

Outline: Estimate the Impact of Opioid Control Policies

Team 9: Raza Lamb, Robert Wan, Roza Ogurlu, and Xiaoyu Chen

For this project, the goal is to analyze the effectiveness of three specific opioid policy changes in three different states in the United States. The three states in question are Florida, Texas, and Washington. In order to evaluate the effect of the policy changes, we will conduct both a pre-post analysis and a difference-in-difference analysis. In the case of Texas and Washington, we will attempt to establish whether or not the trend in the shipment of opioids changed after the policy was implemented. In the case of Florida, we will track the same changes and investigate the trend in opioid deaths. The following outline displays the backward design strategy employed to create an action plan for this analysis.

Final Data

In order to make these plots, here we outline what the final dataset will need to look like. The two key variables are (1) the number of opioid shipments to each county per year, normalized by county population, and (2) the number of opioid overdose death per county per year, normalized by county population. We will need data for all counties within the three states we'd like to investigate. In addition, to establish a comparison group for the difference-in-difference analysis, we will also need data for appropriate counties from other states that did not undergo significant opioid policy changes. Not all counties can be used as a comparison. As we conduct the analysis, we'll need to define what makes a county in the comparison group similar to counties in the three states of interest (Florida, Texas, and Washington). To begin with, we will include the data for all states (and counties within those states) in the U.S.

Raw Data Sets Required

To obtain the final key variables of interest we will need population, the number of opioids shipped to that location per year, and the number of opioid overdose deaths per year for each county. We will need a sample of all counties in the three states of interest (Florida, Texas, and Washington). We will also need appropriate counties from other states that did not undergo significant opioid policy changes. Each row will need to be a county-level observation: i.e., one row will have the number of opioid overdose deaths for a specific county. The policy change years are 2010, 2007, and 2012 for Florida, Texas, and Washington, respectively. The opioid shipment dataset we will be looking at is from 2006-2012, and the overdose dataset is from 2003-2015. For our third dataset, population by county, we will be considering the years 2003-2015 to include data for all the years we need for the analysis.

Taking one step backward, in order to get this final dataset, we will need to start with three raw data sets. First, the raw data for mortality comes from the US Vital Statistics agency. There is a separate data set for each year, so they will need to be combined as part of the cleaning process. Each row in this raw data represents a year-county-type of death observation. This data will also need to be limited to just rows for unintentional overdose deaths. Continually, the second dataset we have is on prescription opioid shipments from the Washington Post. This data is in a significantly different form than our ideal final dataset. Each row represents an opioid shipment, information about the supplier, buyer, date, and location. Therefore, this data will need to be aggregated at the county level. Continually, the information needed on the quantity of drugs differs by row. The quantity can come in several different forms. As part of data cleaning, we will need to standardize the quantity so that they are comparable. Finally, the last dataset we will use is annual population-level estimates for counties in the U.S. This data comes from the U.S. Census. This data is available in multiple forms—for our purposes, we will download two datasets, one containing population estimates for all counties in the U.S. for each year between 2000 and 2010 and another containing the same information for years between 2010 and 2020. Because each row contains multiple years of data, we will need to expand each row into multiple rows and limit the data to the years of interest (i.e., 2003-2015). Subsequently, this data will be merged with the shipment and death data for final analysis. One more item of importance is considering that for efficiency and the sake of best practice, we will store the results of the above cleaning in intermediate datasets on GitHub.

Execution Plan

In order to execute this analysis, we have delegated initial roles and responsibilities for this project, which are subject to change. First, both the mortality data and shipment data will need to be cleaned and transformed into intermediate datasets. Xiaoyu will clean the mortality data, and Robert will clean the shipment data. After these steps are complete, Raza will conduct a code review on the mortality dataset cleaning and Roza will conduct a code review on the prescription dataset cleaning. Concurrently, Raza will clean the population data from the Census to a suitable format, and when ready, will merge the population dataset with the mortality and shipment dataset. The population dataset will be merged into each of the other datasets through State and County variables. For this step, Xiaoyu will conduct the code review. After all of these steps, the data will then be in a suitable place for analysis. Raza and Robert will conduct the analysis on both the mortality and opioid shipment data in Florida, and Roza and Xiaoyu will analyze the opioid shipment data for Washington and Texas.