Causal Recreational Cannabis

Patrick Massey and Elliott Metzler

5/13/2022

Abstract

Can legalizing the recreational use of cannabis reduce rates of drug poisoning deaths? This paper seeks to study the relationship between changes in state-level policy toward recreational cannabis use and drug poisoning death rates. Specifically, we evaluate the impact of such policy change on the first state to legalize, Colorado. We implement a synthetic control design in which we use demographic data for each state in the United States, including Washington D.C., and estimate the causal impact of Colorado's legalization policy shift in 2012. We find [[insert findings and explanation of implications]].

1 Introduction and Background

In 1970 with the passage of the Controlled Substances Act cannabis was labeled a Schedule 1 drug and outlawed in all US states. Since then, the debate on cannabis legalization has continued to rise. People against legalization claim that cannabis is a gateway drug that will lead individuals into more serious and harder drugs. Proponents of legalization claim that regulation will lead to a safer product for consumption. According to the 2019 National Survey on Drug Use and Health, cannabis is the most commonly used illegal drug with 17.5% of Americans aged 12 or older having answered yes to using the drug at least once in 2019. It is a reasonable to assume that there is certain percentage of that population consuming cannabis in a state where it is currently illegal, and they are doing it for recreational purposes. The way these individuals obtain the illicit drugs is through a black market with obviously no regulation. With the lack of regulation there is a not insignificant probability of cannabis obtained through a black market being laced with a much stronger and potentially lethal drug.

For twenty years cannabis remained totally and completely illegal in the US, however in 1991 the first dispensaries appeared in the country. In 1991, the city of San Francisco legalized cannabis for medical use, and California followed in 1996. This would set off a chain of states passing and regulating cannabis for medical use. However, cannabis would remain illegal for recreational use until 2012. In 2012, Colorado and Washington State legalized cannabis for recreational use paving the way for dispensaries, and regulation for the safe consumption of cannabis. Following Colorado and Washington State's path multiple states have passed similar legislation legalizing the recreational use of cannabis. In Table 1 below, we summarize the rollout of cannabis legalization up until 2018 which is when our study ends.

Table 1: Marijuana Legalization by State

State	Year
Colorado	2012
Washington	2012
Alaska	2014
Oregon	2014
California	2016
Maine	2016
Massachusetts	2016
Nevada	2016
Michigan	2018
Vermont	2018

The studies on the implications of legalization are mixed which is a natural outcome of a topic that impacts many aspects of our lives, directly or indirectly. In a study published by Stacy Salomonsen-Sautel et. al. the authors analyze the effect of legalization of medical cannabis on fatal motor vehicle crashes in Colorado. In their results they find that a significant positive change in drivers testing positive for cannabis in fatal vehicle crashes after the law changed. In contrast however, Livingston et. al. examine the effects of recreational legalization on the number of opioid-related deaths. Livingston et. al. find that after legalization there was a significant decrease in the number of opioid-related deaths per month. In this study we seek to expand on the analysis done by Livingston et. al., and to add to the discussion of legalization.

We will be examining drug poisoning deaths across the US from 2000 to 2018. The Center for Disease Control (CDC) classifies a drug poisoning death as a drug-related death that is unintentional, a homicide, a suicide, or of unknown intent. From Figure 1 below we see that drug poisoning deaths have been on the with a large increase in recent years. We also note that we do not see the same trend with Colorado.

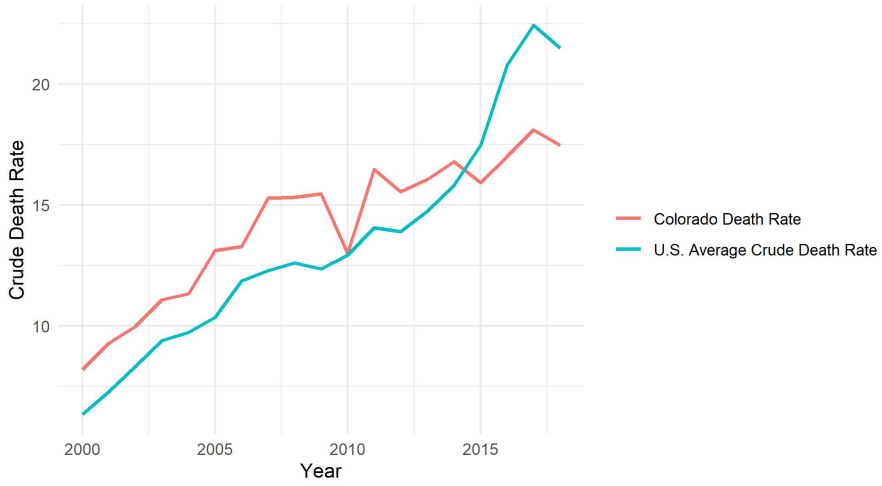

Figure 1: Drug Poisoning Death Rates Trend (2000-2018)

This paper seeks to estimate the causal effect of cannabis legalization on drug poisoning deaths in Colorado. We aim to do this by using synthetic control, a quasi-experimental research design, to examine the changes in the crude death rate for drug poisoning deaths in Colorado vs. states that have not legalized.

2 Data

The data employed for this paper originates from two sources. First, we source data on drug poisoning deaths, population, and an estimate of the crude drug poisoning death rate from the United States Center for Disease Control ("CDC"). Second, we source state-level demographic data from IPUMS USA, which is a database harmonizing, organizing, and structuring data from various population surveys in the United States over time.

The first data set contains summary statistics by state and by year from 1999 to 2018. For each state in each year, the data reports the number of deaths due to drug poisoning, the population, and the ratio of the two previous variables as the crude death rate. Additionally, this data reports an estimate of the standard error for the crude death rate, and a lower and upper confidence interval around the point estimate. In addition to these base estimates,

the data originally includes age-adjusted rates with standard errors and confidence intervals and the United State-wide crude death rate and United State-wide age-adjusted death rate.

From this data set, we focus on the crude death rate for the state and extract this variable along with the state and year variables for data merging purposes. We focus our attention on the crude death rate statistic for a couple of important reasons. First, using the death rate as opposed to the total number of deaths allows us to immediately control for differences in state population when estimating our model. Second, we exclude the age-adjusted rate because our model also accounts for age demographics explicitly as covariates, so it would be inappropriate to use this as our outcome variable of interest.

The second data set contains nearly 28 million entries where each entry represents a person included in [[a survey]] in a particular year. Importantly, this data includes the year and state in which that entry resided along with extensive demographic data. Thus, we are able to use the groups of these entries associated with each year and each state and calculate demographic summary information representative of each year in each state. We include in our demographic summaries the proportion of the population that is listed as male as compared to female [[should footnote simplicity of coding here (I don't remember seeing info about non-binary or choose not to identify)]], a summary of a few significant age buckets, proportions of the population based on race, on marital status, on education, and on employment status. Additionally, we summarize the mean number of hours worked, the median income, the mean number of children, and the mean number of children under 5 years old. We present the mean and standard deviation of each of these demographic variables in Table 2.

To construct our final data set for use in analysis, we combine the two data sources. Since we have summarized the latter data set at the state and year level and the first data set is already at this level, we merge on the state and year. Thus, we have data for each state, inclusive of Washington D.C., for each year between 2000 and 2018 with data on deaths associated with drug poisoning, population, crude death rate, and all of our descriptive demographics data.

Table 2: Variable Summary

Variable	Mean	Std. Dev
Drug Poisonings		
Deaths	777.179	887.743
Population	5995451.840	6730237.227
Crude Death Rate	13.381	7.082
Gender Population Proportions		
Male	0.489	0.012
Female	0.511	0.012
Age Population Proportions		
Under 30	0.176	0.022
Under 50	0.450	0.042
Over 50	0.374	0.045
Race Population Proportions		
American Indian	0.018	0.039
Asian	0.039	0.070
Black	0.088	0.094
Other Race	0.045	0.039
White	0.810	0.131
Hispanic Population Proportions		
Hispanic	0.076	0.081
Not Hispanic	0.924	0.081
Marital Status Population Proportions		
Married	0.609	0.062
Not Married	0.391	0.062
Education Population Proportions		
Less Than High School	0.084	0.029
High School	0.372	0.050
Some College	0.247	0.036
College	0.191	0.034
More Than College	0.106	0.041
Employment Population Proportions and Summary Statistics		
Employed	0.724	0.045
Not Employed	0.276	0.045
Mean Hrs Worked	32.222	2.269
Median Income	22380.650	5043.007
Children Summary Statistics		
Mean Children	0.823	0.107
Mean Children Under 5 years old	0.168	0.029

Please see the appendix for detailed descriptions of our procedure to re-code and summarize variables.

3 Methodology

We utilize the synthetic control method first developed by Abadie and Gardeazabal (2003) and then expanded upon in Abadie et. al (2010)(2015). Similar to Abadie et. al (2010) we are analyzing the impact of a government policy shock. In this analysis we are estimating the impact of Colorado's legalization of recreation cannabis in 2012. For our synthetic control, Colorado is the treatment state of interest, and we have a donor pool of 42 states. In order to create our donor pool of states we remove any states that also experienced a legalization shock during our study period.

Suppose that we have $1, 2...S+1$ states and let $s = 1$ be the treated state, additionally we have $t = 1, 2...T$ time periods where T_0 represents the number of pre-treatment periods, and $T_0 + 1...T$ are the post-treatment periods. With that let Y_{st} be the crude death rate for state s at time t. In a post-treatment time line we are estimating

$$\hat{\alpha} = Y_{1t} - \sum_{s=2}^{S+1} w^* Y_{st}$$

w_s^* is a vector of optimally chosen weights for states from our donor pool to represent our synthetic Colorado. Again deriving from Abadie (2010) w_s^* is the vector of weights that minimizes the following equation

$$(X_1 - X_0 w)' V (X_1 - X_0 w)$$

X_1 is a $(K \times 1)$ vector of state population variables and X_0 is a $(K \times S)$ matrix of state population variables for S donor states. Note that $w_s \geq 0$ and $\sum w_s = 1$.

We utilize the weights generated on states in the donor pool to create a synthetic Colorado. The synthetic Colorado is one where we examine the counterfactual of a Colorado that never legalized recreational cannabis. Additionally we follow Abadie (2010) and use their placebo technique as a robustness check. The placebo technique consists of assigning each state in our donor pool as the treated state and then creating an optimal w_s^*, from the

other donor states as well as the treated state, in the pre-treatment period. Utilizing the weights we calculate the post-treatment root mean squared prediction error (RMSPE) for each state. We then take a ratio between the post and pre-treatment RMSPE for each state. We expect this value to be high for our treated state, which would suggest that the post-treatment RMSPE is large due to the differences between the synthetic and the observed Colorado.

4 Results

By applying the techniques described above, we are able to construct a synthetic Colorado with weights on the donor states such that the synthetic Colorado's crude death rate closely matches actual Colorado"s crude death rate in the period preceding legalization of recreational marijuana. Table 3 displays the weights for each of the donor states with a contribution to Synthetic Colorado over 0.1 percent. We see that Arizona and New Hampshire contribute most significantly to Synthetic Colorado, with weights of approximately 45 and 38 percent, respectively. We also see Washington D.C. (District of Columbia) and Utah contributing at a health level of approximately 10 and 8 percent, respectively. Though many donor states have nonzero weights in our analysis, we exclude them from this table because they contribute at a level below 0.1 percent.

Table 3: Synthetic Weights

Unit	Weight
Arizona	0.454
New Hampshire	0.375
District of Columbia	0.095
Utah	0.076

[a] Table excludes donor states with 0.1 percent model weight.

Next, we present the aggregate level predictors used in the model and compare them across Colorado, Synthetic Colorado, and the overall donor sample. We note that there are some

differences between the observed Colorado and our synthetic Colorado, specifically regarding education level. However, the synthetic Colorado provides a much better estimation of the observed Colorado than our donor pool.

Table 4: Balance Table

Variable	Colorado	Synthetic Colorado	Donor Sample
American Indian Proportion	0.009	0.022	0.014
Asian Proportion	0.026	0.025	0.035
Black Proportion	0.030	0.059	0.098
College Proportion	0.245	0.204	0.184
Employed Proportion	0.753	0.732	0.726
Female Proportion	0.508	0.514	0.515
Hispanic Proportion	0.139	0.120	0.067
High School Proportion	0.304	0.341	0.384
Less Than High School Proportion	0.068	0.084	0.090
Male Proportion	0.492	0.486	0.485
Married Proportion	0.632	0.604	0.627
Mean Children	0.813	0.848	0.844
Mean Children U5	0.183	0.179	0.174
Mean Hrs Worked	33.882	32.677	32.706
Median Income	24523.077	23473.869	21327.955
More Than College Proportion	0.130	0.122	0.100
Not Employed Proportion	0.247	0.268	0.274
Not Hispanic Proportion	0.861	0.880	0.933
Not Married Proportion	0.368	0.396	0.373
Other Race Proportion	0.064	0.056	0.041
Over 50 Proportion	0.337	0.348	0.360
Some College Proportion	0.253	0.250	0.242
Under 30 Proportion	0.180	0.180	0.175
Under 50 Proportion	0.483	0.472	0.466
White Proportion	0.870	0.837	0.813

We now plot the observed Colorado versus our synthetic Colorado in Figure X below. The dashed vertical line indicates when treatment begins. As we noted above our synthetic Colorado is very similar to the observed Colorado, which we see below. In the pre-treatment period our synthetic Colorado tracks closely with the observed Colorado. In the post-treatment period, we see the sharp separation from the synthetic and the observed. In Figure 2 we plot the gap between the synthetic and observed.

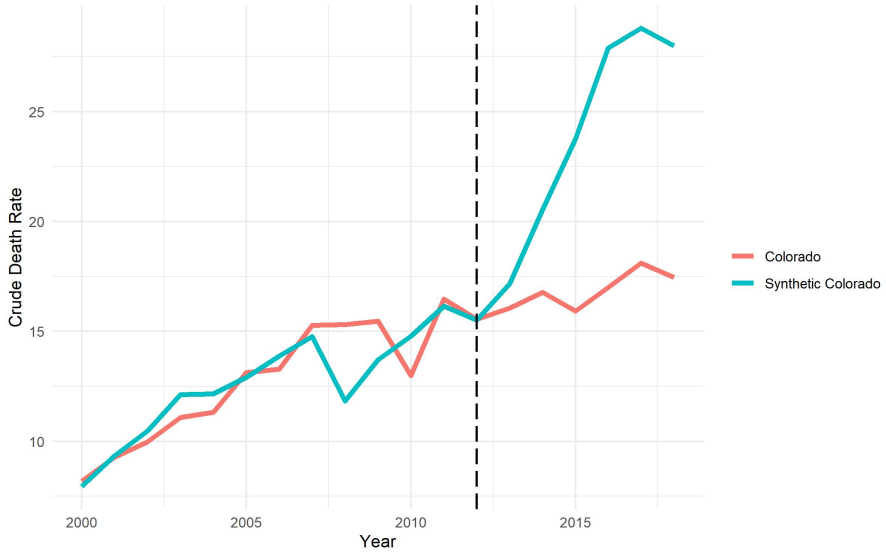

Figure 2: Trends in Crude Death Rate: Colorado vs. Synthetic Colorado

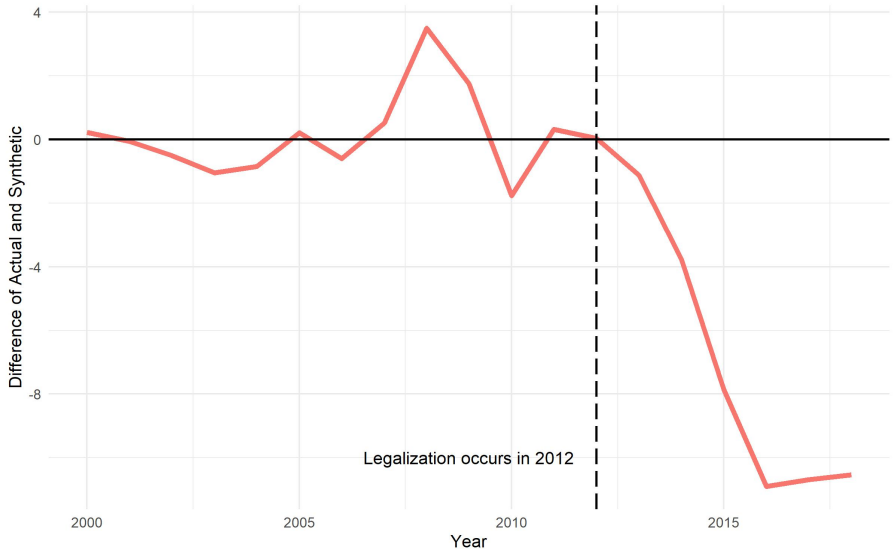

Figure 3: Differences in Crude Death Rate

In Figure 4 below we plot the ratio between the post-treatment RMSPE and the pre-treatment RMSPE. What we expect to see is that the treated unit will have a small pre-treatment RMSPE which would indicate a close fit between the observed and synthetic units

in the pre-treatment period. We would also expect to see the post-treatment RMSPE to be large as that would indicate a separation between the observed and synthetic units. We see that below [[Need to somehow explain new Hampshire]]

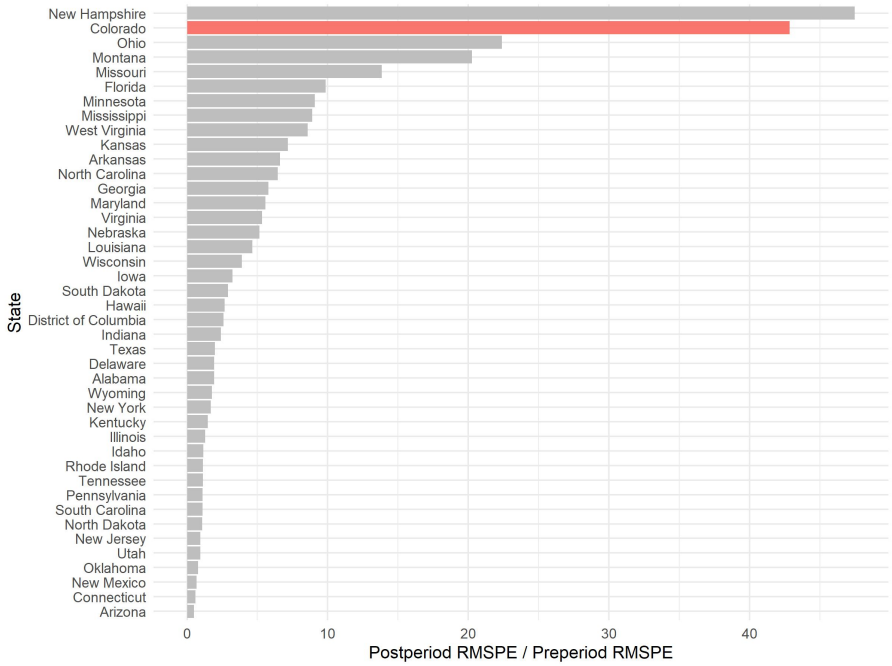

Figure 4: RMSPE Ratio Plot

As a robustness check we perform the placebo test as suggested by Abadie et al. (2010). When performing the placebo, we reassign the treatment to each state in the donor pool while placing Colorado into the donor pool. We then run synthetic control on all the donor states and plot their outcomes in Figure 5 below. The placebo test confirms that our results are strong. Only one placebo had seemed to standout. [[Can we find out which state that is explain?]]

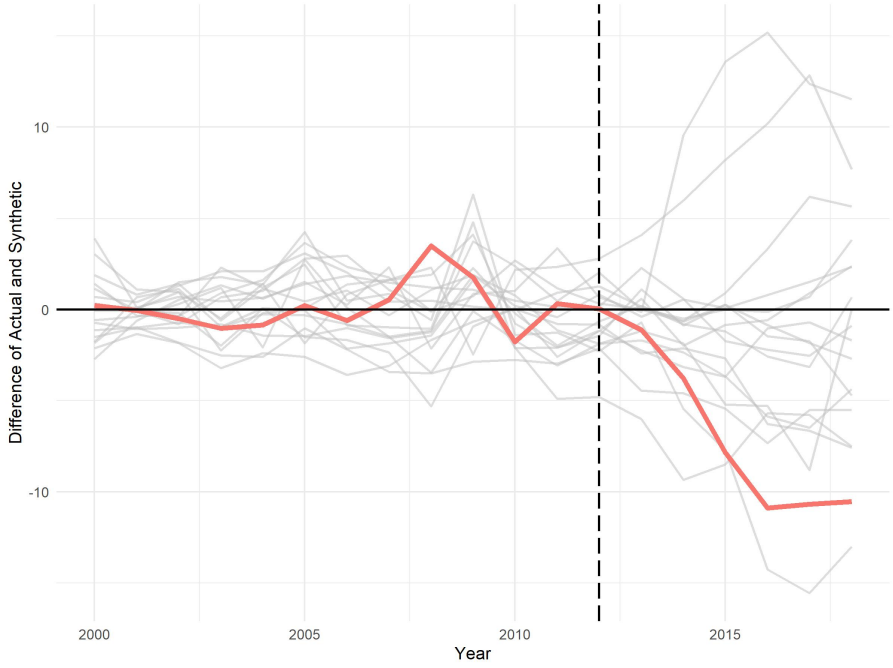

Figure 5: Placebos Plot

5 Discussion

6 Conclusion

7 Appendix

7.1 Variable Re-coding Procedure Detail

- Gender (*sex*)

  Original coding:

  - 1 = Male
  - 2 = Female

  Summarized by calculating proportion of total that is male and proportion of total

that is female.

- Age (*age*)

  Original coding:

  - Actual age of entry

  Summarized by categorizing into the following age groups and calculating the proportion of the total that fall into that age group.

  - Under 30 years old
  - Between 30 years old and under 50 years old
  - Over 50 years old

- Need to do - finish transferring notes from the markdown file into here.