# Opioids: Project Plan/Backwards Design/Strategy

Graded

**Group**

Si Min Loo
Chris Moreira
Shiyue Zhou
...and 1 more
View or edit group

**Total Points**

**1 / 1 pts**

**Question 1**

## Backwards Design

**1** / 1 pt

✔ **+ 1 pt** Stellar! Well thought through.

✔ **+ 0 pts** As discussed in class, deciding how many units to include as controls is a balancing act — more controls means more statistical power, but to get more controls you have to include controls that are of decreasing similarity. So you have to justify your choice on that topic.

However, I *will* say you should work with *at least* three control states per state being studied.

1

7

✔ **+ 0 pts** The idea here is to specify the variables that would be an input to your analysis, not to speak in general terms about data. Can you make the plots above with "number of opioids prescribed per county per year"? or "opioid volume"? (no).

Many people have answers where, reading between the lines, I can see they probably get it, but I think an ideal response here would be "opioid shipments per capita in [units]". That kind of specificity helps readers notice precisely what they need to compute that.

Don't straight into potential sources and datasets, or types of data before doing the really granular, concrete part.

6 You're getting there, don't get me wrong.

Don't need TX

2 Is FIPS in all of these?

3 Oops — did I leave TX in the writeup somewhere? You only have to do WA and FL.

4 Control states?

5 can you?

No questions assigned to the following page.

# Backwards Design

Duke University
PDS 720 - Data Queens and King Group
Chris Moreira, Ailina Aniwan, Loo Si Min (Lucy), Shiyue (Cynthia) Zhou

## 1 Topic

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

**Problem Statement:**
This project focuses on evaluating the effectiveness of opioid control policies implemented to reduce opioid abuse in Washington, Florida, and Te[illegible]. By examining changes in opioid prescription rates and overdose mortality following each state's policy implementation, we aim to understand the broader impact of these interventions. Using pre-policy data as a baseline, we will compare trends in each state's opioid-related metrics to those in states without similar policies to ensure comparable pre-policy conditions. Our analysis employs both pre-post and difference-in-difference methods to assess the effectiveness of each intervention, providing insights into their influence on prescription volumes and overdose rates.

## 2 Project Question

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

**Primary Question:**
What are the changes in opioid prescription rates (Y1) and overdose mortality (Y2) following the implementation of opioid policies?

**Sub-questions:**
1. How does the policy in each state individually impact opioid prescriptions and overdose mortality over time?
2. Are the observed changes in policy-implementing states significantly different from changes in comparison states?

## 3 Project Hypothesis

*What is your hypothesized answer to your question?*

**Main Statement:**
Implementing stricter opioid prescription control regulations will reduce both the volume of opioid prescriptions and the rate of overdose deaths.

**Null Hypotheses (H0)**
**Pre-Post Analysis (PP)**: There are no significant changes in opioid prescription rates ($Y_1$) and overdose mortality following the policy implementation within the same state.

No questions assigned to the following page.

**Difference-in-Differences (DiD)**: There are no significant differences in changes to opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) when comparing states with implemented policies to states without such policies.

**Alternative Hypotheses (H1)**
**Pre-Post Analysis (PP)**: There are significant decreases in opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) within each state following policy implementation, with lower rates observed in the three years after compared to the two years before.
**Difference-in-Differences (DiD)**: There are greater reductions in opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) in states with implemented policies compared to states without such policies over the same period.

# 4 Model Results

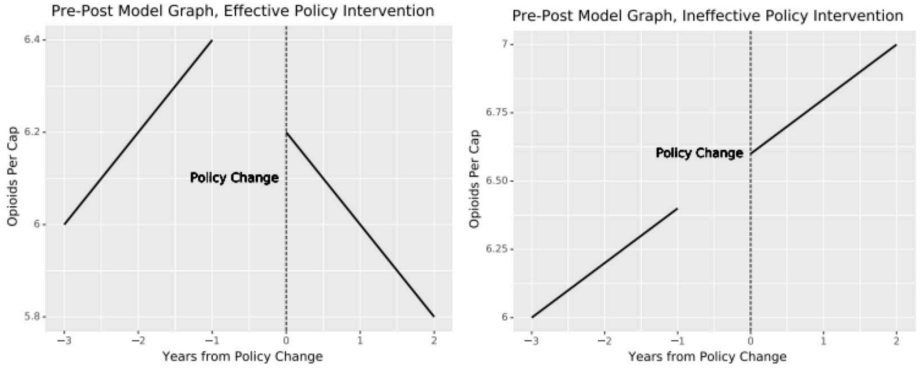
*One of the hardest parts of developing a good data science project is developing a question that is actually answerable. Perhaps the best way to figure out if your question is answerable is to see if you can imagine what an answer to your question would look like. Below, draw the graph, regression table, 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 (e.g. "what is the level of inequality in the United States?"), draw it for high values (high inequality) and low values (low inequality)).*

**Pre-Post Comparison**: Displaying trends in opioid prescriptions and overdose rates pre-policy and post-policy.

(1) Compare Texas Drug situation (overdose deaths) before Dec 2004 - Dec 2006 and Jan 2007 - Jan 2010
(2) Compare Florida Drug situation (opioid shipments, overdose deaths) before Dec 2007 - Dec 2009 and Jan 2010 - Jan 2013
(3) Compare Washington Drug situation (opioid shipments, overdose deaths) before Dec 2009 - Dec 2011 and Jan 2012 - Jan 2015

**Sample Plots for Pre-Post**

Figure 2: Pre-Post Model Figures



No questions assigned to the following page.

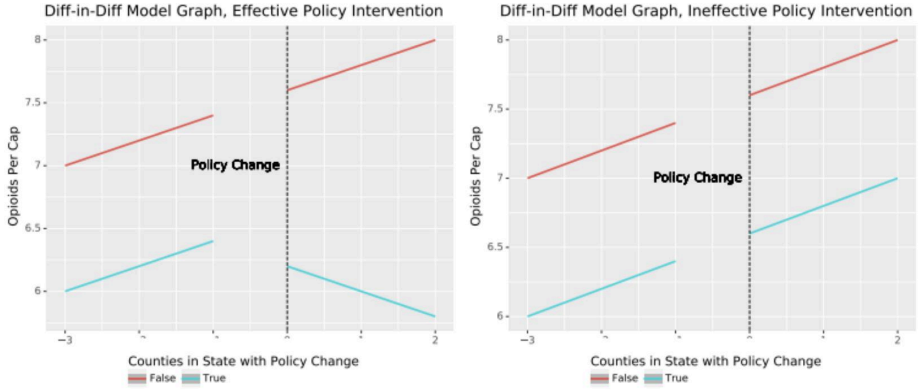
**Difference-in-Difference (DiD) Analysis**: The Difference-in-Difference method will be used to compare changes in opioid-related metrics between states with opioid policy interventions (Texas, Florida, Washington) and states without such interventions (Oklahoma, Oregon, Georgia) over the same time period (January 2012 to January 2015).

**States:**

- Texas vs. Oklahoma
- Washington vs. Oregon
- Florida vs. Georgia

**Sample Plots for DiD**

Figure 3: Difference-in-Difference Example Plots




**Results (if Hypothesis is True)**

**Pre-Post Analysis (PP):**

- There is a noticeable decrease in opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) within the same state following the implementation of opioid control policies.
- Specifically, the data will show that opioid prescription rates and overdose mortality are lower in the three years post-policy implementation compared to the two years preceding it.

**Difference-in-Differences (DiD):**

- There are significant differences in the changes to opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) between states with and without opioid control policies.
- States with implemented policies will demonstrate a greater reduction in both opioid prescription rates and overdose mortality compared to states without these policies over the same period.

**Result (if your hypothesis is false)**

**Pre-Post Analysis (PP):**

- There are minimal or no changes in opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) within the same state following the policy implementation.

No questions assigned to the following page.

- Opioid prescription rates and overdose mortality do not show a significant decrease in the three years after the policy compared to the two years prior.

**Difference-in-Differences (DiD):**

- There is no significant difference in the changes to opioid prescription rates ($Y_1$) and overdose mortality ($Y_2$) between states with and without implemented policies.
- States with policies do not exhibit a notably greater decrease in opioid prescription rates or overdose mortality compared to states without such policies, suggesting that the policy had limited or no impact.

# 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?*
*You don't have to know the exact variable and dataset ("I need ANNPOP2008 from the NHGIS 2019 census 1% sample release"), but you should be specific enough that all properties that are critical to your analysis are fully specified (e.g., I need individual-level annual income data for a nationally representative sample of all working US citizens).*
*Basically, this should be a list of all the variables (not in terms of specific variable names, but the substantive content of each variable) you want in your final dataset and an explanation of the population represented in that data (i.e., what is represented by each row of your data, and what is the group of entities included in the data).*

**Data Elements Needed:**

- **Opioid Prescription Data:** Including volume, shipment dates, and drug identification (such as REPORTER_STATE, REPORTER_COUNTY, BUYER_STATE, BUYER_COUNTY, DRUG_CODE, and TRANSACTION_DATE).
- **Overdose Mortality Rates:** Annual opioid-related overdose deaths by county.
- **Demographic Data:** Annual county-level data on population, education, income, or poverty rates as control variables for socioeconomic context.

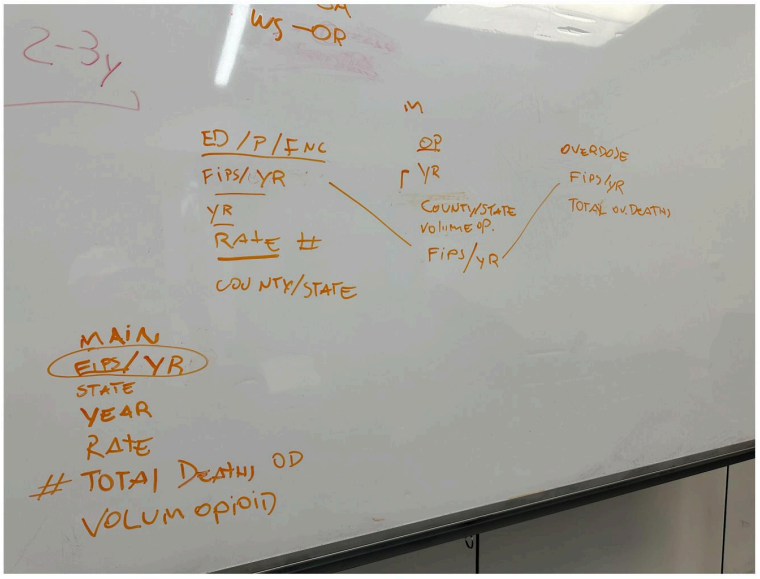
**Population Representation:**
Each row of the final dataset represents a unique County-Year observation, covering relevant U.S. counties from 2004 to 2015 for Texas, Florida, and Washington, as well as comparable non-policy states.

| Variable | Description |
|---|---|
| **FIPS** | Unique ID for each state |
| **Opioid Prescriptions** REPORTER_STATE, REPORTER_COUNTY, BUYER_STATE, and BUYER_COUNTY | Identifies the location of facilities and recipients, essential for assessing policy effects by state or count |

No questions assigned to the following page.

| | |
|---|---|
| **Opioid Prescriptions** DRUG_CODE, NDC_NO, DRUG_NAME | Drug identification codes and names (e.g., oxycodone, hydrocodone), crucial for categorizing opioid types |
| **Opioid Prescriptions** TRANSACTION_DATE | The exact date of shipments, allowing for time-series analysis to track changes before and after policy implementation |
| **Vital Statistics Mortality Data** | Mortality (overdose deaths) |
| **USDA_Education.name (county_name), USDA_Education [2008-2012]** | County-level data on education metrics for socioeconomic analysis over the specified period |

* Dropping state of Alaska
*Geographic unit of analysis: Counties
*Temporal unit of analysis: Years




# 6 Data Sources

*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 datasets, 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*

No questions assigned to the following page.

*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!)*
*If you are familiar with Entity Relational Diagrams for databases, something similar is a great way of diagramming how you expect datasets you plan to use will relate to one another.*

**List of Datasets Needed for Analysis:**

1. **Opioid Prescriptions**: all prescription opioid drug shipments in the United States from 2006 to 2019. https://gfx-data.news-engineering.aws.wapo.pub/ne-static/arcos/v2/bulk/arcos_all.zip
2. **Vital Statistics Mortality Data**: Summary of mortality for drug and non-drug-related causes for every US county from 2003-2015. https://www.dropbox.com/s/kad4dwebr88l3ud/US_VitalStatistics.zip?dl=0
3. **Demographic Data** (Education or Unemployment or Population or Poverty): County-level education, income, and poverty data from the USDA, which will be linked using county FIPS codes to provide socioeconomic control variables. https://www.ers.usda.gov/data-products/county-level-data-sets/

**Data Merge Strategy:**
We will merge datasets by FIPS codes (county identifiers) and year to create a cohesive dataset, ensuring that each row represents a unique County-Year observation.

# 7 Division of Labor

*Now that you have identified what needs to be done, how do you plan to divide those tasks among team members?*

**Data Collection & Cleaning**

- Opioid Prescription Data: Lucy
- Overdose Mortality Data: Ailina
- Demographic Data (Education/Income/Poverty): Cynthia
- FIPS County Code Data: Chris

**Data Merging and Structuring**

- Chris will handle the initial merging of datasets using FIPS and ensure consistency across data sources.

**Analysis & Visualization:**

- Pre-Post Analysis: Chris and Cynthia will analyze and visualize pre- and post-policy trends for each state.
- Difference-in-Differences (DiD) Analysis: Ailina and Lucy will conduct and visualize DiD analyses, comparing policy vs. non-policy states.

**Documentation & Reporting**

- All team members will contribute to report writing, with sections divided to cover methodology, analysis, and visualizations.