

Backwards Design

Beibei Du, Wafiakmal Miftah, Suzanna Thompson, Alisa Tian

November 8, 2022

1 Topic:

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

We are trying to address the problem of rising opioids usage and its connection with death by drugs in the United States.

2 Project Question

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

What effect did the opioid policies in Texas, Washington, and Florida have on the number of opioid deaths and overall drug deaths in those states? In other words, how do opioid deaths and overall drug deaths differ in Texas, Washington, and Florida from our chosen control state, (insert state here)?

3 Project Hypothesis

What is your hypothesized answer to your question?

Our hypothesis is that there is a causal effect between policy change in opioids use and the case of overdose deaths in Florida, Washington, and Texas State. We also hypothesize that while opioid prescriptions decrease, drug-related deaths on the whole increase or remain stable. We suggest this result as there is evidence to suggest that removing a drug from an addict does not necessarily imply that the addict will recover. In some cases, the addict will switch to a new drug of chose. In other cases, the opioid addict will be in recovery and later overdose once they're able to find an opioid. This is a well known phenomena as addicts may forget that their tolerance for the opioid decreases during times of recovery and then try to use the same opioid dose as they were using active addiction. In turn, the addict will accidentally overdose.

We also suppose that opioid prescriptions decrease in Texas, Washington, and Florida.

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)).

Result if your hypothesis is true

Three pre-post graphs for an effective policy, showing a rapid increase line in the pre-policy era for drug deaths and a slow increase line in the post-policy era. Three pre-post graphs for prescription rates for an effective policy, showing an increasing line in the pre-policy era and a decreasing line for post-policy era. Three difference-in-difference plots showing a slowing increase in prescription of opioids and stable or increasing rate of death in Florida, Texas, and Washington, while strictly increasing in other states.

Result if your hypothesis is false

Three pre-post graphs for an effective policy, showing an increasing line for drug deaths in the pre-policy era and an decreasing line in the post-policy era. Three pre-post graphs for prescription rates where in the pre-policy era rates are increasing, and in the post-policy era, rates continue to increase. Three difference-in-difference plots showing a stable or decrease rate of death in Florida, Texas, and Washington, while a comparable increase in other states. Three difference-in-difference plots showing an increase of prescription rate for Florida, Texas, and Washington.

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).

We select the state, the volume of opioids prescribed, population, prescription rate, drug overdose deaths, year, month, period, median household income, and the variable indicating policy change (binary) as final variables. We need the drug overdose deaths data (per 100,000 population) for a sample of US population in different states. We also need the prescription rate and the median household income data based on samples of the US population by states.

6 Data Sources

The link to the data resources is here: [Data Resources here](#)

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!)*

With the education and income data that we find, we will merge variables on the 'state' and 'year'. We are including income(median household income) and education level in order to control for confounding variables per state. If a well educated and affluent state always has a lower death rate and prescription rate compared to a less educated and poorer state, then it is possible that the rates are being influenced by the aforementioned factors instead of the policies that we are evaluating.