Looking at Schizophrenia-Related Suicide Data

Victor Chen, Katie Chuang

Introduction

What groups or demographics are more susceptible to schizophrenia-related suicide? A better understanding of the answer to this question can help further understanding of cause, awareness, and prevention of schizophrenia-related suicide.

Our group analyzed a dataset of schizophrenia-related suicide deaths that we extracted from the Centers for Disease Control and Prevention's WONDER database of United States deaths and death factors (https://wonder.cdc.gov/mcd.html). We were curious about how gender, race, age, and time (year) trend with schizophrenic death by suicide? In other words, when and what kind of people were / are more susceptible to schizophrenia related suicide?

Variables/Hypotheses

We isolated each covariate and predicted the impact of each one on the predicted number of deaths separately:

1. There are more white individuals who have died from schizophrenia related suicide than compared to other races.
2. Schizophrenia related suicide has become more common in recent years.
3. More death by suicide by those who identify as male, as previous studies have shown suicide rates in males are around 3 times more common than in females generally.
4. Most deaths will be from those of 45-54 and/or the 55-64 years age group as studies have also shown that middle-aged men are the population that make up the largest chunk of suicide commits.

To reiterate, the response variable we considered is the number of deaths by suicide of schizophrenics from 1999-2020. The covariates we considered for prediction are the following: year, age, gender, and race.

Findings

Regarding our hypotheses, some of them were affirmed by the data, while others were contracted. On our hypothesis about race, we found that in general, someone who died from schizophrenia-related suicide is more likely to be white than another race group, which aligned with our prediction (slide 6). We suspect this to be because white individuals make up a larger portion of the total population than other races. Our hypothesis about gender was also

confirmed, we saw that this dataset had nearly 4 times as many male deaths than female deaths. We suspect this to be because the trend in suicides overall is mostly male. On the other hand, our hypothesis about age was contradicted. While we predicted the most deaths being from the 45-54/55-64 year age groups, the most deaths in total, came from the 25-34 age group (slide 4). We suspect this is because this age group is where schizophrenia is most common. Our hypothesis on time (year) was also contradicted. We predicted that the rate of suicides would increase over time, however our visualizations (slide 7) suggested that the rate of suicide has not significantly increased, and instead has stayed consistent over time. This does not match the general trends in rates of schizophrenia and suicide seperately which have both increased over time. Based on these findings, our takeaway to prevent more deaths by suicide in schizophrenic community is to consider monitoring and targeting awareness at white males who fall in the 25-34 year age range.

Regression Analysis

We decided to take what we found in our preliminary analysis that the most deaths have occurred in 25-34 year olds over the years and set age to be constant as well as year, where we took a look at the most recent years, both 2019 and 2020 in case COVID skewed our response in any way, to see if our findings for race and gender still hold. If they do, we hope to draw some conclusions and suggest possible target groups for therapy or other forms of treatment to prevent deaths by suicide in the schizophrenic community in the future (slide 9)

We examined the following regression equation:
$Death_i = \beta_0 + \beta_1 gender\_male_i + \beta_2 race\_aapi_i + \beta_3 race\_black\&aa_i + \beta_4 race\_white_i + \varepsilon_i$

Since the confidence intervals obtained for the coefficient on Gender (Male) and Race (White) both did not include 0, (with H_0 assumed to be $H_0 : \beta_i = 0$) the coefficient on these respective variables is statistically significant. These findings support both our gender and race hypotheses under both year 2019 and year 2020 (slide 8).

Looking back on it, we understand that this regression is flawed due to the predicted variable being too reliant on differences in population size. Doing this regression with a predicted variable representing a rate or proportion would have been more useful.