# **Project Strategy Outline**

Sam Sloate, Caleb O’Neel, Tommy Tseng

## **Question and Outputs:**

The main question we seek to answer is: “What is the effect of opioid policy changes on opioid prescriptions and mortality rates?”

A finished product will contain four pre-post analysis graphs:

* + Florida in February 2010 (on opioid shipments and overdose deaths)
  + Texas in January 2007 (on overdose deaths)
  + Washington in January 2012 (on overdose deaths)

As well as four difference-in-difference (DID) analysis graphs:

* + Florida in February 2010 (on opioid shipments and overdose deaths)
  + Texas in January 2007 (on overdose deaths)
  + Washington in January 2012 (on overdose deaths)

If the pre-post analysis graphs indicate changing trends before and after the policies were implemented, a difference-in-difference analysis graph can show whether these trends were a result of the policies themselves or other factors. Each state in the DID graphs will be compared to at least three other baseline states that did not implement opioid policies but that may share similar external situations.

Our hypothesis is that drug shipments and drug overdoses will decrease in Florida, Texas, and Washington after the policies are implemented. If this is true, the pre/post analysis will show a sharp drop in the overdose rates after the years the policies were put in place, and the DID analysis will show the decreases in FL, TX, and WA while other states continue their trends with no change.

## **Final Datasets**

To complete the above analysis, we need two final datasets: One on Opioid Shipments and one on Opioid Mortalities.

Within the Opioid Shipment data, we need a column with years 2006-2012 (which will be on the X-axis of our graph) and a column with a standardized amount[[1]](#footnote-1) of opioids[[2]](#footnote-2) per capita going to each state (for the Y-axis). We will need to think about which states to use as comparisons, likely states near Florida that may be experiencing the same geographic shocks or conditions. We will also examine state and national trends for all other states before Florida’s 2010 policy was implemented to see if any other states mirror Florida’s trends well. The dataset will need to be split by pre- and post-2010.

Within the Opioid Mortality data, we will need a column with years 2003-2015 (X-axis) and a column with the number of per capita overdoses due to opioids in each state (Y-axis). Again, we will need to select useful regions to compare against for the DID analysis, starting with geographically close states and then also examining other states with similar trends. The dataset for each state and their group of comparison states will need to be split before and after the policy year in each state (2010 for Florida, 2007 for Texas, and 2012 for Washington).

## **Intermediate Datasets**

To get to the final Opioid Shipment data listed above, we will need a dataset that has a column with standardized opioid shipments for all counties in each state, for all years 2006-2012. There also needs to be a column with the population of each county per year, in order to calculate per capita shipments. We will likely standardize by the morphine milligram equivalent per 1,000 people[[3]](#footnote-3). We also need to set the index as or have a column with FIPS codes, to avoid issues with duplicate county names.

To get to the final Opioid Mortality data listed above, we will need a dataset that has a column with opioid overdoses for all counties in each state, for all years 2003-2015. There also needs to be a column with the population of each county per year, in order to calculate per capita deaths. We will likely standardize by overdoses per 1,000 people. We also need to set the index as or have a column with FIPS codes, to avoid issues with duplicate county names.

## **Beginning Datasets**

We have two beginning datasets: Opioid Shipments and Opioid Overdoses (several files each with one year). We need to merge all these files together. We also need to find two datasets to merge in: FIPS counties and a dataset with populations per county per year, likely from the U.S. Census Bureau.

To clean the Opioid Overdoses data, we need to drop all observations that are not drug-induced deaths (keep Unintentional, Suicide, and Undetermined causes).

To clean the Opioid Shipment data, we only need to keep standardized amounts based on morphine milligram equivalents (MME Conversion\_Factor).

## **Task Assignment**

1. Put Source Data on Dropbox: Tommy
2. Cleaning up mortality dataset: Caleb
   1. Need to merge all files together
   2. End product: Dataframe (DF) with useful columns as listed above
3. Cleaning up prescription dataset: Tommy
   1. End product: DF with useful columns as listed above
4. Finding and merging population/FIPS datasets: Sam
   1. End product 1: Opioid Prescription DF with a column for population per county per year and one for calculating the percent of deaths per capita
   2. End product 2: Opioid Shipment DF with a column for population per county per year and one for calculating the amount of active ingredient per capita
5. Graphing Pre-Post Data: meet in person; all
6. Graphing DID: meet in person; all
7. Reports
   1. Outline: Sam
   2. Data description sections: Caleb and Tommy
   3. Analysis sections: All

1. We will need to determine a way of standardizing the amount of active ingredients going to each state. Current plan: Using the MME Conversion\_Factor variable in the Opioid Prescription dataset from the Washington Post. [↑](#footnote-ref-1)
2. The data set looks like it has three kinds of opioids. [↑](#footnote-ref-2)
3. As of now, we plan on keeping buyer\_county, buyer\_state, drug\_name, drug\_code, transaction\_date, MME Conversion\_Factor, and transaction\_id, as well as dosage\_unit and calc\_base\_wt\_in\_gm and Ingredient\_Name if relevant. [↑](#footnote-ref-3)