

1. Introduction

Background

Opioids are a drug class that are primarily used for pain relief, but are highly addictive and can result in fatal overdose. Opioids are “controlled substances”, meaning that they are regulated by the government. However, due to addiction and illicit use coupled with malfeasance from drug manufacturers, distributors, and doctors there have been 399,212 opioid related deaths in the US between the years of 1999 and 2017 [1]. These deaths are related to the overprescription of medication and are arguably for the sake of increased profits of pharmaceutical companies, at the cost of human life. This has to a declaration of an opioid “epidemic” and the implementation of public policies aimed at reducing the amount of opioid-related deaths by controlling over-prescription.

The motivation for the project

A database maintained by the Drug Enforcement Administration that tracks pain pills was recently made available as a result of a court order from the Washington Post and HD Media, who publishes the Charleston Gazette-Mail in West Virginia.

From the Automation of Reports and Consolidated Orders System (ARCOS), all opioid pills that were distributed from 2006 to 2012 were recorded. Using this data in comparison with other metrics, causal analysis will be performed to evaluate the effectiveness of opioid policies implemented in three states: Florida (implemented in 2010), Texas (2007), Washington State (2012).

2. Data

Data Handling

Three datasets were combined in preparation for analysis: the ARCOS drug tracking database, mortality data to identify causes of death, and population data from the U.S. Census Bureau. These datasets were cleaned and merged together, with the intermediate goal for each dataset to be merged onto a schema crosswalk to provide a uniform unique identifier. In the case for these three datasets, each county in the US has a FIPS (Federal Information Processing Standards) Code [3], which was merged onto each table prior to all three tables being merged together using their FIPS code.

(a) ARCOS Dataset

This data was downloaded by State from the Washington Post. Each file was reduced down to the following columns:

| State | County | Transaction Date | Calc Base Weight in Grams | MME Conversion Factor |
|-------|--------|------------------|---------------------------|-----------------------|
|-------|--------|------------------|---------------------------|-----------------------|

“Calc Base Weight in Gram” is a value that is generated by ARCOS database that is a value that identifies the final gram conversion of the controlled drug substance, taking into effect of Quantity, Dosage Unit and Dosage Strength information. Each opioid was scaled to a Morphine Milligram Equivalent using its “MME Conversion Factor”. A new value, “MME Strength” was calculated by multiplying “Calc Base Weight in Gram” and “MME Conversion Factor” [2].

After calculating MME Conversion Factor, the Transaction Date was converted to the transaction year, and finally a groupby function was applied on State, County, and Year to have an output table “State”, “County”, “Year”, and “MME Strength”.

These files were concatenated together and then merged with the FIPS table on the “State” and “County” columns for later analysis. The merge was validated using an outer join and 1:1 matching. The final ARCOS Table was:

| FIPS | State | County | Year | MME Strength |
|------|-------|--------|------|--------------|
|------|-------|--------|------|--------------|

(b) Mortality

Mortality data was gathered from 2003 to 2015. The data came in 13 files, one for each year and was merged into one dataframe. The important columns in the source file is shown in the table below:

| County | State | Year | Drug/Alcohol Induced Cause | Deaths |
|--------|-------|------|----------------------------|--------|
|--------|-------|------|----------------------------|--------|

All non-drug related deaths were filtered out of the dataframe because this study was interested in only deaths due to drug-related causes. A group-by operation was performed

to aggregate the data by “County”, “State” , and “Year” to generate the total number of deaths due to all drug-related causes. This table was then merged onto the related FIPS code. We performed a validation test by performing checking that the number of rows of the new table after joining the aggregated mortality table with the FIPS codes would be the same as the number of rows we had in the aggregated mortality table.

(c) Population

The population data was used to generate per capita data. The two original data frames contained population data on a county level from 2000 to 2010 and 2010 to 2018. The important columns in the source file is shown in the table below:

| County Code | State Code | State Name | County Name | Census 2000/2010 | Pop Estimate 2001/2011 | ... |
|-------------|------------|------------|-------------|------------------|------------------------|-----|
|-------------|------------|------------|-------------|------------------|------------------------|-----|

The U.S. Census Bureau conducts the census every 10 years. The last census was conducted in 2010. Based on the 2000 Census and 2010 Census, the Census Bureau estimates the population data from 2001 to 2009 and 2011 to 2018. Thus we used the official census and estimations as population data [4]. Unrelated columns for years not in the 2003 to 2015 time frame were removed. The FIPS code was included in the Census data.

To make merging and analysis process easier for other datasets, we rearranged, renamed some columns and melted the data according to “FIPS”, “County Name” and “State Name”. This allowed us to reshape the data frame by pivoting objects and make an index of years identify individual observations

3. Methods and Metrics

Pre-post Comparison

Pre-post comparison aims to achieve summary statistics for two different time periods. Since we want to see how the opioid policy affected the prescription and overdose death, pre-post analysis shows us a general intuitive change in state. Take Florida as an example: if the prescription and overdose death decreases after applying the policy in

2010, this opioid policy is more likely to be determined successful. If the trend of prescription and overdose death remain unchanged, the policy is possibly ineffective.

We completed the pre-post comparison analysis for overdose deaths were performed for Washington and Texas, and pre-post comparisons for both overdose deaths and for prescriptions were performed for Florida. Analysis for prescriptions were not performed for Washington and Texas due to the lack of opioid shipment data. The analysis aims to directly show the changes of overdose deaths and prescription in the pre-post graphs.

Difference-in-difference

Samples of different groups may have prior differences before the implementation of the policy. A single pre-post comparison may ignore these differences, which lead to a biased estimation of the effect of the policy implementation. The DID model is based on data obtained from natural experiments. It effectively controls the pre-existing differences between the research objects through modeling, and effectively separates the true results of policy influences. Still take Florida as an example, we assume there would be a decrease in prescription and overdose deaths in Florida after policy change. We also want to estimate if there also exist large difference after that period of policy change in other states, which have a similar trend before but with no policy change.

In this project, we chose the whole USA as control group to complete the DID analysis, which indicates the average level and has almost parallel trends before policy year. Also, we completed the DID analysis for overdose deaths in Washington, Texas and Florida, and only for prescription in Florida, based on both graphs and regressions.

The assumptions of DID model:

- 1) The releasing of policy was not determined by overdose deaths and prescription.
- 2) The treatment groups (Florida, Texas and Washington) and control group (the USA) have parallel trends in outcome (have similar trend in overdose deaths and prescription before policy year).
- 3) The composition of both treatment groups and control group is stable.
- 4) The policy changes for all the three states have no spillover effects (the policy would only produce the expected effect, no other effects on the people or society outside the states).

The DID model only allowing for level changes:

$$Y_{c,t} = \alpha + \varphi_c + \beta_1 post_t + \beta_2 post_t policy_state_c + \varepsilon_{c,t}$$

$Y_{c,t}$: overdose deaths or opioid shipment per capita

φ_c : county fixed effect

$post_t$: an indicator for whether we are in a period of after implementation of the policy change

$policy_state_c$: an indicator for whether a given county is in a state that experienced a policy change

In this regression model ,we would like to interpret the coefficient β_2 to show the effect of policy change. The coefficient β_2 includes two levels of difference. This first one is the difference of whether the group is the state of a policy implementation, which could be calculated by $\beta_2 - \beta_1$ mathematically. The second one is the difference of whether the group is after the year of policy change, which could be indicated by $\beta_1 - 0$. If we calculate these two differences together, it will give us a ‘difference-in-difference’ estimate including both of them, which could be calculated by $\beta_2 - \beta_1 + \beta_1 - 0 = \beta_2$.

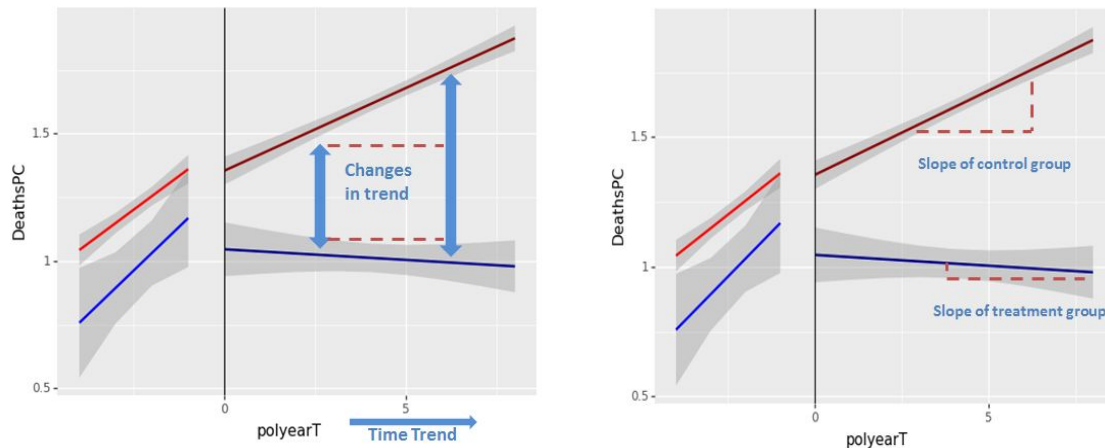
The DID model allowing for both changes in levels and for changes in trends:

$$Y_{c,t} = \alpha + \varphi_c + \beta_1 year_t + \beta_2 post_t + \beta_3 post_t year_t + \beta_4 post_t policy_state_c + \beta_5 post_t year_t policy_state_c + \varepsilon_{c,t}$$

$year_t$: adjust year, which has a value of 0 in the year the policy goes into effect

In this regression model ,we would like to interpret $\beta_4 - \beta_2$ and $\beta_6 - \beta_3$. Similar as the regression model before, $\beta_4 - \beta_2$ just shows us the difference of whether the group is the state of a policy implementation. $\beta_6 - \beta_3$ tells us the changes of outcome after the policy year through the time trend.

For example:

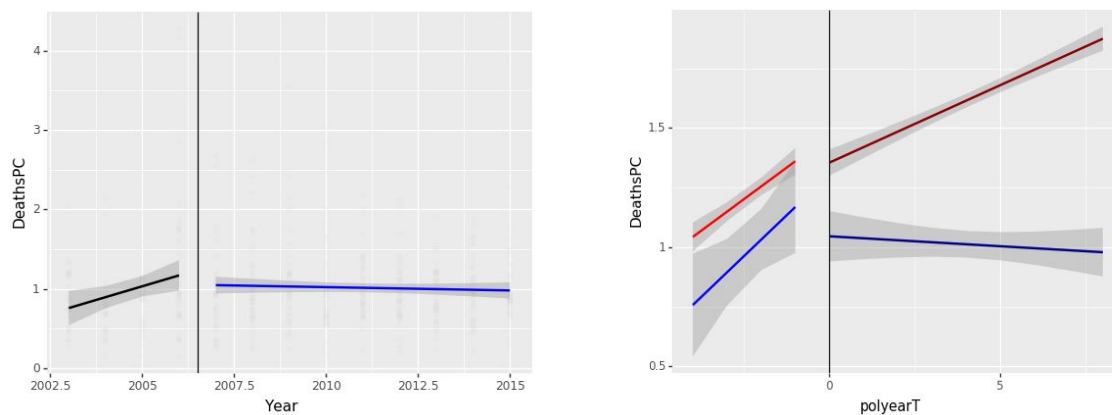


Firstly, $\beta_6 - \beta_3$ can give us the difference after the policy year, where the dummy variable 'post' equals to one instead of zero. What's more, the interaction post*year*state and the interaction post* year could provide the difference between control group and treatment group. As shown in the pictures above, these differences could be larger or smaller through the time trend. We calculate $\beta_6 - \beta_3$ to get the slope difference between the groups, which could be used to return changes after years.

4. Findings

Texas:

Pre-post Comparison and DID graphs for overdose deaths



In Texas we see that there is a change. The slope changes after the policy (indicated by the black line). In the Difference in Difference graph we can see that the US has a clear upward trend while Texas shows a downward trend after the policy.

This means we can interpret these results as evidence for the policy having a causal effect in the decrease of these drug-related deaths in Texas.

DID Regression Analysis for overdose deaths

```

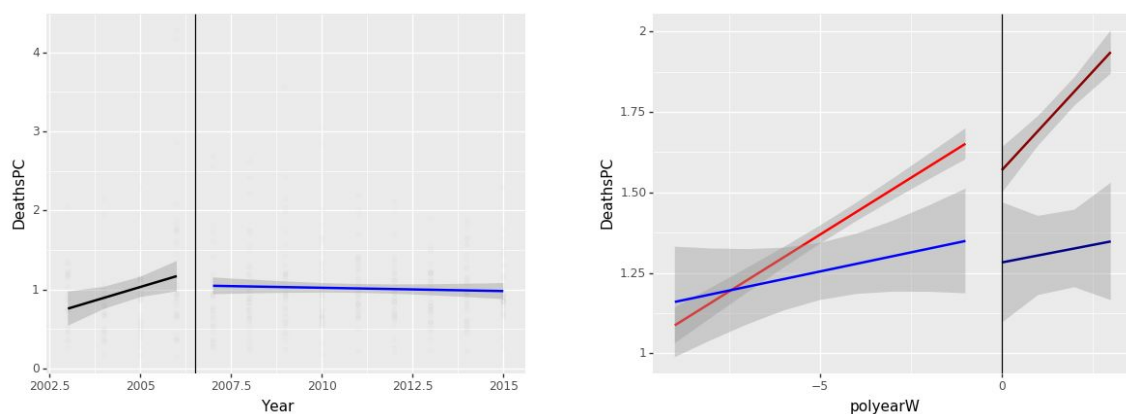
=====
                        OLS Regression Results
=====
Dep. Variable:          DeathsPC      R-squared:                0.004
Model:                  OLS           Adj. R-squared:           0.003
Method:                 Least Squares  F-statistic:             14.29
Date:                   Sun, 10 Nov 2019  Prob (F-statistic):       6.40e-07
Time:                   15:19:18        Log-Likelihood:          -11838.
No. Observations:       7917           AIC:                    2.368e+04
Df Residuals:           7914           BIC:                    2.370e+04
Df Model:                2
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept              1.5137      0.025     60.731     0.000      1.465      1.563
Policy_T               0.0087      0.029      0.302     0.763     -0.048      0.065
Policy_T:State_T       0.3528      0.067      5.276     0.000      0.222      0.484
=====
Omnibus:               5512.832    Durbin-Watson:           0.672
Prob(Omnibus):          0.000    Jarque-Bera (JB):        104455.185
Skew:                   3.129    Prob(JB):                0.00
Kurtosis:              19.658    Cond. No.                7.10
=====

```

| OLS Regression Results | | | | | | |
|-----------------------------|------------------|---------------------|------------|-------|--------|--------|
| ===== | | | | | | |
| Dep. Variable: | DeathsPC | R-squared: | 0.008 | | | |
| Model: | OLS | Adj. R-squared: | 0.007 | | | |
| Method: | Least Squares | F-statistic: | 10.02 | | | |
| Date: | Sun, 10 Nov 2019 | Prob (F-statistic): | 4.70e-11 | | | |
| Time: | 15:21:31 | Log-Likelihood: | -11823. | | | |
| No. Observations: | 7917 | AIC: | 2.366e+04 | | | |
| Df Residuals: | 7910 | BIC: | 2.371e+04 | | | |
| Df Model: | 6 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| Intercept | 1.5175 | 0.060 | 25.475 | 0.000 | 1.401 | 1.634 |
| adj_year_T | -0.0048 | 0.023 | -0.214 | 0.831 | -0.049 | 0.039 |
| Policy_T | 0.0703 | 0.065 | 1.074 | 0.283 | -0.058 | 0.199 |
| Policy_T:State_T | 0.6792 | 0.122 | 5.551 | 0.000 | 0.439 | 0.919 |
| Policy_T:adj_year_T | -0.0108 | 0.023 | -0.467 | 0.640 | -0.056 | 0.035 |
| State_T:adj_year_T | 0.1252 | 0.044 | 2.828 | 0.005 | 0.038 | 0.212 |
| Policy_T:State_T:adj_year_T | -0.2076 | 0.051 | -4.060 | 0.000 | -0.308 | -0.107 |
| ===== | | | | | | |
| Omnibus: | 5509.892 | Durbin-Watson: | 0.674 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 104728.294 | | | |
| Skew: | 3.125 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 19.686 | Cond. No. | 64.1 | | | |
| ===== | | | | | | |

Washington:

Pre-post Comparison and DID graphs for overdose deaths



Since we have little proof for the policy effect according to the DID graph, the difference in level change could not be explained by regression.

Before the policy, Washington shows different behavior than the rest of the country. Due to the nature of scale in the graph we see a little shift in the two graphs but this is not as important. With the confidence or error margins we can see that there is a good chance that there is no difference between the Washington trend before and after the policy. The fact that we can draw a line that can go straight through both intervals indicates that there is a reasonable chance that there is no difference from the policy.

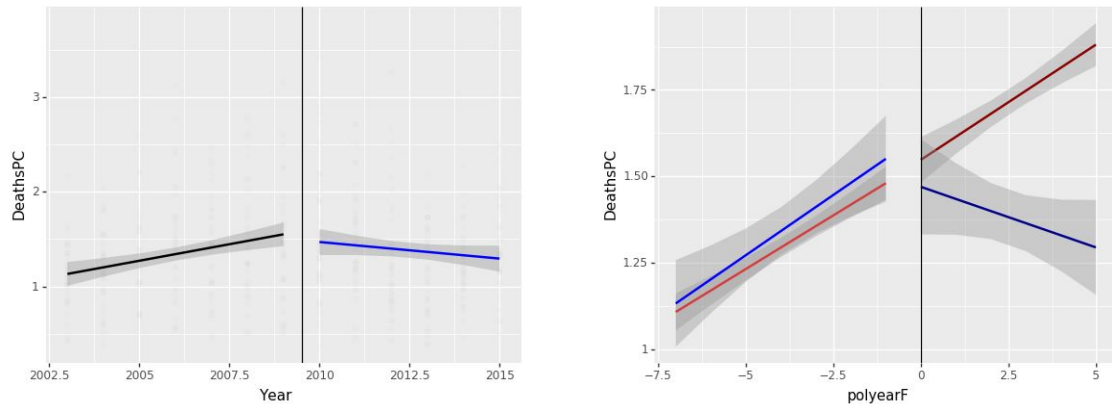
```

=====
                        OLS Regression Results
=====
Dep. Variable:          DeathsPC      R-squared:                0.003
Model:                  OLS           Adj. R-squared:           0.002
Method:                 Least Squares  F-statistic:             4.241
Date:                  Sun, 10 Nov 2019  Prob (F-statistic):      0.000287
Time:                  15:25:12        Log-Likelihood:          -11840.
No. Observations:      7917           AIC:                    2.369e+04
Df Residuals:          7910           BIC:                    2.374e+04
Df Model:              6
Covariance Type:       nonrobust
=====
                        coef      std err      t      P>|t|      [0.025      0.975]
-----
Intercept              1.5951      0.031     51.041    0.000      1.534      1.656
adj_year_W             0.0069      0.006      1.157    0.247     -0.005      0.019
Policy_W              -0.1021      0.047     -2.167    0.030     -0.194     -0.010
Policy_W:State_W       -0.8421      0.343     -2.453    0.014     -1.515     -0.169
Policy_W:adj_year_W    -0.0094      0.019     -0.486    0.627     -0.047      0.029
State_W:adj_year_W      0.0680      0.024      2.792    0.005      0.020      0.116
Policy_W:State_W:adj_year_W  0.1776      0.185      0.960    0.337     -0.185      0.540
=====
Omnibus:              5514.488    Durbin-Watson:           0.671
Prob(Omnibus):         0.000    Jarque-Bera (JB):       103743.276
Skew:                  3.133    Prob(JB):                0.00
Kurtosis:              19.590    Cond. No.                136.
=====

```

Florida:

Pre-post Comparison and DID graphs for overdose deaths



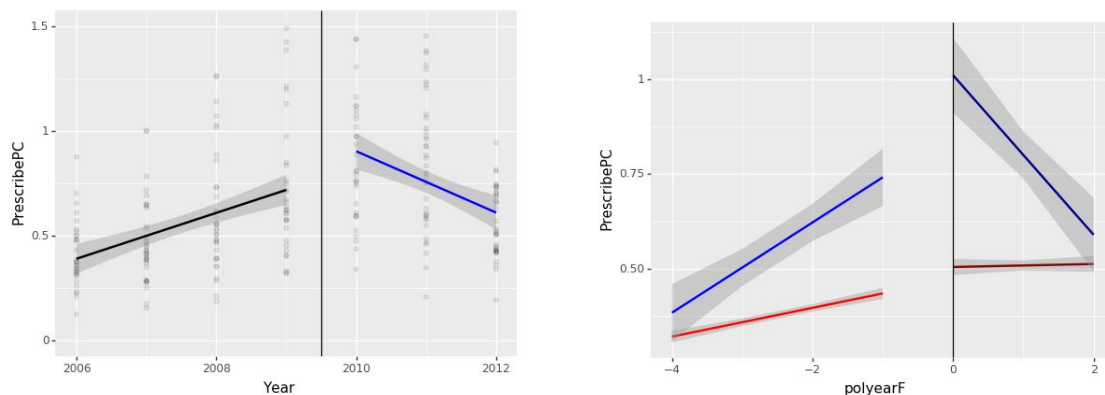
Here we can see that Florida seems to have turned an upward trend to a downward one after the policy enactment. It becomes even clearer in the DID graph. With the Florida counties and USA counties seemingly following the same trend before the policy we can reasonably assume that they have matching behavior in a lot of respects. After the policy, it seems that the USA trend continues while Florida has a complete change. The difference in differences is very stark here. This graph is strong evidence that the policy caused there to be less of these specific kinds of drug-related deaths in Florida.

DID Regression Analysis for overdose deaths

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|------------|-------|--------|--------|
| ===== | | | | | | |
| Dep. Variable: | DeathsPC | R-squared: | 0.001 | | | |
| Model: | OLS | Adj. R-squared: | 0.001 | | | |
| Method: | Least Squares | F-statistic: | 3.331 | | | |
| Date: | Sun, 10 Nov 2019 | Prob (F-statistic): | 0.0358 | | | |
| Time: | 15:14:44 | Log-Likelihood: | -11849. | | | |
| No. Observations: | 7917 | AIC: | 2.370e+04 | | | |
| Df Residuals: | 7914 | BIC: | 2.373e+04 | | | |
| Df Model: | 2 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| Intercept | 1.5599 | 0.018 | 88.058 | 0.000 | 1.525 | 1.595 |
| Policy | -0.0567 | 0.025 | -2.308 | 0.021 | -0.105 | -0.009 |
| Policy:State | 0.1266 | 0.087 | 1.457 | 0.145 | -0.044 | 0.297 |
| ===== | | | | | | |
| Omnibus: | 5512.206 | Durbin-Watson: | 0.670 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 103839.972 | | | |
| Skew: | 3.131 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 19.601 | Cond. No. | 8.30 | | | |
| ----- | | | | | | |

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|------------|-------|--------|--------|
| ===== | | | | | | |
| Dep. Variable: | DeathsPC | R-squared: | 0.004 | | | |
| Model: | OLS | Adj. R-squared: | 0.003 | | | |
| Method: | Least Squares | F-statistic: | 5.061 | | | |
| Date: | Sun, 10 Nov 2019 | Prob (F-statistic): | 3.42e-05 | | | |
| Time: | 15:14:45 | Log-Likelihood: | -11837. | | | |
| No. Observations: | 7917 | AIC: | 2.369e+04 | | | |
| Df Residuals: | 7910 | BIC: | 2.374e+04 | | | |
| Df Model: | 6 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| Intercept | 1.6393 | 0.038 | 43.432 | 0.000 | 1.565 | 1.713 |
| adj_year | 0.0186 | 0.009 | 2.049 | 0.041 | 0.001 | 0.036 |
| Policy | -0.0946 | 0.049 | -1.945 | 0.052 | -0.190 | 0.001 |
| Policy:State | 0.3842 | 0.156 | 2.460 | 0.014 | 0.078 | 0.690 |
| Policy:adj_year | -0.0347 | 0.013 | -2.582 | 0.010 | -0.061 | -0.008 |
| State:adj_year | 0.0588 | 0.019 | 3.071 | 0.002 | 0.021 | 0.096 |
| Policy:State:adj_year | -0.1593 | 0.054 | -2.944 | 0.003 | -0.265 | -0.053 |
| ===== | | | | | | |
| Omnibus: | 5491.417 | Durbin-Watson: | 0.672 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 102512.515 | | | |
| Skew: | 3.117 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 19.489 | Cond. No. | 53.4 | | | |
| ===== | | | | | | |

Pre-post comparison and DID graphs for prescription



With prescriptions we see similar behavior than the deaths. In the PrePost graph we can identify the same switch in slope, where after the policy it becomes negative. When we continue to the DID graph we have a similar comparison but it seems that Florida counties have a much more exaggerated behavior than that of the rest of the US. When we consider the post-policy years, Florida has a negative slope while the US slope seems

to be somewhat flat. This means that the policy managed to change the trend to negative. We cannot know yet, however, whether this effect is sustainable or if Florida is approaching USA levels of prescribed opioids.

DID Regression Analysis for prescription

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|------------|-------|--------|--------|
| ===== | | | | | | |
| Dep. Variable: | PrescribePC | R-squared: | 0.019 | | | |
| Model: | OLS | Adj. R-squared: | 0.019 | | | |
| Method: | Least Squares | F-statistic: | 543.2 | | | |
| Date: | Sun, 10 Nov 2019 | Prob (F-statistic): | 2.34e-234 | | | |
| Time: | 15:30:47 | Log-Likelihood: | -5490.0 | | | |
| No. Observations: | 55419 | AIC: | 1.099e+04 | | | |
| Df Residuals: | 55416 | BIC: | 1.101e+04 | | | |
| Df Model: | 2 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| Intercept | 0.4417 | 0.002 | 263.774 | 0.000 | 0.438 | 0.445 |
| Policy | -0.0319 | 0.002 | -13.850 | 0.000 | -0.036 | -0.027 |
| Policy:State | 0.2216 | 0.007 | 31.703 | 0.000 | 0.208 | 0.235 |
| ===== | | | | | | |
| Omnibus: | 28868.651 | Durbin-Watson: | 0.265 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 249011.523 | | | |
| Skew: | 2.370 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 12.240 | Cond. No. | 7.19 | | | |
| ===== | | | | | | |

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|------------|-------|--------|--------|
| ===== | | | | | | |
| Dep. Variable: | PrescribePC | R-squared: | 0.043 | | | |
| Model: | OLS | Adj. R-squared: | 0.043 | | | |
| Method: | Least Squares | F-statistic: | 417.3 | | | |
| Date: | Sun, 10 Nov 2019 | Prob (F-statistic): | 0.00 | | | |
| Time: | 15:31:34 | Log-Likelihood: | -4803.3 | | | |
| No. Observations: | 55419 | AIC: | 9621. | | | |
| Df Residuals: | 55412 | BIC: | 9683. | | | |
| Df Model: | 6 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| Intercept | 0.4231 | 0.004 | 119.062 | 0.000 | 0.416 | 0.430 |
| adj_year | -0.0012 | 0.001 | -1.394 | 0.163 | -0.003 | 0.000 |
| Policy | 0.0001 | 0.005 | 0.029 | 0.977 | -0.009 | 0.009 |
| Policy:State | 0.2413 | 0.012 | 19.716 | 0.000 | 0.217 | 0.265 |
| Policy:adj_year | -0.0040 | 0.001 | -3.181 | 0.001 | -0.006 | -0.002 |
| State:adj_year | -0.0544 | 0.002 | -36.069 | 0.000 | -0.057 | -0.051 |
| Policy:State:adj_year | 0.0463 | 0.004 | 10.775 | 0.000 | 0.038 | 0.055 |
| ===== | | | | | | |
| Omnibus: | 28815.198 | Durbin-Watson: | 0.272 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 253345.452 | | | |
| Skew: | 2.356 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 12.355 | Cond. No. | 45.9 | | | |
| ===== | | | | | | |

References

- [1] "Overdose Death Rates." *NIDA*, National Institute on Drug Abuse, 29 Jan. 2019, www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates.
- [2] "ARCOS Registrant Handbook." *ARCOS Registrant Handbook - Appendix 5*, www.deadiversion.usdoj.gov/arcos/handbook/appendix5.htm.
- [3] "FIPS Code." *NRCS*, www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?&cid=nrcs143_013697.
- [4] "US Census Population Data." *Index of /Programs-Surveys/Popest/Datasets*, www2.census.gov/programs-surveys/popest/datasets/.