IDS 720 Final Project: Backwards Design

November 8, 2022

by: Suzy Anil, Sukhpreet Sahota, Yuanjing Zhu, Xianchi Zhang

1 Topic

What problem are you (or your stakeholder) trying to address?

For this project, our primary goal and objective is understanding the effectiveness of policy interventions specifically targeted to limiting illegal quantities of opioids within the United States. This will enable us to analyze the upward mortality trend seen in commonly prescribed opioids, heroin, and other synthetic opioids and answer the problem: do policy interventions regarding opioid drug prescriptions help prevent the over-prescription of opioids and reduce the amount of prescription overdose deaths?

2 Project Question

What specific question are you seeking to answer with this project?

There are three states that implemented policies designed to address the opioid pandemic: Florida, Washington, and Texas. Within Florida, the state legislature began by addressing the opioid problem through registration of pain clinics within the state to ensure proper certification and then working with DEA and enforcement agencies to conduct Operation Pill Nation (raids on uncertified/improperly operated clinics). The state of Texas implemented a very different approach by putting the patient at the forefront through medical evaluations, obtaining consent, periodic reviews of opioid treatments and maintaining proper medical records. Likewise, Washington's Department of Health regulated opioid intake by a patient by placing restrictions on quantities and clearing consultation sessions.

As a result of these policy changes and with the limited set of data, we will be analyzing the impacts to not only the volume/quantities of opioids prescribed but drug overdose deaths/the mortality rate within both Florida and Washington. In particular, we aim to find answers to the following questions:

1. Did the policies placed around opioid drug prescriptions influence the volume of opioids prescribed and drug overdose deaths after being passed in states Florida and Washington?

2. Is there a significant difference in the volume of opioids prescribed and drug overdose deaths between policy-change states (Florida and Washington) and non-policy-change states?

In our comparison, we would like to compare the changes (if any) within Florida to changes in nearby neighboring states, such as Georgia, North Carolina, and Alabama. For Washington, it can be assumed that states within the West/Mountain time zone are more comparable. Therefore, we would like to compare the changes (if any) within Washington to changes in nearby neighboring states, such as Oregon, Idaho, and Colorado.

For the state of Texas, our primary focus is to address questions on the patient's health and the drug overdose deaths/the mortality rate after the state began regulating patient medical evaluations:

1. Is there a significant difference in the amount of opioid overdose deaths between Texas and non-policy-change states?

Similar to Florida and Washington, we would like to compare the changes (if any) within Texas to changes in similar states, such as California, New York, and Pennsylvania. We will attempt to analyze the policy impacts on prescription shipments/volume pending cleansing and merging activities of the data at the month level.

3 Project Hypothesis

What is your hypothesized answer to your question?

As we have crafted questions that will allow us to address the opioid problems within Florida and Washington respectively, it is crucial we outline and benchmark our expectations with appropriate hypotheses. As we are analyzing both the prescription shipments/supply and the mortality rate for Florida and Washington, the following null hypothesis will create a scenario as though policy changes were not implemented/created at all as they were ineffective or didn't achieve their intended purpose after its actual implementation:

- Opioids prescription shipments/supply null hypothesis: The policy changes have zero or negative effect on reducing volume of opioids prescription.
- Drug overdose deaths/mortality rate null hypothesis: The policy changes have zero or negative effect on reducing drug overdose deaths.

On the contrary, if the policies in each of these states proved effective, we would see positive impacts and downward trends in both opioid volumes and mortality rate:

- Opioids prescription shipments/supply alternative hypothesis: The policy changes have a positive effect on reducing volume of opioids prescription.
- Drug overdose deaths/mortality rate alternative hypothesis: The policy changes have a positive effect on reducing drug overdose deaths.

Along the same thought process, we will evaluate the impact of the policy within Texas through an alike null hypothesis and alternative hypothesis (however, as aforementioned, our primary focus will only be on mortality rate with our attempt of opioid shipments pending data cleansing and merging activities):

- Drug overdose deaths/mortality rate null hypothesis: The policy changes have zero or negative effect on reducing drug overdose deaths.
- Drug overdose deaths/mortality rate alternative hypothesis: The policy changes have a positive effect on reducing drug overdose deaths.

4 Model Results

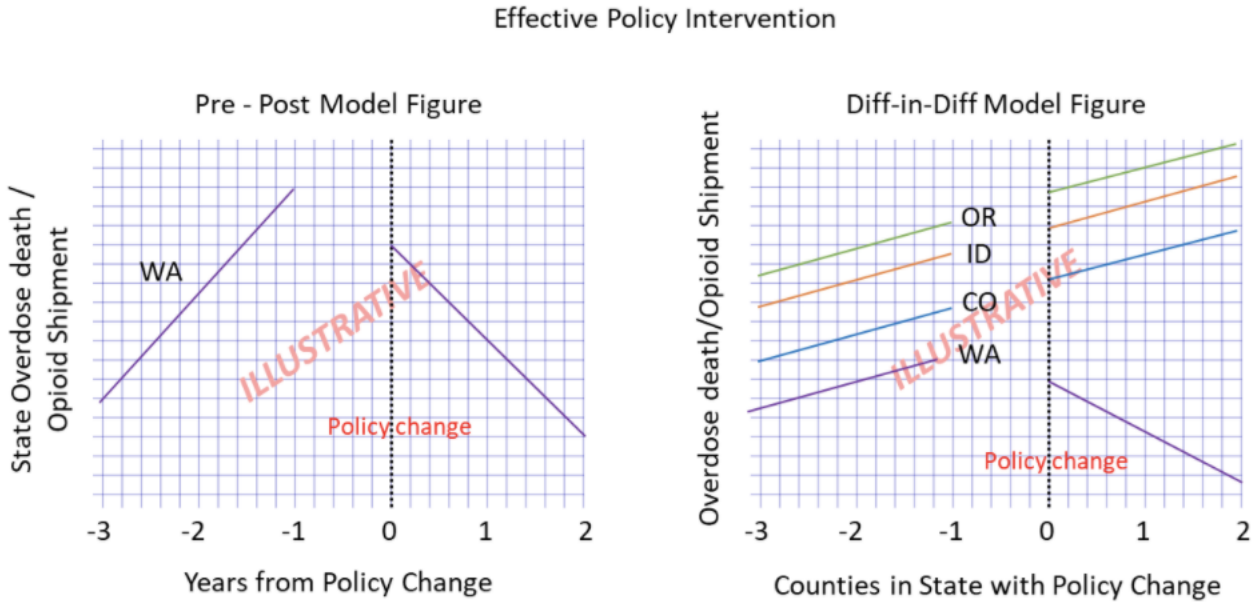
One of the hardest parts of developing a good data science project is developing a question that is actually answerable. Perhaps the best way to figure out if your question is answerable is to see if you can imagine what an answer to your question would look like. Below, draw the graph, regression table, etc. that you would consider to be an answer to your question. Then draw it again, so you have a model result for if your hypothesized answer is true, and a model result for if your hypothesized answer is false. (If the answer to your question is continuous, not discrete (like: what is the level of inequality in the United States), draw it for high values (high inequality) and low values (low inequality)).

If the policies prove to be effective, we will observe a trend that is illustrative below - in which the Pre-Post analysis and Difference-in-Difference analysis will show a downward trend on mortality rate/opioid shipments.

On the contrary, if the policies prove to be ineffective, we will observe a trend that is illustrative below - in which the Pre-Post analysis and Difference-in-Difference analysis will show a continuing upward trend on mortality rate/opioid shipments (or possibly a downward trend that is not attributed to the policy changes and demonstrated nationally/regionally).

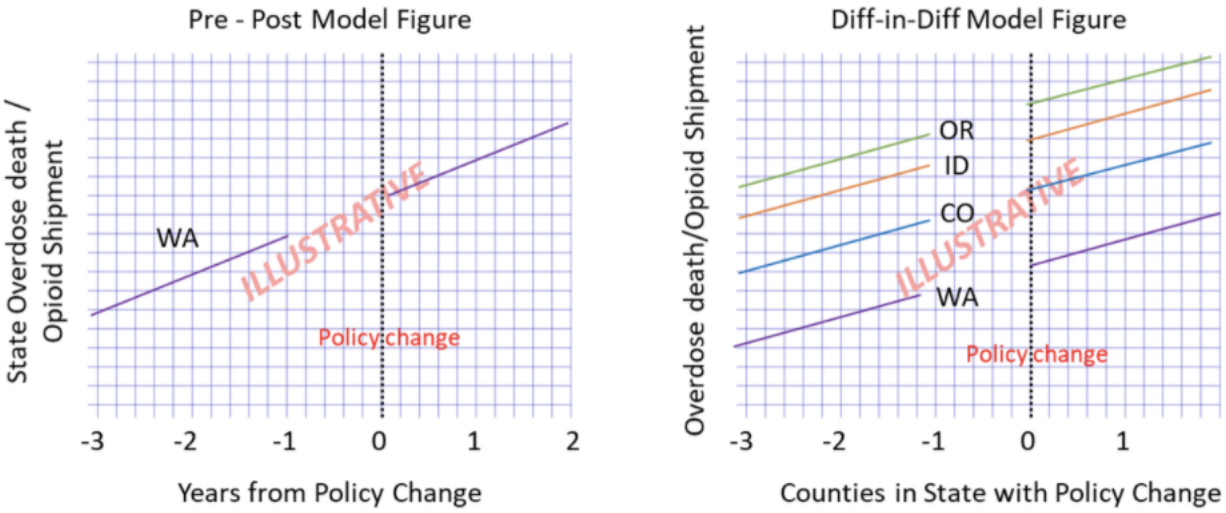
The key for the difference-in-difference model is we will try to look for states that show a similar trend as our states in question. This will be represented through parallel.

Result if your hypothesis is true



Result if your hypothesis is false

Ineffective Policy Intervention



5 Final Variables Required

Now that you've specified what an answer to your question looks like, what data do you need to generate that answer.

For each variable, define both the variable you need ***and*** *the population for which you need the variables to be defined.*

You don't have to be too specific ("I need variable 7 from the NHGIS 2019 census 1% sample release") – just define it in the most general terms that are still specific enough to meet your needs (e.g. I need income data for a nationally representative sample of US citizens).

Based on the questions we designed, we require the following data variables regarding the opioid crisis to generate an answer.

1. All FIPS codes for the states and counties we will use for analysis. These identifiers will serve better as unique primary keys to merge our data sets.
2. A record of deaths marked by whether they were due to an opioid overdose or not ranging from two years before and two years after each policy was placed (2008-2014). It should be separated by counties from both policy (Florida and Washington) and no-policy states.
3. We need data on the volume of opioid prescriptions distributed across the country, separated by state then county during the opioid epidemic.
4. We also require population data for the states we are analyzing separated by county from 2008 – 2014.

6 Data Sources

Finally, given the variables you need for your analysis, what actual data sources do you think will have the data you need?

*In specifying the datasets you need, if you list more than one **also** indicate how you think you can relate these datasets (i.e. if you're gonna merge them, what variables do you think those datasets will provide that will allow you merge them? There's no use saying "I'll merge this political survey with medical records of who has received bad care" if the political survey doesn't provide identifying information you can use to link survey respondents to medical records, even if you have both the survey and medical records!)*

Our data comes from 4 sources:

1. Prescription opioid drug shipments in the United States from 2006 to 2014. It is provided by the Washington Post, which obtained the data through a Freedom of Information Act (FOIA) request to the US Drug Enforcement Agency. In the DEA pain pills database page, we downloaded the state-wide datasets of Florida, Texas, Washington, and their aforementioned comparative states (i.e. OR, ID, CO, etc.).
2. Drug overdose death records from the US Vital Statistics records, which include data on every death in the United States. NOTE: The US Vital Statistics agency censors some data for privacy. If the number of people in a given category (i.e. one county / year / cause of death category) is less than 10, that data does not appear in the data.
3. Monthly buyers' records from 2006 to 2014 from ARCOS API. Here is the GitHub link for it: *https://github.com/wpinvestigative/arcos-api.*
4. Population per county per year from 2003 to 2015, which is retrieved from the Country Population Data by the National Historical Geographic Information System (NHGIS).
5. A dataset/website linking each state and county to its unique FIPS code.

We plan to merge the 4 data sets mentioned above on the FIPS code (with will represent the State and County) and Year, and we'll perform some data manipulation to make sure our datasets have these 3 columns to merge upon. We will also look into the use of FIPS codes to ensure our data merges are accurate/based on an identifier rather than concatenation. Within the Opioid Prescriptions dataset, each year record is stored in one file, so we'll merge these files together and create a year column for each. We plan to do the same with the Monthly buyers' records dataset. For the Vital Statistics Mortality Data, county is recorded as "county, state" (e.g. Autauga County, AL), so we'll split the column by ",", then convert into two columns, one is county and the other is state, and merge it with other datasets.

7 Responsibilities/Division of Work

Outline who will be responsible for writing initial code for each step of the project and who will review each set of code.

For this project, there are 5 critical data activities: 1) Importation of the data, 2) Data merging and cleansing activities, 3) Generation of summary statistics, 4) Pre-Post Analysis, and 5) Difference-in-Difference Analysis.

Upon recognizing these main activities, our team has decided to divide up and delegate the work accordingly:

1. Import data – getting data and adding Dropbox
   (a) Lead: Suzy Anil
   (b) Support: Yuanjing Zhu

2. Organizing of Data / Data Merging / Data Cleansing
    (a) Suzy Anil, Sukhpreet Sahota, Xianchi Zhang, Yuanjing Zhu
3. Summary Statistics
    (a) Lead: Xianchi Zhang
    (b) Support: Suzy Anil
4. Pre-Post Analysis
    (a) Lead: Yuanjing Zhu
    (b) Support: Sukhpreet Sahota
5. Difference-in-Difference Analysis
    (a) Leads: Suzy Anil and Sukhpreet Sahota
    (b) Support: Xianchi Zhang and Yuanjing Zhu

The leads will be responsible for the associated main activity, while the support will provide any additional help as well as conduct PR/code reviews. This will ensure the code is not only reviewed, but can be adjusted accordingly.