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 a significant decline in drug poisoning death rate ranging from around 3.8 to 10.9 [[Could we calculate an average?]] deaths per 100,000 depending on when measuring after treatment.

1 Introduction and Background

In 1970, with the passage of the Controlled Substances Act, cannabis was labeled a Schedule 1 drug and outlawed across the United States ("U.S."). 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, while proponents of legalization claim that legalization and 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 percent 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 a certain percentage of that population consuming cannabis in a state where it is currently illegal, and they are doing so for recreational purposes. The way these individuals obtain the illicit drugs is through a black market which obviously lacks regulation. Of course, due to the lack of regulation, the likelihood of obtaining cannabis laced with potentially more lethal drugs increases. Furthermore, since a buyer must go to the black market to obtain their cannabis to begin with, the likelihood of substitution toward alternative riskier drugs is bound to increase.

For twenty years cannabis remained completely illegal in the U.S. However, in 1991, the first dispensaries appeared in San Francisco as the city legalized the use of cannabis for medical purposes, with the state of California (in which San Francisco is located) following suite in 1996. This set off a chain of states legalizing cannabis for medical use. Despite the warming around the country to the medical benefits of cannabis, it would remain illegal for recreational use until 2012. In 2012, Colorado and Washington legalized cannabis for recreational use paving the way for dispensaries, and regulation for the safe consumption of cannabis. Following Colorado and Washington's path, eight additional states have passed similar legislation legalizing the recreational use of cannabis. In Table 1 below, we summarize the roll out of recreational cannabis legalization up through 2018 (the final year of our study). [[Elliott left off here]]

Table 1: Legalization of Recreational Cannabis 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 Salomonsen-Sautel et al. (2014) 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. (2017) examine the effects of recreational legalization on the number of opioid-related deaths. The authors 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 rise with a large increase in recent years. We also note that we do not see the same trend with Colorado.
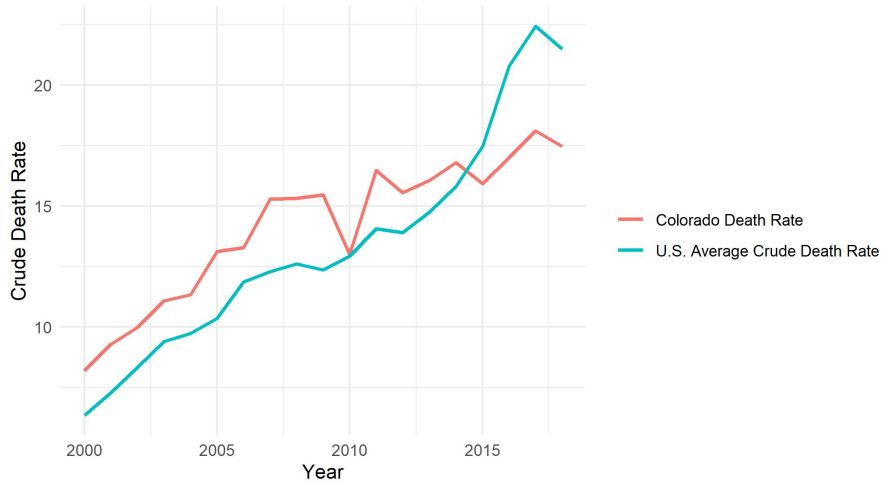
Figure 1: Drug Poisoning Death Rates Trend (2000-2018)

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" (National Center for Health Statistics, 2019). 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.(Ruggles et al., 2021)

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 the American Community 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

5

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

3 Methodology

We utilize the synthetic control method first developed by Abadie & Gardeazabal (2003) and then expanded upon in Abadie et al. (2010) and Abadie et al. (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 recreational 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 et al. (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

4.1 Model

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 Colorado Donor State Weights

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

[a] Table excludes donor states with less than 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: Synthetic Colorado Predictor Variable Balance Summary

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

4.2 Estimated Impact

We now plot the observed Colorado versus our synthetic Colorado in Figure 2 below. The dashed vertical line indicates when treatment begins. As we noted above our synthetic

9

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.
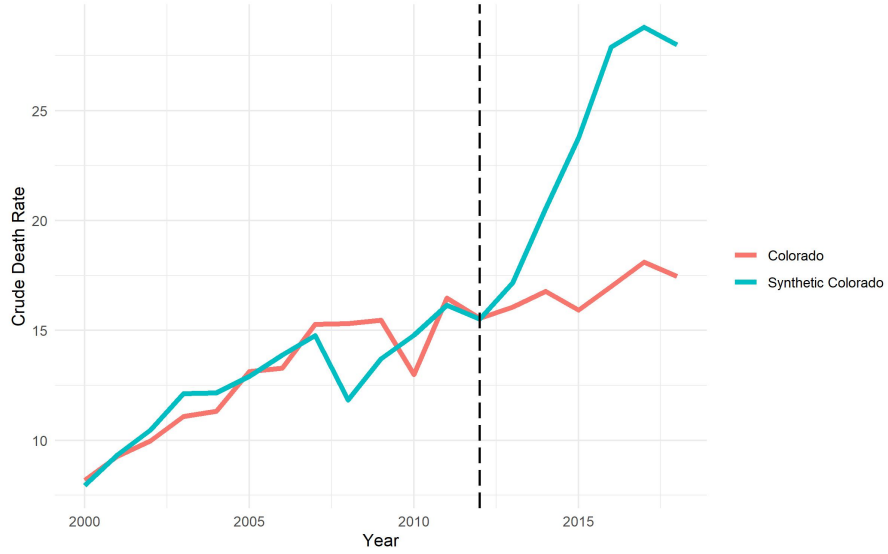

Figure 2: Trends in Drug Poisoning Death Rate: Colorado vs. Synthetic Colorado

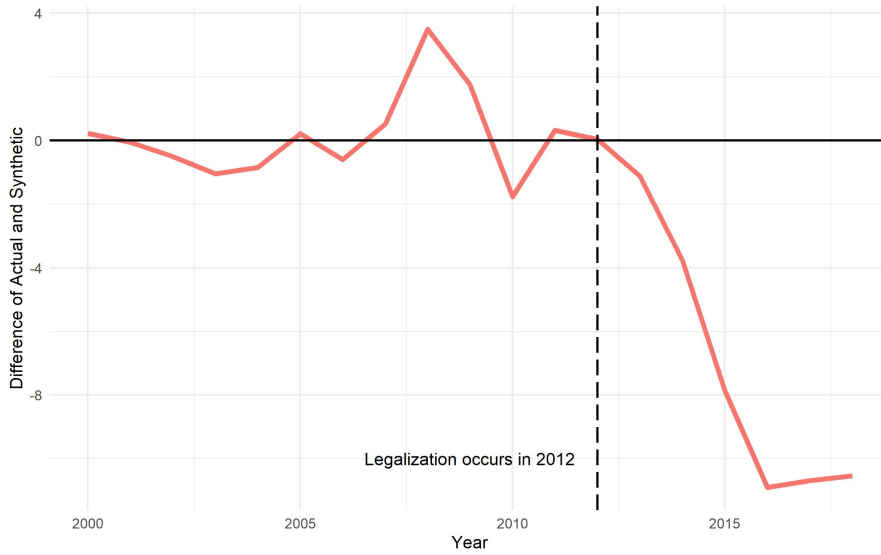In Figure 3 we plot the gap between the synthetic and observed.

Figure 3: Differences in Drug Poisoning Death Rate

Finally, we present the numerical results associated with Figure 3 in 5. We estimate the causal effect of legalization of Marijuana on crude death rate in Colorado to be a decline of approximately 3.75 two years after legalization, decline of approximately 10.9 four years after legalization, and a decline of approximately 10.5 six years after legalization.

Table 5: Estimated Impact of Legalization on Drug Poisoning Death Rate: Colorado

| Years After Legalization | Drug Poisoning Death Rate | | |
	Actual	Synthetic	Estimated Causal Effect
2	16.785	20.552	-3.767
4	17.002	27.899	-10.897
6	17.470	28.002	-10.532

4.3   Robustness Checks

In Figure 4 below we plot the ratio between the post-treatment RMSPE and the pre-treatment MSPE. 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

11

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 generally holds in the figure below, with Colorado standing well differentiated from each other state with the exception of New Hampshire. This is still a strong indication, if a state were picked at random from our sample, the probability of picking a state with a ratio as large as Colorado would be $2/43 \approx .047$.
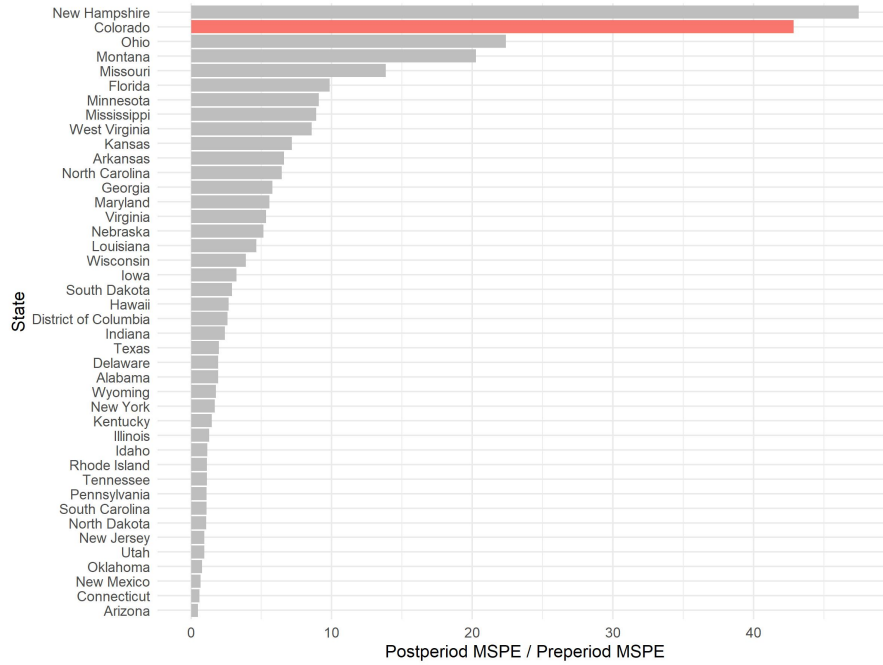

Figure 4: Root Mean Squared Prediction Error Ratio Plot - Colorado

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.
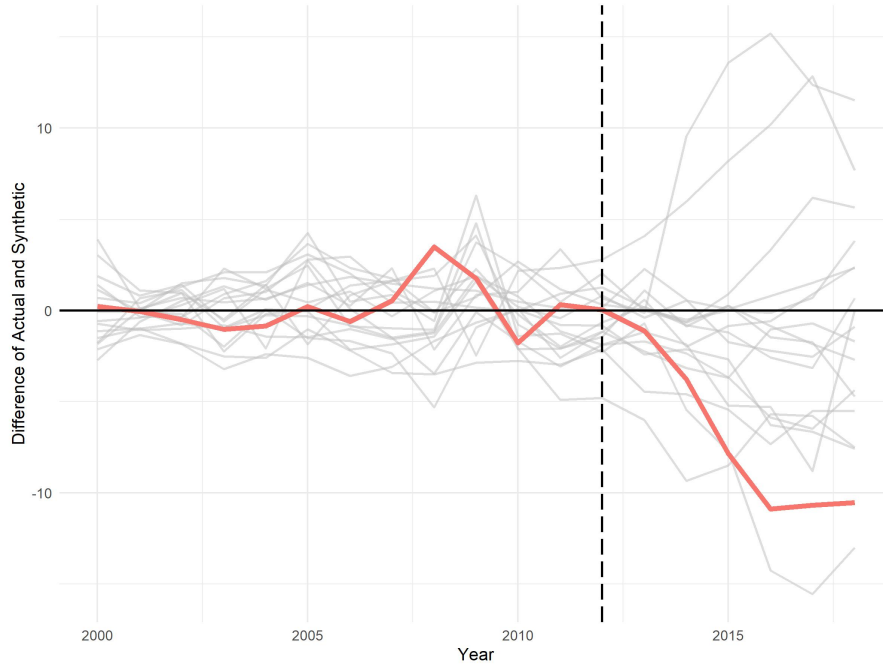
Figure 5: Placebo Analysis Plot - Colorado

To ensure we are comparing our result for Colorado with comparably fit placebos, we only chart and assess placebos for which their pretreatment period RMSPE is less than or equal to double the Colorado pre-treatment period RMSPE. This is because we are interested in assessing models where the fit is relatively good in the pre-treatment period, not those that we were not able to fit in the pre-treatment period. As shown in the figure above, the placebo test confirms the strength of our results, with only one placebo falling below Colorado. Montana ... [[Explain why?]]

For our final robustness check, we apply the same methodology described here and above to another state, Washington, that legalized marijuana at the same time as Colorado (2012). We present detailed results of this analysis in the appendix. When applying our synthetic control model to Washington, we find similarly statistically significant results, albeit with less sizable estimated causal effects. We estimate a decline in crude death rate of 2.3 two years removed from treatment, 6.4 four years after treatment, and 6.4 again six years after

13

treatment.

5 Discussion

Our findings in this paper strongly support the notion that legalization of marijuana has *some* positive societal benefit. Entirely separate from the potential economic gains of a new high growth industry and the job and tax revenue creation that implies, we have shown that there are potential social benefits. Perhaps it is not the case that marijuana is simply a gateway drug - perhaps it is rather the case that marijuana can provide a more gentle release than other intoxicating substitutes, and actually decrease drug risk overall. Our finding of a non negligible decrease in drug poisoning death rate in both Colorado and Washington in years following legalization of recreational marijuana use supports this notion, and potentially supports legalization more broadly across the United States.

We recognize that an interesting limit on the general applicability of our results is the heterogeneity of states and regions in the United States, and that broadly applying a legalization policy may not have identical positive effects. State and regional preferences, attitudes, and political leanings often impact how these sort of policies are received and can drastically change usage outcomes. It is worth noting that Colorado and Washington both are historically known as states that are more friendly and curious regarding recreational use of drugs, which may skew results relative to a generally applied law. Nonetheless, our finding is logically supported by accessibility-driven decision making. Naturally, if something is more accessible, we're likely to reach for that instead as long as we view it as a viable substitute. If recreational marijuana is available and accessible, perhaps fewer people are going to the trouble of finding a black market drug supplier and acquiring harder, riskier drugs. Furthermore, the risk of death from overdosing on marijuana is nearly nonexistent, particularly when compared to other psychoactive drugs.

Additionally, we recognize that our findings carry some important limitations. One key limitation is in our data. We are only able to use data at the state and year level, reducing some of the granularity in our results. Furthermore, we are limited primarily to demographic

14

data for this study, which is limiting our the ability to fit our model. Though matching a synthetic state based on broad demographics is a good start, we would ideally expand our model predictors to provide more of a rich feature set. Ultimately, we are estimating drug poisoning death rate, so additional data relating to this outcome would be very useful in producing a richer and more accurate model. Another limitation is in the number of years of data we have to study. Though we have many years in the pre-treatment period, we only have six years after treatment to assess impact. Ideally, we would have more years in the post-treatment period.

In this paper we have isolated Colorado and looked at the impact that the legalization of recreational cannabis had on the crude death rate for drug poisoning deaths. We reaffirmed our findings by performing a separate synthetic control on Washington which legalized recreational cannabis at the same time. However, as we stated earlier there have been many states that have legalized at various times. However in Ben-Michael et al. (2021), the authors present a unique version of synthetic control they call "Augmented Synthetic Control". In the synthetic control method we utilized we are limited to exploring a single treated unit, this is not the case with the augmented synthetic control. With augmented synthetic control one could get the causal effect with multiple treated units even if there is staggered adoption as we have in this case. A useful extension would be to apply the augmented synthetic control method and estimate the causal impacts of legalization with all of the treated states.

6 Appendix - Robustness check with Washington

First, we present the donor unit weights for the Synthetic Washington in Table 6.

Table 6: Synthetic Washington Donor State Weights

Unit	Weight
Florida	0.282
Utah	0.281
Connecticut	0.212
Montana	0.115
Hawaii	0.048
Arizona	0.047
District of Columbia	0.015

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

We see that Synthetic Washington is comprised of seven donor states with more than 0.1 percent contribution. Three donor states stand out with over 20 percent contribution: Florida, Utah, and Connecticut. Additionally, we see three of the key states overlap with donor states for Synthetic Colorado: Arizona, Washington D.C. (District of Columbia), and Utah. Though New Hampshire carries the second largest weight for Synthetic Colorado, it is absent from the key donors list for Washington.

Next, we present the analogous balance table below in Table 7. In the balance table we see similar results that we saw in the Colorado analysis. The synthetic Washington results in a much closer fit to the observed Washington, than our donor sample. In fact our synthetic Washington matches the observed Washington better than the synthetic and observed Colorado.

Table 7: Synthetic Washington Predictor Variable Balance Summary

Variable	Washington	Synthetic Washington	Donor Sample
American Indian Proportion	0.014	0.014	0.014
Asian Proportion	0.071	0.049	0.035
Black Proportion	0.026	0.061	0.098
College Proportion	0.212	0.200	0.184
Employed Proportion	0.719	0.728	0.726
Female Proportion	0.510	0.513	0.515
Hispanic Proportion	0.063	0.106	0.067
High School Proportion	0.332	0.356	0.384
Less Than High School Proportion	0.065	0.072	0.090
Male Proportion	0.490	0.487	0.485
Married Proportion	0.626	0.638	0.627
Mean Children	0.817	0.935	0.844
Mean Children U5	0.173	0.202	0.174
Mean Hrs Worked	31.858	32.302	32.706
Median Income	22900.000	21956.634	21327.955
More Than College Proportion	0.110	0.108	0.100
Not Employed Proportion	0.281	0.272	0.274
Not Hispanic Proportion	0.937	0.894	0.933
Not Married Proportion	0.374	0.362	0.373
Other Race Proportion	0.053	0.050	0.041
Over 50 Proportion	0.355	0.349	0.360
Some College Proportion	0.280	0.263	0.242
Under 30 Proportion	0.173	0.186	0.175
Under 50 Proportion	0.472	0.466	0.466
White Proportion	0.837	0.826	0.813

When looking to the pair of trend and difference charts for Washington, we see a similar picture as we found for Colorado. Where the Synthetic Washington and Synthetic Colorado increase sharply after the treatment in 2012, the actual Washington and Colorado level off.
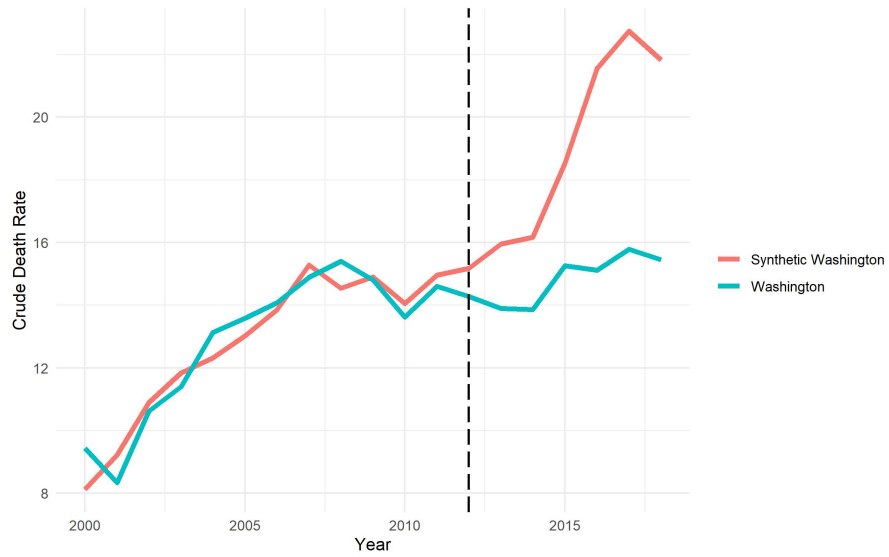
Figure 6: Trends in Drug Poisoning Death Rate: Washington vs. Synthetic Washington

This leveling off is further highlighted in the difference chart, where we see a notable negative gap.
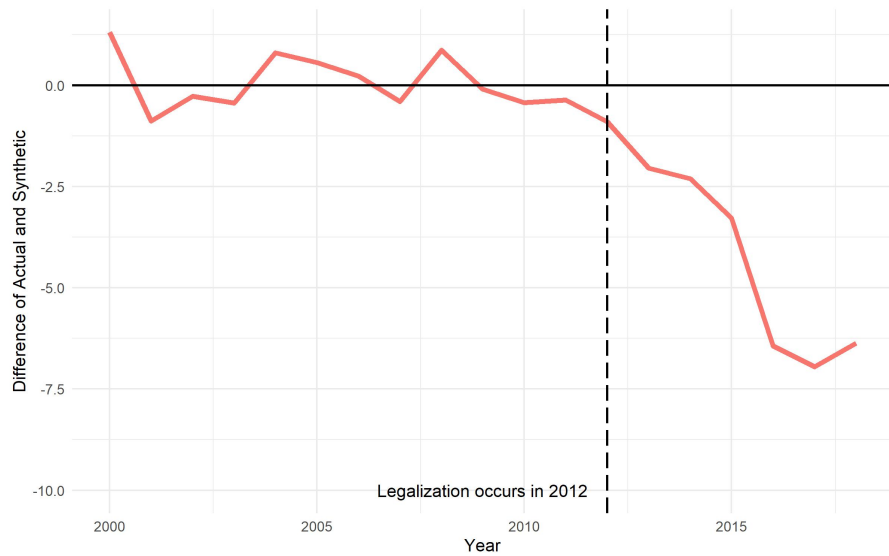

Figure 7: Differences in Drug Poisoning Death Rate

To summarize our ultimate causal estimates, we present a table analogous to 5 as 8. We see

18

that in the years following legalization we estimate a negative causal impact of treatment. Though the results for Washington are not as large in order of magnitude as Colorado, it is encouraging for our model that we estimate a similar results path and same signed outcomes.

Table 8: Estimated Impact of Legalization on Drug Poisoning Death Rate: Washington

Years After Legalization	Drug Poisoning Death Rate		
	Actual	Synthetic	Estimated Causal Effect
2	13.864	16.167	-2.303
4	15.121	21.554	-6.433
6	15.447	21.819	-6.372

Finally, we present the analogous MSPE ratio plot and placebos plot for Washington. Unlike Colorado, Washington's ratio of post-treatment MSPE to pre-treatment MSPE stands well apart from the placebos.
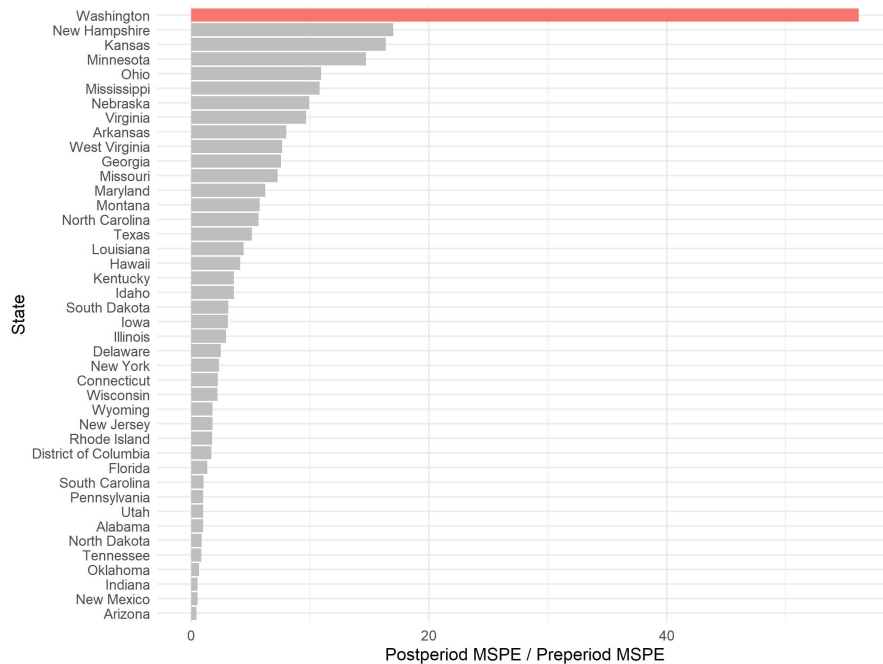

Figure 8: Mean Squared Prediction Error Ratio Plot- Washington

We note that in the plot of differences between actual and synthetic for the placebos, we see that when our model is applied to Washington we qualitatively find a lower fit quality. Indeed, only a small handful of placebo states fit the pruning criteria in our Washington model, as compared to nearly half of the states passing the pruning in our Colorado model.
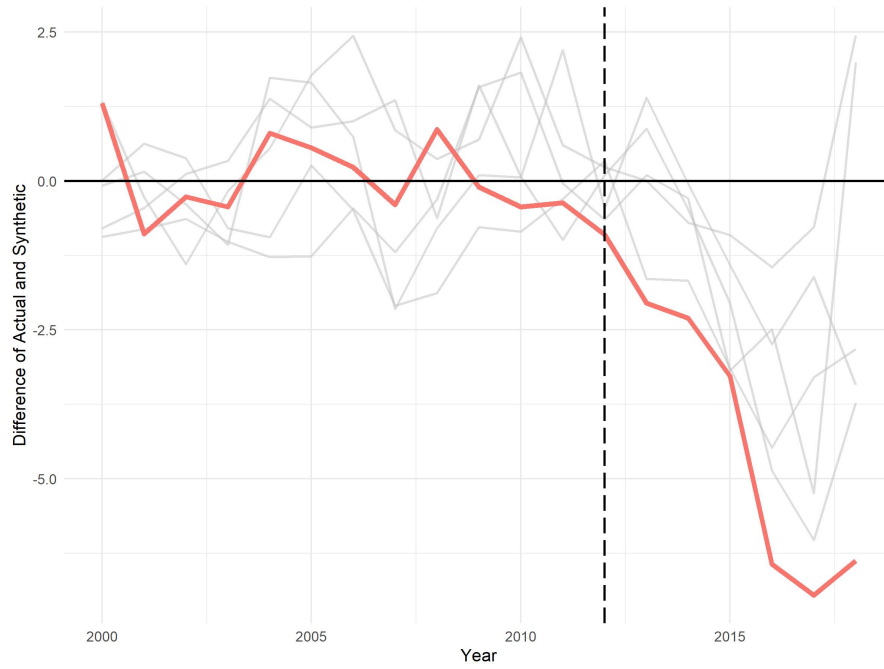

Figure 9: Placebo Analysis Plot - Washington

References

National Center for Health Statistics. (2019). *Nchs - drug poisoning mortality by state: United states.* (data retrieved from: CDC, https://data.cdc.gov/d/44rk-q6r2.)

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493-505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American journal of political science*, *59*(2), 495-510.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *The American economic review*, *93*(1), 113-132.

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789-1803. Retrieved from https://doi.org/10.1080/01621459.2021.1929245 doi: 10.1080/01621459.2021.1929245

Livingston, M. D., Barnett, T. E., Delcher, C., & Wagenaar, A. C. (2017). Recreational cannabis legalization and opioid-related deaths in colorado, 2000-2015. *American journal of public health (1971)*, *107*(11), 1827-1829.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., & Sobek, M. (2021). *Ipums usa: Version 11.0 [dataset]. minneapolis, mn: Ipums.* (data retrieved from: IPUMS USA, https://doi.org/10.18128/D010.V11.0)

Salomonsen-Sautel, S., Min, S.-J., Sakai, J. T., Thurstone, C., & Hopfer, C. (2014). Trends in fatal motor vehicle crashes before and after marijuana commercialization in colorado. *Drug and alcohol dependence*, *140*, 137-144.