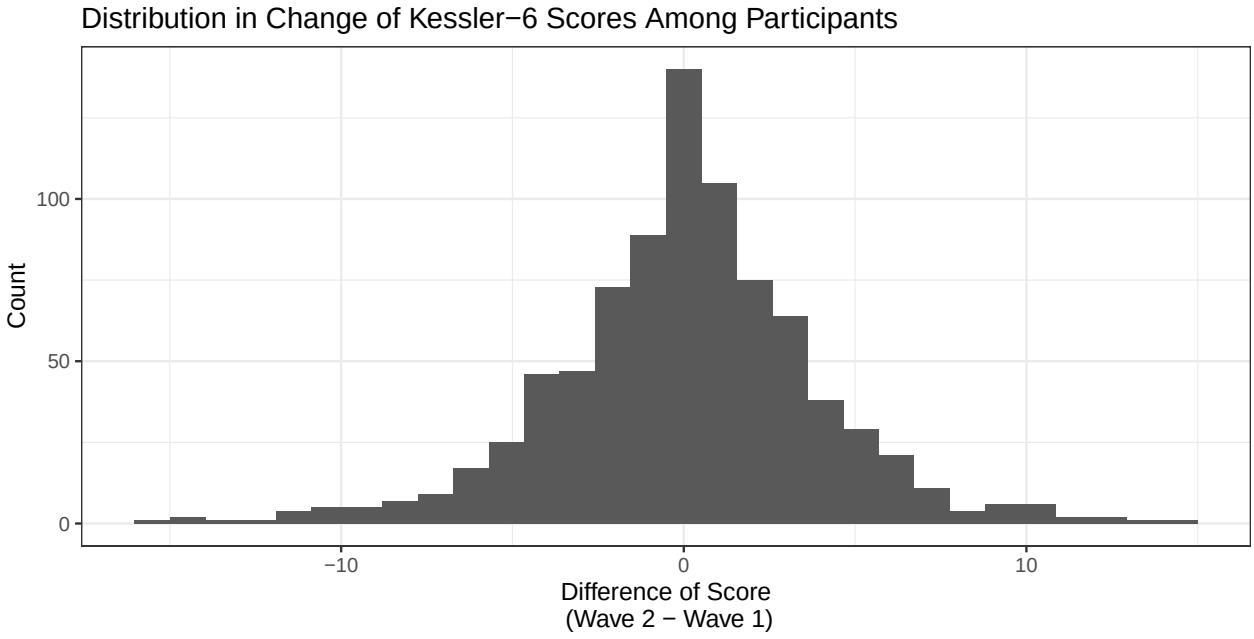

Figure 1: Distribution in change of Kessler−6 Scores from Wave 1 survey to Wave 2 survey. As seen above, the distribution of difference in score appears to be approximately normal.

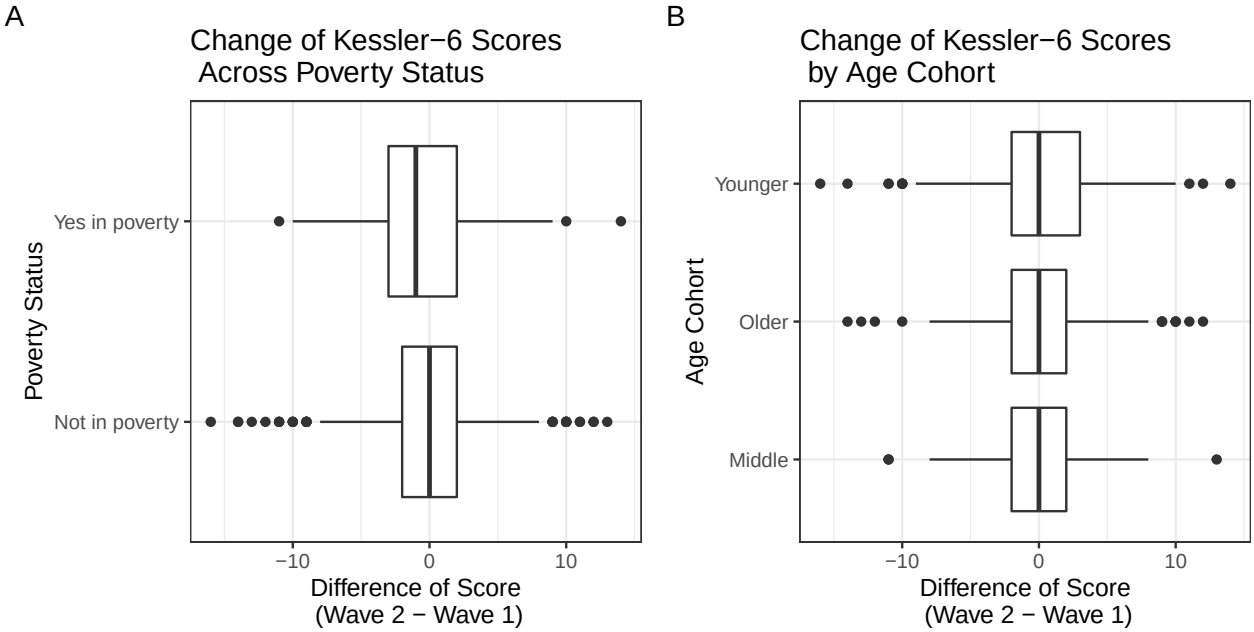

Figure 2: Distribution of changes in Kessler−6 scores from Wave 1 survey to Wave 2 survey amongst participants with varying demographic characeristics of poverty status and age cohort.

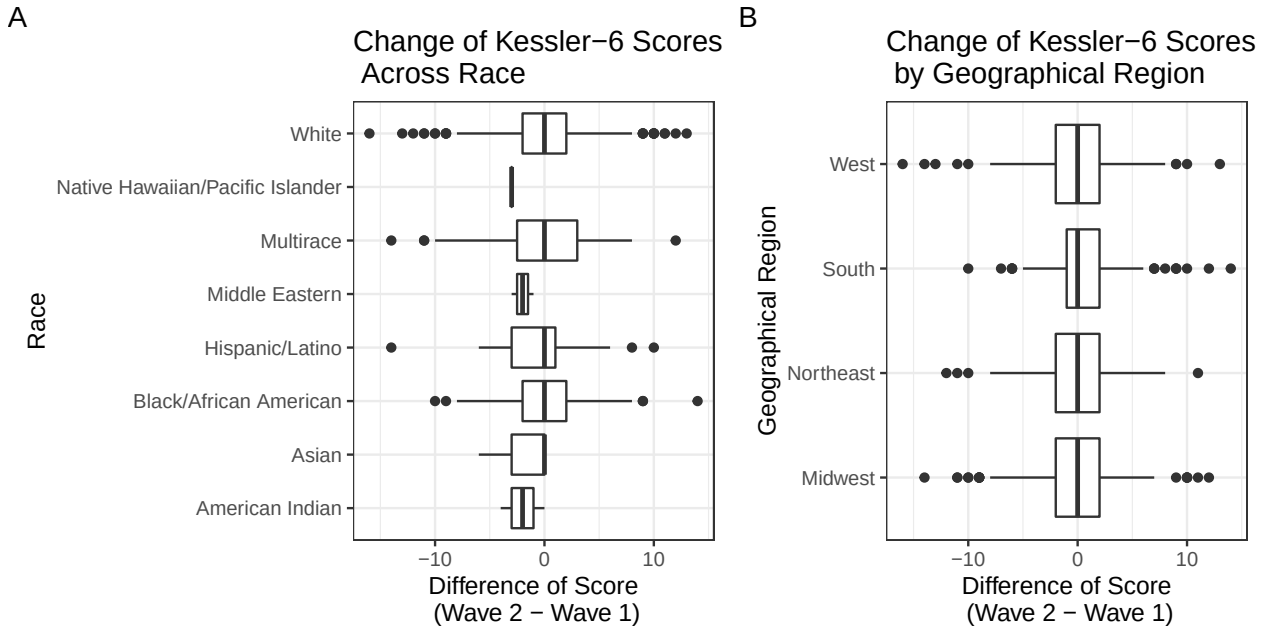

Figure 3: Distribution of changes in Kessler−6 scores from Wave 1 survey to Wave 2 survey amongst participants with varying demographic characeristics of racial identity and geographic region of residence.

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

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

$$\text{where } \beta_{0i} \sim N(0, \sigma_0) \perp \epsilon_{ij} \sim N(0, \sigma)$$

Model Diagnostics

Results

Table 1: Summary of Parameter Posterior Distributions

Variables	Posterior Mean	2.5th Percentile	97.5th Percentile
Intercept	0.317	-1.595	2.214
Sexual Identity Coefficients			
Bisexual	-0.152	-0.812	0.488
Other Sexual Minority	0.149	-0.736	1.033
Education Coefficients			
Some College	-0.471	-1.275	0.325
College	-0.327	-1.063	0.378
More than College	-0.004	-0.652	0.618
Urbanicity Coefficients			
Urban	0.099	-0.719	0.887
Geographical Region Coefficients			
Midwest	0.025	-0.791	0.850
South	0.935	0.175	1.681
West	0.210	-0.538	0.963
Age Cohort Coefficients			
Middle Age Cohort	0.126	-0.592	0.816
Older Age Cohort	0.093	-0.605	0.825
Sex and Gender Coefficients			
Men, Non-transgender	-1.738	-3.896	0.379
Genderqueer Non-binary, Female	-0.770	-2.248	0.655
Genderqueer Non-binary, Male	-0.561	-1.949	0.805
Poverty Status Coefficients			
Below FPL	-0.414	-1.275	0.460
Random Intercepts by Race			
Sigma for Race Intercept	0.402	0.014	1.319
Asian	-0.038	-1.143	0.942
Black/African American	-0.093	-1.286	0.829
Hispanic/Latino	0.106	-0.536	0.977
Middle Eastern	-0.105	-0.871	0.559
Native Hawaiian/Pacific Islander	-0.040	-1.205	0.977
White	-0.037	-0.745	0.665
American Indian	-0.050	-1.270	0.985
Multiracial	0.216	-0.293	1.023

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	-0.5827167	-0.1152042	0.1162828
Native Hawaiian/Pacific Islander - Black/African American	-0.5655682	-0.0498079	0.2303455
Middle Eastern - Black/African American	-0.5675447	-0.0504845	0.2446301
Multiracial - Black/African American	-0.4955545	-0.0758937	0.1594725
Hispanic/Latino - American Indian	-0.4811314	-0.0308782	0.3086171
Asian - American Indian	-0.4883211	-0.0090667	0.3616218
Native Hawaiian/Pacific Islander - Multiracial	-0.3824648	0.0054739	0.3836728
Hispanic/Latino - Asian	-0.4282597	-0.0126571	0.3779903
Native Hawaiian/Pacific Islander - American Indian	-0.4496811	0.0001356	0.3988327
Native Hawaiian/Pacific Islander - Middle Eastern	-0.4350340	-0.0011114	0.4197258
Middle Eastern - American Indian	-0.4099026	-0.0036404	0.4151577
Multiracial - American Indian	-0.3981948	-0.0130832	0.3897499
Multiracial - Middle Eastern	-0.4007788	-0.0072151	0.3790797
Native Hawaiian/Pacific Islander - Asian	-0.3787812	0.0093681	0.4549788
Middle Eastern - Asian	-0.3655739	0.0096456	0.4847659
Native Hawaiian/Pacific Islander - Hispanic/Latino	-0.3134268	0.0306686	0.4599756
Multiracial - Asian	-0.3453053	0.0019533	0.4392377
Middle Eastern - Hispanic/Latino	-0.3051738	0.0329289	0.4869333
Multiracial - Hispanic/Latino	-0.2282448	0.0225537	0.3907637
White - Black/African American	-0.1505775	0.0686158	0.4050884
Black/African American - American Indian	-0.2432279	0.0452449	0.5471086
Black/African American - Asian	-0.2094858	0.0712043	0.6410606
White - Multiracial	-0.0289933	0.1961758	0.5771406
White - American Indian	-0.1119127	0.1282479	0.6912064
White - Middle Eastern	-0.1276746	0.1255324	0.7065473
White - Native Hawaiian/Pacific Islander	-0.1208983	0.1276097	0.6944758
White - Asian	-0.0878421	0.1560064	0.7754680
White - Hispanic/Latino	-0.0154703	0.2423211	0.7101067

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	-1.2950253	-0.1152042	0.5642120
Native Hawaiian/Pacific Islander - Black/African American	-1.7484422	-0.0498079	0.9734841
Middle Eastern - Black/African American	-1.6670188	-0.0504845	0.9570659
Multiracial - Black/African American	-1.0620249	-0.0758937	0.6160414
Hispanic/Latino - American Indian	-1.3654874	-0.0308782	1.2774968
Asian - American Indian	-1.5419146	-0.0090667	1.2912494
Native Hawaiian/Pacific Islander - Multiracial	-1.3942974	0.0054739	1.1928661
Hispanic/Latino - Asian	-1.2280869	-0.0126571	1.2688411
Native Hawaiian/Pacific Islander - American Indian	-1.5289784	0.0001356	1.5001777
Native Hawaiian/Pacific Islander - Middle Eastern	-1.5731264	-0.0011114	1.5264837
Middle Eastern - American Indian	-1.4857054	-0.0036404	1.4472780
Multiracial - American Indian	-1.1725294	-0.0130832	1.3675559
Multiracial - Middle Eastern	-1.1821528	-0.0072151	1.4382263
Native Hawaiian/Pacific Islander - Asian	-1.3562921	0.0093681	1.6604798
Middle Eastern - Asian	-1.3983340	0.0096456	1.5898930
Native Hawaiian/Pacific Islander - Hispanic/Latino	-1.2580513	0.0306686	1.3165347
Multiracial - Asian	-1.0578507	0.0019533	1.4761841
Middle Eastern - Hispanic/Latino	-1.2985162	0.0329289	1.3805923
Multiracial - Hispanic/Latino	-0.7280848	0.0225537	0.9588722
White - Black/African American	-0.5448463	0.0686158	0.8678598
Black/African American - American Indian	-0.9374653	0.0452449	1.6490699
Black/African American - Asian	-0.8527941	0.0712043	1.7630601
White - Multiracial	-0.3200038	0.1961758	1.0208959
White - American Indian	-0.7354615	0.1282479	1.7227078
White - Middle Eastern	-0.8254826	0.1255324	1.7928109
White - Native Hawaiian/Pacific Islander	-0.7287545	0.1276097	1.8078220
White - Asian	-0.6427828	0.1560064	1.8078458
White - Hispanic/Latino	-0.2667810	0.2423211	1.2117924

Discussion

Conclusion

Limitations and Future Work

Summary

Appendix

Model Diagnostics

Plot of Conditional Mean of Posterior Comparisons for Race Intercepts