Project Strategy

Team

Pooja Kabber (Team Lead), Chenying Li, Andrew Kroening, Dingkun Yang

Topic

Since the use and abuse of prescription opioids have led to a considerable rise in opioid addiction and overdose deaths, several policies have been implemented to limit opioid abuse. However, the impact of policies may not be apparent under the influence of multiple factors. There exists the possibility that the regulations' influence on opioid addiction and overdose deaths may be the opposite of what is expected or intended.

For this assignment, we are trying to estimate the effectiveness of three opioid prescription regulation policies on the volume of opioids prescribed and drug overdose deaths. We will conduct this analysis using a pre-post and difference-in-difference analysis, focusing on the states of Florida, Texas, and Washington.

Project Questions

Between 2007 and 2012, Florida, Texas, and Washington undertook policy initiatives to limit opioid abuse. The general concept of these policies is to provide additional oversight and regulation concerning the use and rate of prescribed opioids. The research questions for this analysis are divided due to data availability:

1. What effects do policy changes in Florida and Washington have on the volume of opioids prescribed and the drug overdose deaths in both states?
2. What effect does the policy in Texas have on drug overdose deaths?

The analysis will examine pre-post comparison results for these three policies on the volume of opioids prescribed and drug overdose deaths. We will also explore the difference-in-difference results for these three policies against a body of comparison states.

Project Hypothesis

The nature of opioid prescription and abuse of illegal variants makes the topic of outcomes challenging. The prevalence of alternative substances with exceptionally high mortality rates from overdose is a particularly challenging dynamic of this project. Our hypothesized outcomes are:

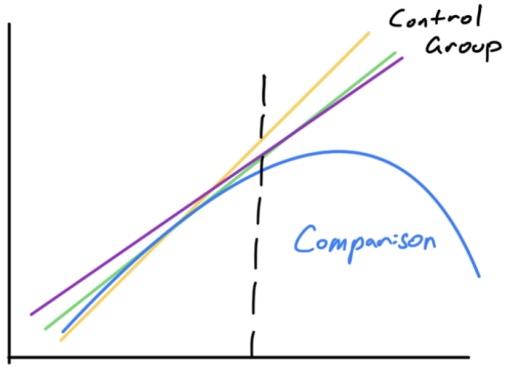
1. Pre-Post Comparison: After comparing drug prescriptions before and after policy changes, we should observe a decreasing number of drug prescriptions. Due to the increased use of more dangerous, illegal substitutes by the addicted, drug overdose deaths should theoretically increase. If the policy is unsuccessful, the drug prescriptions should either stay the same or increase with no change to drug overdose deaths.

2. Difference-in-Difference: Drug prescriptions and overdoses in the experiment group will be compared against the control group. The states with policy changes should have a noticeable effect on these variables, more significant than the changes in the states without policy changes over the same period. This indicates a successful policy. If the difference is not more prominent, then the decrease in drug prescriptions and the increase in overdoses cannot be attributed to the policy.

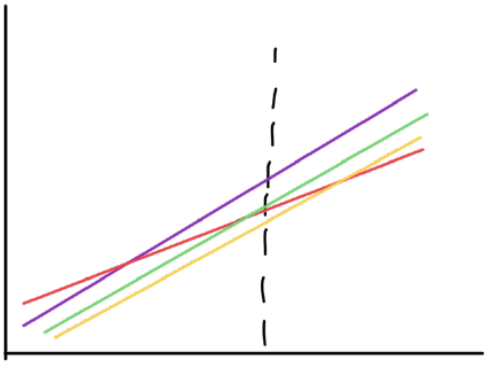
Model Results

Difference-in-difference:

If the hypothesized impact of policy changes on **drug prescriptions is true**, the prescriptions should decrease after the policy while other areas should not change.

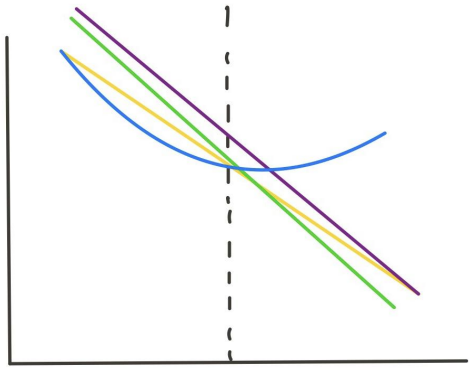


If the hypothesis is *true*

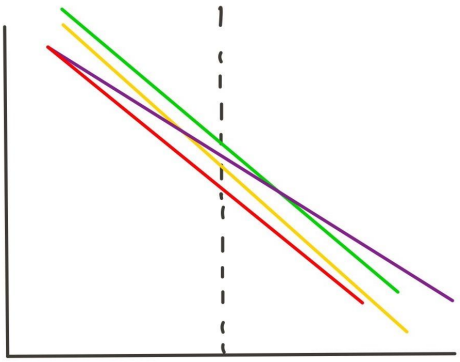


If the hypothesis is *false*

If the hypothesized impact of policy changes on **drug-related overdose deaths is true**, the rate of fatalities should increase after the policy while other areas should not change.



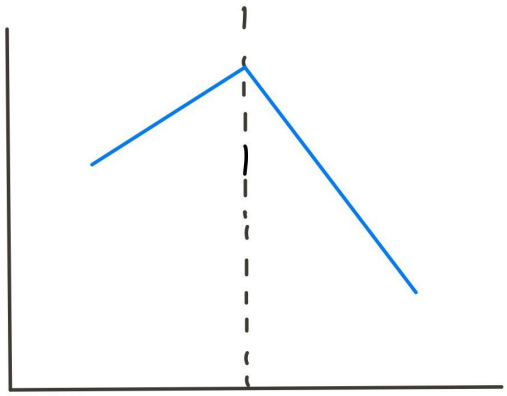
If the hypothesis is *true*



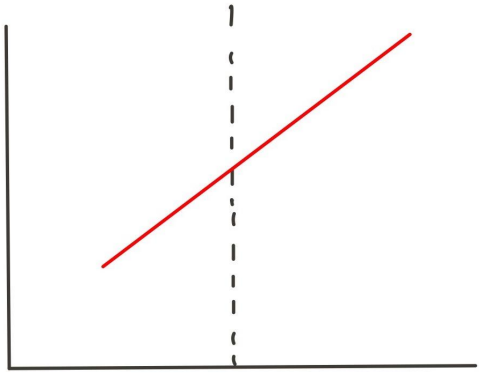
If the hypothesis is *false*

Pre-post comparison:

If the hypothesized impact of policy changes on **drug prescriptions is true**, the prescription rates should decrease after the policy while other areas should not change.

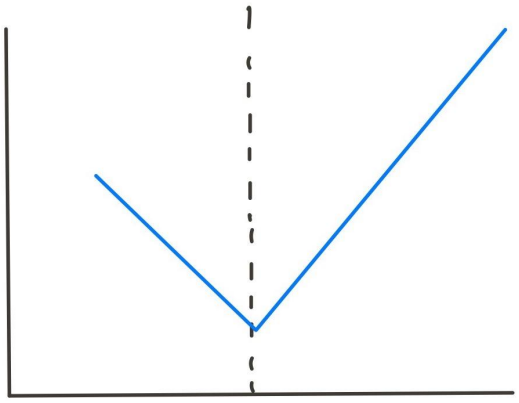


If the hypothesis is *true*

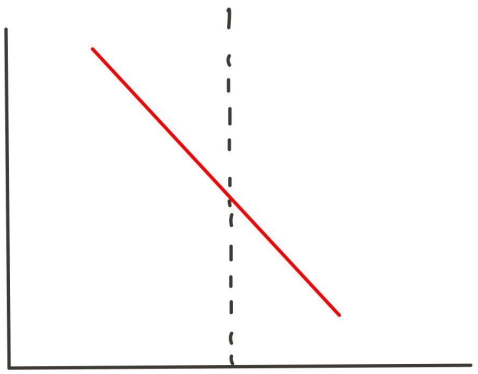


If the hypothesis is *false*

If the hypothesized impact of policy changes on **drug-related overdose deaths is true**, the rate of fatalities should increase after the policy while other areas should not change.



If the hypothesis is *true*



If the hypothesis is *false*

Final Variables Required

To complete an analysis of the effectiveness of the specified policy interventions, we will utilize a data structure that is approximately represented as:

County FIPS Code	Year	State	County Name	Drug Deaths	Drug Shipment Volume	Population Estimate	Death Rate	Ship Rate

- **County FIPS Code**, **Year**, and **State** are the identifiers we will use to enable the analysis and comparison plots. To achieve these keys, we anticipate several merges and transformations from provided or resourced data (described in the next section). The base data used for this generation is the U.S. Vital Statistics dataset.
- **County Name** is preserved from the U.S. Vital Statistics and WAPO FOIA datasets.
- **Drug Deaths** are derived from the U.S. Vital Statistics and will be transformed to meet the structure required for this final set-up.
- **Drug Shipment Volume** is a variable derived from the WAPO FOIA dataset transformed to provide an aggregated volume of drug shipments to a particular county in a specified year.
- **Population Estimate** is going to be constructed from the 2010 U.S. Census Data to enable information about occurrence rates.
- **Death Rate** and **Ship Rate** are the final variables we will produce for the analysis. They are an outcome of Drug Deaths divided by Population and Drug Ship Volume divided by Population for each county.

This data structure will provide the analysis team with incidence rates for Drug Deaths and Drug Shipments into particular counties and states controlled for population size. From this data structure, the team will be able to conduct state-level analysis of Florida, Washington, and Texas by using simple group aggregations and plotting data. A comparison group will be constructed from states selected because of their geographic proximity and approximate size relationship to the experimental group.

Data Sources

1. **Federal Information Processing System (FIPS) Codes for States and Counties**
   a. **Source:**
      - .txt files of Vital Statistics Mortality Summary Data organized by Dr. Nick Eubank [Reorganize “County Code” in the files to 5-digit FIPS codes (with zero padding)]: Dropbox Download Link
      - .txt file from *Federal Communications Commission (FCC)* [Cross Reference/Validation with Vital Statistics Mortality Data]: Download Link
      - .pdf file from *U.S. Census Bureau* (2010 version) [Cross Reference/Validation with Vital Statistics Mortality Data]: Appendix A—State and County Codes and Names
   b. **Summary / Variables:** This dataset will provide us with the unique codes for states and counties which we can use with the other datasets to ensure no dependence on

the names of the states and counties which can be subject to change from one dataset to another.

c. **Merging:** This dataset can be merged with the population data, the prescriptions data and the mortality data using the state and county names.

2. **Population Data**
   a. **Source:**
      - .csv file from *U.S. Census Bureau* (2010 Census Data) [Merge Total Population Count for Each State with other datasets]: Census Bureau Tables
   b. **Summary / Variables:** This dataset provides us with the population of each state and county.
   c. **Merging:** This dataset can be merged with the FIPS dataset first using county and state names, then with the prescriptions and mortality data using the FIPS codes.

3. **Opioid Prescriptions Data (dataset of all prescription opioid drug shipments in the United States from 2006 to 2014)**
   a. **Source:**
      - .tsv.gz file from *Washington Post*: How to download and use the DEA pain pills database - The Washington Post
   b. **Summary / Variables:** This dataset provides us with the weight and number of drugs shipped to each county in the specified time period.
   c. **Merging:** This dataset can be merged with the FIPS dataset first using county and state names, then with the mortality and population data using the FIPS codes.

4. **Vital Statistics Mortality Data (summary of mortality for drug and non-drug related causes for every US county from 2003-2015)**
   a. **Source:**
      - .txt files organized by Dr. Nick Eubank: Dropbox Download Link
   b. **Summary / Variables:** This dataset provides us with the number of drug related deaths in each county in the specified time period.
   c. **Merging:** This dataset can be merged with the FIPS dataset first using county and state names, then with the prescription and population data using the FIPS codes.

Division of Responsibilities

1. **Data Wrangling:** chunking data and extracting required variables
   a. **Assignees:** Dingkun Yang and Andrew Kroening
   b. **Code Review:** Chenying Li and Pooja Kabber

2. **Merging:** merge the above four datasets
   a. **Assignees:** Chenying Li and Pooja Kabber
   b. **Code Review:** Dingkun Yang and Andrew Kroening

3. **Integrity Test:** data sanity test, red teaming
   a. **Assignees:** Dingkun Yang and Andrew Kroening
   b. **Code Review:** Chenying Li and Pooja Kabber

4. **Modeling / Plots:** draw plots for the two types of comparisons

    a. **Assignees:** Chenying Li and Pooja Kabber
    b. **Code Review:** Dingkun Yang and Andrew Kroening

5. **Reports:** two reports, powerpoint presentation, video
    a. **Assignees:** Dingkun Yang, Andrew Kroening, Chenying Li and Pooja Kabber