

625FinalPaper

Caitlin Seibel

2023-12-11

Introduction

Illicit substance use is known to have several health related consequences, including cardiovascular disease, cancer, and death [1]. Opioid misuse has risen in the United States from the first wave of prescription opioid overdoses in the 1990s to the rise of synthetic opioid use (i.e. fentanyl) in 2013 [2], calling for concern on (potentially) increasing deaths from drug or alcohol related causes. We analyzed data obtained from the Centers for Disease Control and Prevention (CDC) [3] and the U.S. Census Bureau [4] to make inferences on drug and alcohol related death counts, adjusting for age, gender, racial strata, and an indicator of the COVID-19 pandemic.

Table 1 shows counts of demographic variables we have available in this data set, as well as the counts for the cause of death. Note that 65% of deaths are from non-drug and non-alcohol induced causes since these include all deaths unrelated to drugs or alcohol. Also note that the standard deviation of the deaths is about 3800 with a approximate mean of 957, indicating high variability in the deaths. Plotting two visualizations of this shows that there is high variability in the death counts.

Table 1: Descriptive statistics

	Overall
n	12064
Deaths (mean (SD))	956.53 (3799.72)
Cause_of_Death (%)	
All other non-drug and non-alcohol causes	7879 (65.3)
Alcohol-induced causes	1821 (15.1)
Drug-induced causes	2364 (19.6)
Age (%)	
Under 5 years	436 (3.6)
5-9 years	194 (1.6)
10-14 years	193 (1.6)
15-19 years	439 (3.6)
20-24 years	606 (5.0)
25-29 years	763 (6.3)
30-34 years	854 (7.1)
35-39 years	896 (7.4)
40-44 years	889 (7.4)
45-49 years	944 (7.8)
50-54 years	979 (8.1)
55-59 years	992 (8.2)
65-69 years	927 (7.7)
70-74 years	862 (7.1)
75-79 years	769 (6.4)
80-84 years	717 (5.9)
85 years and over	604 (5.0)
Race (%)	
White	3960 (32.8)
American Indian or Alaska Native	1577 (13.1)
Asian	1437 (11.9)
Black or African American	3025 (25.1)
More than one race	1397 (11.6)
Native Hawaiian or Other Pacific Islander	668 (5.5)
Gender = Female (%)	5580 (46.3)
COVID = 1 (%)	5864 (48.6)

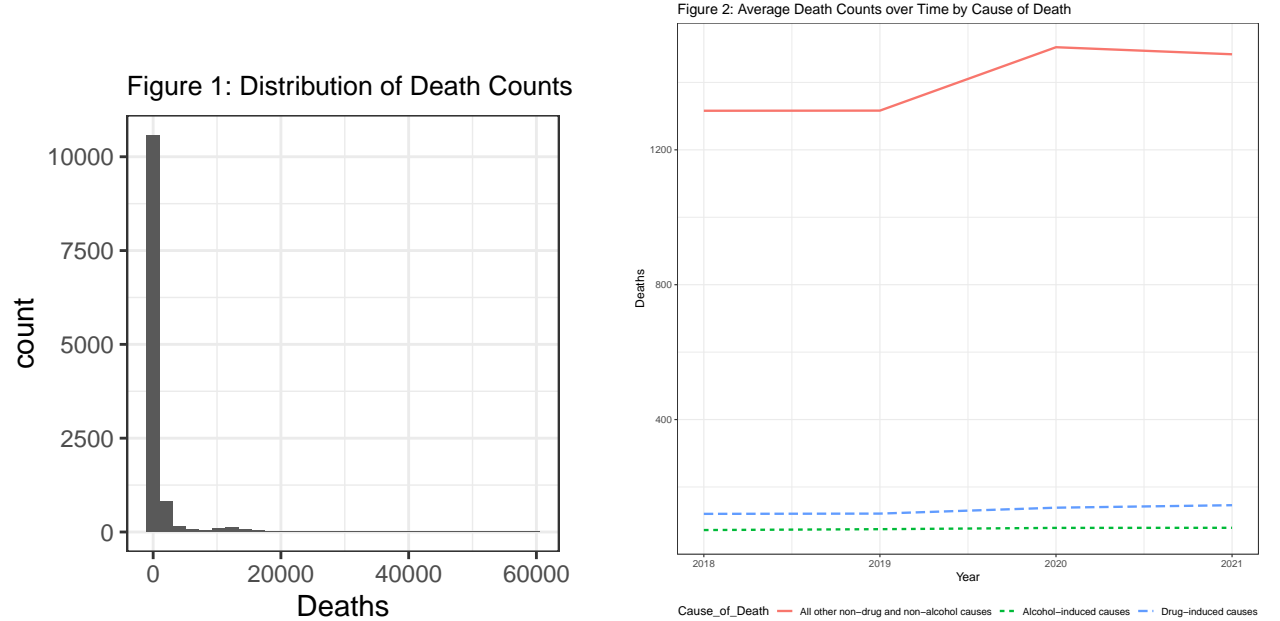
Methods

Data Preprocessing

Deaths from drug or alcohol use is defined as: if drugs or alcohol was the catalyst in a series of events that led to death [2]. 2018-2021 death count data obtained from the CDC was requested based on five-year age categories, gender (male/female), six racial groups (American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, and More Than One Race), cause of

death (Drug, Alcohol, and Non-drug/non-alcohol), and month of death. Population data for these strata groupings were unavailable.

To make inferences about death-rate differences between substrata populations, 2020-2022 data obtained from the Census used similar strata as the CDC. Two differences between the data were the characterizations of the five-year age groups and monthly/yearly population counts. CDC characterized individuals who died before a year old as their own group; they also count all five year age groups up to 100 as unique groups. The Census collapses these into < 5 and > 85 , with unique categories for five-year groups between these ages. CDC death count data was aggregated to match the Census strata. Further, Census data was only available on yearly 2020-2022 estimates, whereas we have monthly CDC data from 2018. We assumed the population is stable throughout a given year and that the rate of change within substrata is constant between years. By these assumptions, we estimated the substrata population counts for 2018 and 2019.



Looking at the dates included in the model, we assumed that the COVID-19 pandemic will skew the death counts upwards. Plotting death counts over time, we see this assumption may hold, so we created an indicator to distinguish months before and after March 2020, which we are taking as the beginning of the COVID-19 pandemic in the U.S.

Model Building

To analyze our data, we choose a Generalized Linear Model (GLM). Our outcome variable of interest is number of deaths per month, referred to in our model simply as “Deaths”. Our main predictor of interests are cause of death: drug related, alcohol related, or non-drug/alcohol related, and indicator of COVID-19 pandemic: 0 if before the start of the COVID-19 pandemic (defined here as March 2020), or 1 if after. Our other demographic covariates include age group, gender, and race. We also want to investigate the possible interaction between cause of death and the indicator of COVID-19 pandemic. Our outcome is a count variable so we initially assume a Poisson distribution. We use $\log(\text{Population})$ as an offset for the model, since we want to model count rather than rate. Our initial GLM model is as follows.

$$\log(\hat{Deaths}) = \hat{\beta}_0 + \log(Population) + \hat{\beta}_1 Cause + \hat{\beta}_2 Gender + \hat{\beta}_3 Age + \hat{\beta}_4 Race + \hat{\beta}_5 COVID + \hat{\beta}_6 Cause \times COVID$$

Since we are using a Poisson distribution, we need to check for overdispersion, or a violation of the assumption of equal mean and variance. To check for this violation, we will run our model and calculate the deviance

divided by the residual degrees of freedom. If the equal mean and variance assumption is met, we expect this value to be approximately 1. From our model, we calculate this metric as 111.4576, which is much larger than 1, so we have evidence of overdispersion. To account for this overdispersion, we will use a negative binomial outcome for our GLM.

Results

RShiny App

To create an interactive platform through which users can visualize different trends in the data, we created an RShiny app that offers a couple different features and methods that can be used to analyze our data. There are two overarching options in our app; regression model building and trend visualization. The main page, titled “Analysis”, is where a regression model can be built and the second page, titled “Exploratory Plots”, is where data trend visualization can be performed.

The main page of the app features a couple of drop down lists to select different aspects of a regression model. You can first select the desired outcome variable from a list of all available variables in the dataset, and then you can select all of the covariates you’re interested in including in the model. There is also an option to include interaction terms in the model, but this is not necessary to run the regression. Once these variables are selected, the app runs a glm in the background with a negative binomial distribution assumption and an offset of the log of the total population (see above for further model description). The first tab on this page shows the summary of the regression model, with the β estimates, p-values, and other relevant model estimates. The second tab on this page shows the residual plot resulting from the model, with the fitted values on the x-axis and the residual values on the y-axis. The final tab on this page shows the ANOVA output of the model, giving relevant statistics for all selected variables and interaction terms.

The second page provides the option to create exploratory plots to visualize trends in the data. There are two drop-down lists from which the y variable and the variable on which to stratify on can be selected. There are four choices in the drop-down menu for the y-variable. The options are the total population, the total death count, only the alcohol-related deaths, and only the drug-related deaths. If either the alcohol-related deaths or the drug-related deaths are chosen, the app works in the background to create a subset of the full data set that consists only of the relevant cause of death. The time variable on the x-axis is automatically set as Years. After creating the plot, the trends in the selected y variable are visible and separated for each group in the selected stratification variable. The output is color-coded and a key is provided.

Discussion

References

- [1] National Institute on Drug Abuse. What are the other health consequences of drug addiction? 2020, URL: <https://nida.nih.gov/publications/drugs-brains-behavior-science-addiction/addiction-health>
- [2] Centers for Disease Control and Prevention [CDC]. Understanding the Opioid Epidemic. 2023. URL: <https://www.cdc.gov/opioids/basics/epidemic.html>
- [3] Centers for Disease Control and Prevention [CDC], Alcohol Use. 2023. URL: <https://www.cdc.gov/nchs/fastats/alcohol.htm>
- [4] United States Census Bureau [USCB]. National Population by Characteristics 2020-2022. 2023. <https://www.census.gov/data/tables/time-series/demo/popest/2020s-national-detail.html>