The Effect of Recreational Marijuana Sales Legalization on Workplace Injuries: Evidence from Oregon*

Xiuming (Audrey) Dong†

November 12, 2019

Abstract

Recreational marijuana will soon be available to nearly 1 in 4 Americans under state statutes. This paper estimates the impact of the recreational marijuana sales legalization (RML) on workplace injuries. Using restricted-use Workers' Compensation claims as a proxy for injuries, I exploit variation in the county level implementation of recreational marijuana law in Oregon. Local governments could implement bans if less than 45% voters in their jurisdiction voted in favor of RML. In order to identify the casual impact of RML, I first utilize Difference-in-Difference strategy to compare the injury rate before and after RML for high exposure counties to the same difference for low exposure counties. To relax the parallel trends assumption, I use a data-driven procedure, known as the Synthetic Control Method to construct suitable comparison groups. Lastly, I exploit variation in the vote share rule near the cutoff under a Regression Discontinuity Design. My estimates suggest workplace injury rate is approximately 5%-20% higher for treated relative to control counties post-RML. It also indicates that RML increases work injury costs roughly by $7 to $34 million (or $5 to $24 per capita) per year. Overall, my results suggest recreational marijuana sales legalization may come at the expense of workplace injury.

*I am grateful to my advisors, Jeffrey Kubik and Gary Engelhardt, for their continued guidance and support. I also thank Dan Black, Arindrajit Dube, David Green, David Neumark, Perry Singleton, and seminar participants at Oregon State University, Syracuse University, and University of Hawaii for their valuable suggestions and comments. I appreciate the assistance of the Oregon Department of Consumer and Business Services, Oregon Workers' Compensation Department and Oregon Liquor Control Commission for their data support. The statements, findings, conclusions, views, and opinions contained and expressed herein are not necessarily those of Oregon Workers' Compensation Department, Oregon Occupational Safety and Health Administration or any of their affiliated or subsidiary entities. All errors are on my own.

†Dong: Department of Economics, Syracuse University, xdong102@syr.edu.

1 Introduction

The impact of substance use on labor market outcomes, such as workplace injuries, employment and wages, is a perennial topic of interest in public, labor and health economics (Kaestner and Grossman, 1995, 1998; DeSimone, 2002; Auld, 2005). While there has been longstanding interest in specific policies such as alcohol tax (Ohsfeldt and Morrisey, 1997; Dave and Kaestner, 2002), there has been substantially less emphasis on one of the fastest growing government drug policies, marijuana legalization. In particular, starting from Washington and Colorado, 11 states and the District of Columbia have passed recreational marijuana law since 2012. In other words, recreational marijuana will soon be available to nearly 1 in 4 United States residents. This paper provides the first causal estimates of the effect of recreational marijuana sales legalization (RML) on workplace injuries.

The link between RML and workplace injuries is important. The recreational marijuana market expands rapidly after states pass RML. For example, the monthly recreational marijuana sales went up to over $65 million within two years of legalization in Oregon as shown in Figure 1. Meanwhile, workplace drug test positivity rates have shown a strong increase since the passage of recreational marijuana law. This rise only holds for marijuana and not for other drug categories as presented in Figure 2. In addition, the epidemiological literature in the past two decades documents that marijuana use can have negative health effects, such as impaired cognitive, short-term memory, altered judgment, etc. (Hall and Degenhardt, 2009; Volkow et al., 2014; Hall, 2015). This suggests that marijuana use can create negative externalities in the workplace due to the increase of injury risk.[1] Given on-the-job injuries are costly (estimated cost in the United States was $192 billion in 2007 (Leigh, 2011)), somewhat surprisingly, there is little empirical evidence has been provided on the impact of RML on workplace safety.

[1] According to Grossman (1972)'s health capital model, where health has both consumption and investment aspects, the application to the risky behavior such as marijuana use is that when the individual has solved the constrained maximization problem, the optimal participation in risky behavior will be characterized by an equality of the marginal costs of the risky behavior (both the monetary cost of purchasing market good and the non-pecuniary cost of reduced health and shorter lifespan) and the marginal benefits (such as the instantaneous pleasure derived from consumption). So in this standard neoclassic model, people respond optimally to the costs and benefits of their available choices, and there is no reason to intervene except for externalities (Cawley and Ruhm, 2011).

I study Oregon for the following reasons. First, even though recreational marijuana law is a state policy, regulations are applied at a more local level in Oregon. Local governments that have less than 45 percent of voters who voted in favor of RML have prohibited the establishment of licensed recreational marijuana producers, processors, wholesalers, and/or retailers. Figure 3 shows these banning and allowing counties. This leads to a substantial amount of variation in recreational marijuana sales and licensed retail stores across counties. Second, I obtain detailed restricted-use administrative Workers' Compensation claims that serves as a proxy for workplace injuries, while most of the previous literature uses self-reported survey data such as National Longitudinal Survey of Youth.

Examining this link empirically is difficult because: (i) finding direct measure of marijuana usage is challenging; (ii) potential unobserved heterogeneity across states/counties both from the altitude towards legalization and workplace safety complicates the identification of the causal effects. To address the first challenge, I construct three measures of recreational marijuana exposure for each county in Oregon. The first measure is calculated by using monthly administrative recreational marijuana sales data from the Oregon Liquor Control Commission. It is defined as the total sales divided by the population within a county. Since cities also can ban marijuana sales following the same vote share rule, I create the second exposure measure defined as population share lives in the jurisdictions that allow recreational marijuana sales within a county. The last measure interacts the supply and latent demand for recreational marijuana in each county. Specifically, I multiply population share that has access to local recreational marijuana store(s) in a county with the baseline year workplace marijuana drug positivity rate.

To identify the causal impact of RML, I first employ a standard Difference-in-Difference (DiD) model, and compare the injury rate before and after RML for high exposure counties to the same difference for low exposure counties. In order to taking into account of the time-varing effect of the observed and unobserved predictors of the injury rate (such as rurality), I advocate a data-driven procedures, known as the Synthetic Control Method (SCM), to construct suitable comparison groups as in Abadie et al. (2010). The main underlying idea is that instead of choosing between low recreational marijuana exposure counties, or taking a simple average of injury rate in those counties, the SCM chooses weights for each of the those low exposure counties so that the weighted average best approximates pre-RML injury rate and other county characteristics of the treated county.[2]

Finally, I exploit variation in the 2014 electoral rules to estimate the effect of RML on the workplace injury rate under a Regression Discontinuity Design (RD). During November 2014, counties with less than 45 percent votes in favor of Measure 91 prohibited the establishment of licensed recreational marijuana producers, processors, wholesalers, and/or retailers.[3] Although counties that pass RML are likely to differ in both observable and unobservable ways from those that do not, these differences can be minimized by focusing on very close elections: a county that opted out of the recreational marijuana sales legalization by one vote is likely to be similar to one that passes RML by the same margin, though their "treatment" status will be quite different.[4]

My estimates suggest a substantial, statistically significant and positive impact of RML on workplace injuries. Overall, the workplace injury rate is approximately 5%-20% higher for treated relative to control counties post-RML. The restricted-use Workers' Compensation claims data allows for the assessment of heterogeneous effects across source of injury, age group, gender, industry and occupation. I find the effect is strongest for (i) younger workers, (ii) males, (iii) construction and services industries, (iv) construction and transportation occupations, and (v) the increase of the injury rate after RML is mainly due to being struck and falling.

What do these estimates mean? With the average baseline monthly claims within the treated counties being approximately 643, the results imply that RML increases injuries by approximately 286 to 424 within the treated counties per year. According to Viscusi and Aldy (2003), the value of statistical injury ranged from \$20,000 to \$70,000 per injury in 2010 (Kniesner and Leeth, 2014). This translates into \$23,000 to \$80,000 in 2018 dollars. Hence, my estimates suggest that RML

[2]The traditional SCM has only one treatment unit, however, there are multiple high recreational marijuana exposure (treatment) counties in the Oregon. Hence, I extend the SCM into multiple treatment units, please read appendix section for formal description.

[3]Prior to December 2015, cities could also implement local bans if less than 45 percent of voters in their jurisdiction voted in favor of Measure 91 without referring the ordinance to the voter.

[4]As in other Regression Discontinuity Designs, the comparison between near "winners" and near "losers" potentially eliminates any confounding selection and omitted variable biases, and allows for credible and transparent estimates of the effect of RML on workplace injury rate.

increases work injury costs by roughly \$7 to \$34 million (or \$5 to \$24 per capita) per year in Oregon. Overall, it suggests that recreational marijuana legalization may come at the expense of workplace injuries.

In what follows, I provide additional background on marijuana policies in Oregon and variation in RML implementation across counties. Section 3 describes the data and construction of empirical settings, and Section 4 presents the research design and identification strategies. After which Section 5 presents result, and Section 6 concludes.

2 Background

Prior to the passage of the Marijuana Taxation Act of 1938, the consumption of marijuana for both recreational and medical purposes is legal.[5] In fact, cannabis was entered into the United States Pharmacopeia in 1850 as a treatment for pain, some infectious diseases, bleeding and other conditions.[6] The Controlled Substance Act of 1970 re-classified marijuana as a Schedule I substance along with heroin and methamphetamine, as a "high potential for abuse and little known medical benefit" drug.

Oregon became the first state to decriminalize the possession of small amounts of marijuana in 1973, although cultivation and distribution of the drug remained felony offenses.[7] In 1996, California became the first state to legalize marijuana for medical use. Currently, 33 states and Washington, D.C allow the cultivation, possession, and use of marijuana by doctor recommendation for patients with certain medical conditions (Oregon legalized medical marijuana use in 1998). Furthermore, since 2012, 11 states and the District of Columbia have legalized personal recreational marijuana use. Despite the increase in marijuana laws for a number of states, cannabis is still illegal under federal law.[8]

[5] https://www.nytimes.com/1854/01/10/archives/our-fashionable-narcotics.html.

[6] Throughout the paper, I will use the terms: marijuana and cannabis interchangeably as they convey the same meaning.

[7] https://www.civilized.life/articles/the-history-of-marijuana-in-oregon/.

[8] The federal government regulates drugs through the Controlled Substances Act (21 U.S.C. §811), which does not recognize the difference between medical and recreational use of cannabis.

Measure 91 was approved by the Oregon General Election on November 4th 2014. It has two main components. The first component is what I refer to as "demand side legalization", which took effect on July 1st, 2015. This allows non-medical cultivation and the possession of small amounts of marijuana for adults over the age of 21. Specifically, Oregonians are allowed to grow up to four plants on their property, possess up to eight ounces of usable marijuana in their homes and up to one ounce on their person.[9] The second component is what I refer to as "supply side legalization", which took effect on October 1st, 2015. This is when the commercial recreational marijuana sales market opened. It allows the manufacture, sale of marijuana by/to adults, subject to state licensing, regulation, and taxation. Specifically, in 2015, Oregon Governor Kate Brown signed an emergency bill declaring marijuana sales legal to recreational users from dispensaries starting October 1st, 2015. State officials began working on establishing a regulatory structure for the sales of marijuana and taxing of such sales, with the OLCC to oversee it. According to the Oregon Department of Revenue, state and local recreational marijuana sales/excise taxes generated $78 million in tax revenue in 2017, up from $60 million in 2016. Effective on January 1, 2017, dispensaries were no longer permitted to sell cannabis for recreational use unless they applied for and received OLCC license for such sales. As of March 1, 2019, there are 621 active marijuana retail licenses approved by OLCC.

However, even though RML is state policy, regulations are applied at a more local level. The law provides cities and counties the opportunity to prohibit recreational marijuana business in their jurisdiction. Counties that have less than 45 percent of voters who voted in favor the Measure 91 prohibited the establishment of Licensed Recreational Marijuana producers, processors, wholesalers, and/or retailers. Due to rounding, the final decision cutoff is 46%. There are 80 cities and 16 (out of 36) counties ban recreational marijuana businesses in Oregon.[10] In addition, prior to December 2015, cities could also implement local bans following the same vote share rules without referring the ordinance to the voter. Cities choose to opt-out should provide the OLCC

[9]For more information go to: www.whatslegaloregon.com.

[10]They are Baker, Crook, Gilliam, Grant, Harney, Jefferson, Klamath, Lake, Marion, Malheur, Morrow, Sherman, Umatilla, Union, Wallowa, and Wheeler Counties. Douglas county had 54.5% vote share who voted against (or 45.5% vote share who voted in favor of) marijuana business, it has been defeated since the final decision is rounded.

with a copy of their ordinance, signed and returned the official "Local Option Opt-Out" form found on *www.marijuana.oregon.gov* by December 27th, 2015. It is important to note that the vote share rules is not a mandate. For example, Marion County has more than 46% voters voted in favor of RML, however, they decide to opted out of marijuana business in their unincorporated land. Cities inside of Marion County, such as Salem, can still allow recreational marijuana sales.

Local bans lead to a substantial amount of variation in recreational marijuana sales and licensed recreational marijuana retailers across counties. As of February 2019, licensed recreational marijuana retailers varies from 0 to 173 across counties in Oregon, according to the OLCC. Moreover, up to December 2017, total recreational marijuana sales ranged from $0 to $183 million across counties in Oregon. Moreover, up to December 2017, 11 counties had no recreational marijuana sales,[11] but counties like Multnomah, Washington and Lane have sold more than $50 million recreational marijuana in total after RML.

As mentioned earlier, this paper will focus on the recreational marijuana sales legalization. One is because supply-side interventions often dominate the discussion surrounding drug policy. The other is because the scope of demand side legalization's impact on marijuana usage is limited. Specifically, the demand side law only allows for the small possession and use of marijuana at home per household. Lastly, there is a time lag in the demand side legalization. Even the Measure 91 was approved in the end of 2014, but did not enter into force until July 2015. Moreover, by October 2015, the Legislature passed four bills, which made comprehensive reforms to Measure 91 and addressed issues of local control, taxation, and early sales. Therefore, by exploiting this supply-side regulation on recreational marijuana facilities across counties, it arguably better approximates how market participants actually interact with the law in most cases than the variation induced by relatively infrequent action by state legislatures.

[11] They are Gilliam, Crook, Wheeler, Grant, Wallowa, Union, Lake, Sherman, Malheur, Klamath and Morrow Counties.

3 Data and Empirical Setting

3.1 Workplace Injury Rate

The main data for this paper comes from the Oregon Department of Consumer and Business Services, Workers' Compensation Division. It is restricted-use accepted disabling workers' compensation claims from 2013-2017. These are work-related, where the worker gets medical care and misses more than 3 days of work.[12] For each claim, the data contains information on claimant's demographic characteristics and information on the date of injury, county of injury, claimant's occupation, industry, job tenure and detailed information on the injury source, and the body part injured. It also has detailed employer's information such as employer's name, address, and whether the employer company is publicly or privately owned.

Table 1 describes the composition of workers' compensation data in the sample. In order to create a monthly injury rate data for each county, I collect employment number from the American Community Survey (ACS) 2010 5-year estimates. Hence, for each month and year, and each county, I calculate the injury rate as the injury number per 1000 employment. Mathematically, $InjuryRate_{jt} = \frac{\#ofInjuries_{jt}}{\#ofEmployment_j}$, for each county j and time t.

3.2 Measure 91 Vote Shares

The election statistics are from the Oregon Secretary of State, Elections Division.[13] It has the official results of the November 4^{th}, 2014 General Election on Measure 91. As mentioned earlier, counties that have less than 45% voters who voted *in favor of* the Measure 91 have prohibited the establishment of licensed recreational marijuana producers, processors, wholesalers, and/or retailers. Since the final decision is rounded, the vote share cutoff of RML is actually at 46%. The vote share is calculated by number of votes in favor of RML divided by total vote counts within

[12] Insurers are not required to report any claim that does not meet all of these requirements. Hence, an employee who gets injured on the job and only misses work for a medical appointment related to the injury, the department would not have that claim on record.

[13] https://sos.oregon.gov/elections/Pages/electionhistory.aspx.

a county. Figure 3 shows the geographic variation in the vote share (panel A) and the vote share cutoff 46% (panel B).

3.3 Licensed Recreational Marijuana Retailers and Sales

The data of monthly administrative marijuana sales records are from the "traceability" system maintained by OLCC. The system was designed to track each step in the marijuana supply chain, enabling state officials to collect taxes and enforce regulations. To ensure accurate data, OLCC employees conduct random in person audits. Violators face penalties that include inventory seizure and destruction. The data set has monthly county level recreational marijuana sales in dollars by product type and quantity of sales. The tracking system records the date, quantity and price of the transaction and generates a unique identifier. The system tracks individual retail sales, which allows me to link the prices, quantities and transaction time of each sale to the product characteristics.

The approved licensed recreational marijuana retailers' information are also obtained from the OLCC. It has each retailer's geographic location, company name, whether it provide delivery services, etc.[14]

3.4 Recreational Marijuana Exposure Measures

I construct three measures of recreational marijuana exposure for each county in Oregon. The first measure (called *sales-per-capita*) is defined as the total sales of recreational marijuana divided by the population in the county, where population is also from the ACS 2010 5-year estimates. This measure serves as a proxy for the interaction of actual demand and supply of recreational marijuana. Mathematically, $SalesPerCapita_j = \frac{TotalSales_j}{Population_j}$, for each county j.[15] Since cities also can ban marijuana sales following the same vote share rules, I create the second marijuana exposure measure (called *PopShare*) defined as the population share lives in the jurisdictions that allow recreational marijuana sales within a county. I obtain the list of cities or counties that have

[14]For more information , go to: https://www.oregon.gov/olcc/marijuana/Pages/default.aspx.

[15]The OLCC starts to collect administrative sales data from July 2016, and my analysis period ends at 2017. Hence, the total sales in the formula is the sum of the sales (in dollars) from July 2016 to December 2017.

prohibited the establishment of licensed recreational marijuana producers, processors, wholesalers, and/or retailers from the OLCC. Within a county, I add the city population that allows marijuana sales divided by the county population. Mathematically, for each city i within county j,

$$PopShare_j = \frac{\sum_{i \in j}(CityPopulation_i * Allow_i)}{CountyPopulation_j}, \quad (1)$$

where $Allow_i$ is an indicator that equals to 1 if city i allows recreational marijuana sales of county j, and zero otherwise. Lastly, I create the *fraction* measure that interacts the supply and latent demand for recreational marijuana in each county. I collect licensed recreational marijuana retailers' city location from the OLCC. Within a county, I add the city population that has licensed recreational marijuana retailers divided by the county population. This generates a population share that has access to local recreational marijuana store(s). Lastly, I multiply the population share by the baseline year's (2013) workplace marijuana drug positivity rate collected from Quest Diagnostics.[16] Mathematically, for each county j and city i that have access to licensed recreational marijuana retailer(s),

$$Fraction_j = \frac{\sum_{i \in j}(CityPopulation_i * \mathbb{1}[\#ofRetailer_i > 0])}{CountyPopulation_j} * PositivityRate_{j,baseline}, \quad (2)$$

where $\mathbb{1}[.]$ is an indicator that equals to 1 if the number of retailers in the city i of county j is positive, and zero otherwise.

Table 2 shows the summary statistics of the main analysis sample. The top panel consists of monthly injury rate data between 2013-2017, for all 36 counties in the Oregon. The bottom panel reports the recreational marijuana exposure measures' mean and standard deviation. I define high marijuana exposure (treatment) counties as those who have marijuana exposure measures above certain threshold.[17] Figure 4 shows the geographic variation in the recreational marijuana exposure measures in Oregon described above. Panel A-C are *sales-per-capita*, *fraction* and *PopShare*

[16] https://www.questdiagnostics.com/home/physicians/health-trends/drug-testing.

[17] Specifically, I define high exposure if (i) sales-per-capita or fraction measures are above the 75$^{\text{th}}$ percentile; (ii) Population share measure is above the 50$^{\text{th}}$ percentile.

measures, respectively.

4 Empirical Strategy

The main purpose of this paper is to identify the causal impact of RML on workplace injuries. Examining this link empirically is difficult because potential unobserved heterogeneity across states/counties both from the altitude towards legalization and workplace safety complicates the identification of the causal effects. As a direct consequence, standard regression techniques lead to biased and inconsistent coefficient estimates. In this paper, I address these endogeneity issues by using following research designs.

4.1 Difference-in-Difference Model

I first estimate the impact of RML on workplace injury rate with the following "native" difference-in-differences (DiD) regression model.

$$\begin{aligned} InjuryRate_{jt} = \alpha + \beta_1 \mathbb{1}(X_j \geq 0.46) * After_t + \beta_2 \mathbb{1}(X_j \geq 0.46) + \beta_3 After_t \\ + \delta_j + \theta_t + X_{jt} + \varepsilon_{jt}. \end{aligned} \tag{3}$$

Where $InjuryRate_{jt}$ represents injury number per 1000 employment for county j in time t. $\mathbb{1}(X_j \geq 0.46)$ is the indicator for being above the threshold 46 percent vote share in favor of RML. $After_t$ is a dummy variable that corresponds to all periods starting from the October 2015 when recreational sales became legalized. δ_j and θ_t account for county, time (year, month, and year-by-month) fixed effects, respectively. X_{jt} is a vector of time-varing county characteristics such as the unemployment rate and median income. ε_{jt} is the error term. The standard errors are clustered around the county level. Thus, the coefficient β_1 is the parameter of interests.

However, there is a lot more variation across counties as shown in Figure 4. Hence, the preferred specifications are using the constructed marijuana exposure measures to estimate the impact of RML on workplace injury rate with the following more "advanced" DiD regression model.[18]

$$\begin{aligned} InjuryRate_{jt} = \alpha + \beta_1 HighExposure_j * After_t + \beta_2 HighExposure_j + \beta_3 After_t \\ + \delta_j + \theta_t + X_{jt} + \varepsilon_{jt}. \end{aligned} \tag{4}$$

Where the variables are the same as equation (3) except for $HighExposure_j$, which is a dummy variable that takes a value of 1 when counties who have marijuana exposure measures above the certain threshold, zero otherwise. Thus, the coefficient β_1 provides a reduced form estimate of the causal effect of RML on workplace injury rate.

The fundamental identifying assumption in the DiD framework is that, in the absence of treatment, the average injury rate for the treated and control counties would have followed parallel trends over time. This pretrend graph is presented in the result section. The assumption allows the averages of the time-invariant unobserved variables to differ between treated and control groups. However, in the RML settings, the parallel trends assumption may be implausible if unobserved confounders, such as rurality, have time-varying effects on the injury rate.

Synthetic Control Method

An alternative strategy is to employ a data-driven search for a comparison group based on pre-RML injury rates and trends, known as the Synthetic Control Method (SCM).[19] In contrast to DiD, the synthetic control approach moves away from a simple average of control units, and instead uses a weighted average of the set of controls. In other words, instead of choosing between low recreational marijuana exposure counties, or taking a simple average of injury rate in those counties, the synthetic control approach chooses weights for each of the those low exposure counties so that the weighted average best approximates pre-RML injury rate and other county characteristics of

[18] As mentioned above, three recreational marijuana exposure measures were used to capture the differential recreational marijuana usage after RML across counties in Oregon. I only report the *sales-per-capita* measure in the result section, the other measure results can be found in the appendix.

[19] See Abadie and Gardeazabal (2003); Abadie et al. (2010) and Abadie et al. (2015) for detailed method explanation.

the treated county. It formally identifies the control group called the “synthetic control unit”. [20]

In stead of DiD’s parallel trend assumption, the SCM allows the effects of observed and unobserved predictors of the injury rate to change over time (such as urban vs. rural differential trends), while assuming the pre-intervention covariates have a linear relationship with injury rate post-treatment. The traditional SCM has only one treatment unit, however, there are multiple high recreational marijuana exposure (treatment) counties in the Oregon. Hence, I extend the SCM into multiple treatment units, please read appendix section for formal description.

4.2 Regression Discontinuity Design

Last, I utilizing regression discontinuity design (RD) by exploiting the discontinuity in marijuana exposure generated by RML vote share rules. The county treatment status is determined by the running variable.[21] If the running variable crosses the vote share cutoff c (in this case, is 46% in favor of RML), treatment is (partially) switched on or off. Validity of this design is that counties are unable to precisely manipulate the running variable near the vote share cutoff and therefore randomly assigning them into a treatment ad control group (Lee and Lemieux, 2010).

The measure 91 vote share rule is not mandatory for the local government. Hence, the discontinuity in the probability of (high) exposure is smaller than 100% (“imperfect compliance”). This setup is naturally lead to a fuzzy RD design where the running variable is the county-specific vote share *in favor of* RML denoted as X_j that partially determines (high) recreational marijuana exposure. Formally, I estimate the following equations:

$$InjuryRate_{jt} = \theta + \theta_1 f(\tilde{X_j}) + \kappa_0 D_j + \pi_1 D_j g(\tilde{X_j}) + \xi_{jt}. \tag{5}$$

$$D_j = \gamma + \gamma_1 f(\tilde{X_j}) + \pi_0 T_j + \pi_1 T_j g(\tilde{X_j}) + \mu_j. \tag{6}$$

[20] Abadie et al. (2010) argues that this method can be used to reduce any potential endogeneity problem caused by omitted variables. If a synthetic county can be found such that it matches the pre-RML trajectory of the injury rate for the treated county, then the size of the bias caused by time varying unobserved confounders in the difference between the post-RML injury rate for the treated and the synthetic control county goes towards to zero as the pre-RML period increases.

[21] This also referred to in the literature as the “forcing” or “assignment” variable.

Where $InjuryRate_{jt}$ is injury rate for county j and time (month, year) t. D_j is a high exposure indicator that takes a value of 1 when counties who have marijuana exposure measures certain threshold, zero otherwise. $\tilde{X}_j = (X_j - 0.46)$ is vote share centered at the 46%. I instrument high exposure indicator using the RML vote share cutoff $T_j = \mathbb{1}(\tilde{X}_j \geq 0)$ which equals one if county vote share exceed the official 46% cutoff for marijuana sales legalization. $f(\tilde{X}_j)$ and $g(\tilde{X}_j)$ are flexible controls of the vote share, allowed to be difference above and below the 46% cutoff. The model also includes county and time fixed effects. Finally, ε_{jt}, μ_j are idiosyncratic error terms.

Eq.(5) shows that the treatment effect of RML captured by the parameter κ_0. Eq.(6) is classical first-stage equation linking the endogenous treatment variable D_j to the respective set of exogenous variables and the instrument. As discussed in Lee and Lemieux (2010), there are two ways to estimate the discontinuity parameters in Eq.(5) and (6). First, one can capture the running variable using a parametric function and use all of the available data to estimate these questions via ordinary least squires, typically referred to as the global polynomial approach. Second, one can capture vote share via a linear function of X_j and estimate the equation over a narrow range of data, typically referred to as the local linear approach. The preferred estimation method is the local linear regression approach.[22]

Specifically, the model is estimated using local linear regression (Gelman and Imbens, 2018) within the optimal bandwidth on either side of the cutoff using triangular kernel weights.[23] This avoids the problem of identifying local effects using variation too far away from the passing threshold. The choice of bandwidth is motivated by graphical fit, data-driven optimal bandwidth selectors. I use a Mean Squared Error (MSE)-optimal bandwidth calculated using the procedure suggested by Calonico et al. (2014), to predict the optimal bandwidths. The advantage of estimating the model non-parametrically is that there is no need to specify functional form of $f(X_j)$, as if the functional form is specified incorrectly, the estimates are likely to be biased.

The main effect is estimated one year after treatment (2017). As a falsification test, I check that

[22] The global polynomial approach and results are also presented in the appendix.

[23] The triangular kernel puts more weight on observations closer to the cutoff, as this provides the optimal boundary correction.

there are no pre-existing discontinuities in the outcomes by estimating the same specification on one year before treatment (2013) observations. This is presented in the result section.

5 Results

5.1 Main Results

I begin by presenting the results from the "naive" DiD regression in Table 3. It shows that after RML, counties who pass the 46% vote share threshold increases injury rate by 0.019 per 1000 employment relative to counties who did not pass. As I described earlier, this regression gives up a lot more variation in marijuana exposure captured by the three constructed measures across counties. Hence, I move to show the results from the "advanced" DiD model. Table 4 shows results from Eq.(4) that analyzes the effect of RML on workplace injury rate. Going from column 1 to Column 3 shows an increase in the estimated coefficient after controlling for county-by-year fixed effects, likely due to differences in economic conditions in high exposure versus low exposure counties. It shows that high marijuana exposure counties increase their injury rates by 0.034 to 0.050 per 1000 employment relative to low marijuana exposure counties after RML. The average monthly injury rate in the sample is 0.913 per 1000 employment, so the effects translate into 3.7%-5.5% increase in the injury rate.[24]

What do the estimates mean? Let us talk about how costly the injuries are in the United States first. Nationally, the total cost of workplace injuries in 2017 was $161.5 billion,[25] and 70,000,000 productive days lost due to injuries in the same year (National Safety Council).[26] Moreover, according to Viscusi and Aldy (2003), the value of statistical injury ranged from $20,000 to

[24]The *fraction* measure results are shown in the appendix **??**, the results are consistent with the *sale-per-capita* measure.

[25]This includes wage and productivity losses of $50.7 billion, medical expenses of $34.3 billion and administrative expenses of $52.0 billion. This total also includes employers' uninsured costs of $12.4 billion, including the value of time lost by workers other than those with disabling injuries who are directly or indirectly involved in injuries, and the cost of time required to investigate injuries, write up injury reports and so forth. The total also includes damage to motor vehicles in work-related injuries of $4.9 billion and fire losses of $7.3 billion.

[26]This estimate includes the actual time lost during the year from disabling injuries, but excludes time lost on the day of the injury, time required for further medical treatment or check-ups following the injured person's return to work.

$70,000 per injury in 2010 (Kniesner and Leeth, 2014). This translates into $23,000 to $80,000 in 2018 dollars.

In Oregon, before RML in my sample (Jan.2013-Sep.2015), the sum of the injury claims within the treated counties is 21,209, hence, the average baseline monthly claims in the same group is approximately 643(=21,209÷33 month). My results imply that RML increases injuries approximately by 285.5 (=643*3.7%*12month) to 424.4(=643*5.5%*12month) within the treated counties per year. Therefore, my estimates suggest that RML increases work injury costs by roughly $7(=285.5*$23,000) to $34(=424.4*$80,000) million per year. With the sum of the treated counties population is 1,413,949, my results indicate that RML increases injury costs by $5(=7÷ 1.4 million) to $24(=34÷1.4 million) per capita per year.

The biggest concern of my DiD method is whether the results satisfy parallel trends assumption. In another words, my results are only valid if the injury rate are attributed to changes that occurred in the RML implementation period and not to pre-existing trends. Figure 5 reports this assumption graphically by showing the mean variations of each month injury rate across treatment and control group. They show the difference in means remaining relatively stable pre-RML, the difference substantially increase after RML.

Heterogeneous Effects

The work injury costs also vary by workers' characteristics (such as age or gender), occupation/industry and cause of injury. Studies show that the nonfatal on-the-job injury rate in construction industry was 71% higher than that for all industries (Bureau of Labor Statistics, 2005). Five construction industries accounted for over half the total fatal and nonfatal injury costs.[27] Waehrer et al. (2007) show that the total costs of fatal and nonfatal injuries in the construction industry were estimated to be $11.5 billion in 2002 ($16.2 billion in 2015 dollars), which constitutes 15% of the costs for all private industries. They also calculate the average cost per case of fatal or nonfatal

[27] They are miscellaneous special trade contractors; plumbing; heating and air-conditioning; electrical work; heavy construction except highway; and residential building construction.

injury is $27,000 for construction in 2002 ($38,043 in 2015 dollars[28]), almost double the per-case cost for all industries.

In terms of cause of injury, the top three costly lost-time Workers' Compensation claims are caused from motor vehicle crashes ($73,599), burns ($49,107) and falls or slips ($46,297) in 2015 and 2016 (National Safety Council).[29] Study also estimated the average Workers' Compensation claims costs for falls in construction industry (Occupational Safety and Health Administration, 2012). For example, falls from elevations by roofers cost approximately $106,000 per claim and falls from elevations by carpenters cost over $97,000 per claim.

Hence, in order to better understand the economic costs of RML on workplace injuries, it is important to discover the dis-aggregate effects of RML. The restricted-use Oregon Workers' Compensation claims data allows for the assessment of those heterogeneous effects of RML on age, gender, industry, occupation, ownership and source of injury using the DiD model in Eq.(4). Table 5 to Table 11 show the results to see if particular subsamples drives the positive results from the main analysis.[30]

Table 5 Column 1 to Column 7 show the results by age group. It shows that RML's impact mainly on age 18-34 and over 65 workers. Specifically, RML increases injury rate among 18-24 years old workers for 0.006 per 1000 employment in high exposure counties relative to low exposure counties. The mean injury rate for this age group is 0.096 per 1000 employment, so the effect translate into a 6.3 percent increase. This is consistent with the evidence about marijuana usage age groups. According to the annual National Survey on Drug Use and Health (2013), illicit drug use is highest among people in their late teens and twenties, and it is increasing among people in their sixties.[31]

Table 6 shows the results separately by gender. I find that male has the most impact from RML. This result is also consistent with the evidence that men are more likely than women to use almost

[28]This only takes into account for the Consumer Price Index (CPI) inflation rate, without considering the drug price and other related medical treatment price change.

[29]https://injuryfacts.nsc.org/work/costs/workers-compensation-costs/.

[30]Table A1 to Table A6 show the subsample results using the *fraction* measure, respectively.

[31]https://www.drugabuse.gov/publications/drugfacts/nationwide-trends.

all types of illicit drugs (National Institute on Drug Abuse). In addition, men have higher injury frequency than women, which is often correlated with their concentrated in sectors that have high contact injury rates. Specifically, Column 2 shows that RML increases injury rate among male for 0.024 per 1000 employment in high exposure counties relative to low exposure counties. The mean injury rate for this age group is 0.612 per 1000 employment, so the effect translate into a 3.9 percent increase.

Table 7 and Table 8 show the heterogeneity effects of RML by industry and occupation. They show that industries such as construction and services have the most impact from RML. Specifically, construction industry increases 9.5 percent after RML. Moreover, construction related and transportation, and material moving are the most impacted occupations. For example, construction related occupation increases 8.2 percent, transportation related occupation increases 8.4 percent after RML. These results are consistent with the evidence that the construction industry is one of the most hazardous industries, and marijuana usage was rated the second most common drug abuse among construction workers (Fardhosseini and Esmaeili, 2016). Literature also suggests that marijuana use is associated with a significantly increased risk of being involved in motor vehicle crashes (e.g., Blows et al., 2005; Li et al., 2011).

Table 9 presents the RML's impact by firm ownership. It shows that private firms have the dominant effect from RML. This makes sense since the government employees are often require to stay drug-free in the workplace since marijuana use is still illegal in the federal level (Drug-Free Workplace Act of 1988).

Finally, Table 11 shows the estimates of RML's impact on the source of injury. Each claim has the detailed description on how the worker gets injured for the insurance purpose. The results indicate that the main source of injury after RML are workers who are strucked or fall. Specifically, falling increases 10.5% after RML. While the research on the contribution of psychoactive drugs to fall risk is limited, literature has shown that marijuana can have negative effects on attention, memory, and learning that may last for days or weeks after the acute effects subside. These could be the potential mechanism behind the result of the increase in fall.[32]

SCM Results

An an alternative strategy to construct suitable comparison group, Figure 6 presents graphical results from the SCM. The top panel shows that pre-intervention the injury rate trend for the synthetic control group closely matches the corresponding trend in treated counties. Since it is the high-frequency (monthly) data, I use moving average method to smooth the injury rates. The average pre-intervention difference between treated and synthetic treated is almost 0. In the post-intervention period, Figure 6 (bottom panel) reveals a sizable gap between treated and synthetic treated group.

Table 11 (panel A) presents estimates of the DiD estimators in Synthetic Control Method. For each outcome, the first two columns present the average differences between treated and the synthetic control for pre-intervention periods (January 2013 to September 2015) and post-intervention periods (October 2015 to December 2017). The remaining columns present DiD estimates of the injury rate effect of RML, the p-value from a one-tail test of the likelihood of observing an estimate at least as positive as that for treated county, and the mean of the injury rate. The results report nearly zero pre-intervention differential between treated and synthetic treated and then widen considerably in the post-intervention period. One can use the DiD estimates to calculate the net increase in injury rate caused by the passage and implementation of RML. Specifically, RML increases 0.03 injuries per 1000 employment with the mean injury rate is 0.95 per 1000 employment. This result is consistent with the base case DiD estimates (panel C).

Figure 7 displays the raw data needed to conduct the permutation test of the significance of the relative increase in the treated unit. Specifically, for each of the donor counties as well as the treated county, the figure displays the month-by-month difference between the injury rate for all counties. The differences for each of the donor counties are displayed with the gray lines, while the

[32] Sterke et al. (2008) conduct a systematic review of the literature to investigate which psychoactive drugs increase fall risk. The scarce evidence shows that multiple psychoactive drugs increase fall risk in nursing home populations with residents with dementia.

difference for treated is displayed by the orange line. During the pre-intervention period, January 2013 through September 2015, the treated county data points clearly lie within the distribution of placebo estimates, suggesting that population weighted treatment county is not an outlier during this period. In the post-intervention periods, the treated differences move to the upper of the distribution, it became a visible outlier.

Robustness Checks

Synthetic control method at the county level I also estimate treatment effects for each county at the treatment group, and aggregate the ATT using population weights of the treated counties.[33] Table 11 (panel B) presents the result using the county level SCM. I find that the conclusions are consistent with those from the base case analysis. The somewhat different point estimates can be explained with the worse pre-treatment fit obtained with the county level analysis. Overall, my base case results are robust across SCM and DiD, as shown in Table 11.

Leave-one-out To examine whether the synthetic control results were driven by a few influential control counties, I also assess sensitivity of the estimated treatment effects to the iterative deletion of counties from the donor pool (Abadie et al., 2015). Accounting for the structure of the data, I estimate 12 new synthetic counties, in each of these analysis deleting one of the control. I find the results are robust to the elimination of one control county at a time.

5.2 RD Results

I begin by showing graphical representation of the first-stage effects. Figure 8 shows the population share in a county that allows marijuana sales by vote share in favor of RML. The vote share is centered at 46% threshold. The scatter plots are overlaid with linear fit on both sides of the discontinuity. The graph clearly indicate that above the official vote share threshold, the share of population lives in the county that allows marijuana sales increases sharply indicating a discontinuous jump in the

[33] Please see Appendix for detailed extension on SCM with multiple treatment units.

probability of being high exposure counties.[34]

To show the causal impact of RML on workplace injuries, I start by showing descriptively how the injury rate, defined as # of injuries per 1000 employment, changed over the period from 2013 to 2017 for counties that were under recreational marijuana sales legalization vote share cutoff and for counties that were not. Figure 9 shows the raw county yearly injury rate in the sample one year before (2013, top panel) and one year after (2017, bottom panel) by vote share, with overlaid linear fitted lines.[35] During the pre-treatment period, there is no visible discontinuity around the cutoff, but over the post-treatment period, counties who pass the sales legalization (right of the cutoff) are experiencing a discontinuous increase.

Figure 10 shows the visual equivalent of the RD estimates for the injury rate. The panel on the top shows the falsification RD graph estimated on the sample of pre-treatment periods, while the one on the bottom shows the main RD graph of interest, estimated on post-treatment periods. The points show the average value of the outcome for different bins of the running variable. The line plots the fit from a locally linear regression estimated separately on each side of the discontinuity. The RD graphs show that there was no difference in the injury rate at the discontinuity in the pre-treatment sample. However, after the legalization, counties just above the threshold had a higher injury rate than those just below.

The regression estimates confirm the results. Table 12 presents the effect of sales legalization for a 100%, 150%, 175% and 200% of MSE-optimal bandwidth separately for the pre-treatment sample (Column 1 to 4) and for the post-treatment sample (column 5 to 8). Panel A shows the estimation without covariates, and panel B presents the estimation with county, year and month fixed effects. As shown, there was no statistically significant difference in the injury rate in the pre-treatment, but counties above the threshold had a higher injury rate in the post-treatment with respect to those below. The estimates are robust to the different bandwidth.

The magnitude of the effect is large: looking at the estimates for places within a 0.054 bandwidth

[34] High exposure is an indicator that equals to 1 if county's population share measure is at 50 percentile and above.

[35] In Oregon, Measure 91 was passed in 2014, sales legalization starts at 2015, and at the end of 2016, their is another round of election on marijuana sales legalization. So the pre-period is defined at 2013, and post is defined at 2017.

from the threshold, the coefficient shows a 20%(=0.182/0.907) increase in the injury rate for treated counties after the RML was introduced. To the extent the unobservables vary continuously at the threshold and there are no pre-treatment differences in the socio-economic composition of control and treated counties, the effect is unlikely to be explained by other external factors, suggesting that indeed the result must be explained by sales legalization. One big caveat is that my point estimates in the RD setting lost a lot of precision due to lack of power around the cutoff. However, I check the robustness in the following subsections and the results are consistent. Nonetheless, the lack of power makes the results somewhat less reliable.

RD Validity Checks

The main identification assumption underlying a RD design is that all determinants of future outcomes vary smoothly across the threshold. In that sense, any observed discontinuity at the threshold can be attributed to the causal effect of passing RML. As a consequence of this assumption, all observed and unobserved characteristics should be balanced around the cutoff and treatment is "as good as randomly assigned". In other words, counties below the vote share cutoff represent a valid control group for those just above the threshold and any comparison between groups reveals the local causal effects of interest.

As a first validity check for local random assignment, I investigate the density of the running variable. Figure A1 shows the histogram of vote shares with a normal density estimate. Inspecting the density graph suggests no manipulation of the running variable since it appears to the smooth around the threshold thus reinforcing the validity of the RD approach used in this paper.[36]

Furthermore, I compare predetermined county characteristic just below and above the vote share cutoff to see whether they are locally balanced around the cutoff. In fact, if treatment is locally randomized then counties around the threshold should not differ substantially in observable and unobservable characteristics. Figure A2 shows the scatter plots of education, share of male and

[36]In addition, I conducted the formal McCrary (2008) test which also provides no evidence for a significant discontinuity in the distribution of vote share at the 46 percent threshold.

marital status overlaid with local polynomial smooth plots around the vote share cutoff. The graphs clearly indicate no significant discontinuity in any of the baseline covariates at the threshold pointing towards local random assignment. Overall, the RD validity checks support my identification strategy and provide no evidence for violations of the key underlying assumptions.

RD Robustness Checks

In this section I show that my results are robust to a number of potential concerns. Figure 11 shows the coefficient for the dummy for being above the threshold, together with 95% confidence intervals, for different samples, specifications, and estimation techniques. The relevant comparison is whether each coefficient is difference than the one estimated using the baseline specification at the top of each graph.

Robustness to sample restrictions and specifications. Figure 11 panel A shows that the main result is robust to using different sample restrictions and different specifications. First, the result does not change is I drop population outliers (counties that has lower than 10 percentile population in Oregon). Second, the result is robust to controlling for baseline county characteristics, such as percentage of male, education and marriage status. Third, clustering standard errors at the county-year level to allow errors to be correlated at the county-specific time does not make a difference.

Robustness to estimation. Figure 11 panel B shows that the specific estimation technique used is robust to the alternatives. First, I show robustness to using a triangular and an Epanechnikow kernels. The main result is not affected, although coefficient is large in magnitude using Epanechnikow kernel. Second, I estimate the main specification using locally quadratic regression with a triangular kernel. The result is not quite robust to using a locally quadratic regression. This is in line with Gelman and Imbens (2018), which suggests that RD estimates that are using higher order polynomials of the forcing variable, are sensitive and conventional inference tends to perform poorly in these settings.

6 Conclusion

The landscape of marijuana policies is changing rapidly, this has led to a heated discussion in its impact on social, economic, and public health outcomes, both positive and negative. This paper estimates the effect of recreational marijuana sales legalization on workplace injuries. Using administrative workers' compensation claims as a proxy for workplace injury rate, I answer this question by exploiting variations in the county level implementation of recreational marijuana law in Oregon. Cities and counties could implement local bans if less than 45% voters in their jurisdiction voted in favor of RML. This leads to a substantial amount of variation in recreational marijuana sales and licensed retail stores across counties.

Using three different empirical strategies, I find that workplace injury rate is approximately 5%-20% higher for treated relative to control counties post-RML. With the average baseline monthly claims within the treated counties being approximately 643, the results imply that RML increases injuries by approximately 286 to 424 within the treated counties per year. According to Viscusi and Aldy (2003), the value of statistical injury ranged from $20,000 to $70,000 per injury in 2010 (Kniesner and Leeth, 2014). This translates into $23,000 to $80,000 in 2018 dollars. Hence, my estimates suggest that RML increases work injury costs roughly by $7 to $34 million (or $5 to $24 per capita) per year in Oregon.

This paper has two main contributions: First, it uses several research designs to correct for standard endogeneity issues to identify the causal impact of recreational marijuana sales legalization. Second, it is the first study to analyze recreational marijuana legalization on workplace injury rate, highlighting several important heterogeneity effect across gender, age group, occupation, etc.. From a policy perspective, the findings in this study have important implications that suggesting legalizing recreational marijuana sales may come at the expense of workplace safety.

References

Abadie, A., Diamond, A., and 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., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

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

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, 121(2):368–391.

Auld, M. C. (2005). Smoking, drinking, and income. *Journal of Human Resources*, 40(2):505–518.

Blows, S., Ivers, R. Q., Connor, J., Ameratunga, S., Woodward, M., and Norton, R. (2005). Marijuana use and car crash injury. *Addiction*, 100(5):605–611.

Bohn, S., Lofstrom, M., and Raphael, S. (2014). Did the 2007 legal arizona workers act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics*, 96(2):258–269.

Buchmueller, T. C., DiNardo, J., and Valletta, R. G. (2011). The effect of an employer health insurance mandate on health insurance coverage and the demand for labor: Evidence from hawaii. *American Economic Journal: Economic Policy*, 3(4):25–51.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Cawley, J. and Ruhm, C. J. (2011). The economics of risky health behaviors. In *Handbook of health economics*, volume 2, pages 95–199. Elsevier.

Dave, D. and Kaestner, R. (2002). Alcohol taxes and labor market outcomes. *Journal of Health Economics*, 21(3):357–371.

DeSimone, J. (2002). Illegal drug use and employment. *Journal of Labor Economics*, 20(4):952–977.

Dube, A. and Zipperer, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies, iza discussion papers 8944. *Institute for the Study of Labor (IZA). URL: https://ideas. repec. org/p/iza/izadps/dp8944. html*.

Fardhosseini, M. S. and Esmaeili, B. (2016). The impact of the legalization of recreational marijuana on construction safety. In *Construction Research Congress 2016*, pages 2972–2983.

Gelman, A. and Imbens, G. (2018). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, pages 1–10.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy*, 80(2):223–255.

Hall, W. (2015). What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction*, 110(1):19–35.

Hall, W. and Degenhardt, L. (2009). Adverse health effects of non-medical cannabis use. *The Lancet*, 374(9698):1383–1391.

Kaestner, R. and Grossman, M. (1995). Wages, workers' compensation benefits, and drug use: indirect evidence of the effect of drugs on workplace accidents. *The American economic review*, 85(2):55–60.

Kaestner, R. and Grossman, M. (1998). The effect of drug use on workplace accidents. *Labour Economics*, 5(3):267–294.

Kniesner, T. J. and Leeth, J. D. (2014). Regulating occupational and product risks. In *Handbook of the Economics of Risk and Uncertainty*, volume 1, pages 493–600. Elsevier.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., and Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics*, 25(12):1514–1528.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.

Leigh, J. P. (2011). Economic burden of occupational injury and illness in the united states. *The Milbank Quarterly*, 89(4):728–772.

Li, M.-C., Brady, J. E., DiMaggio, C. J., Lusardi, A. R., Tzong, K. Y., and Li, G. (2011). Marijuana use and motor vehicle crashes. *Epidemiologic reviews*, 34(1):65–72.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714.

Ohsfeldt, R. L. and Morrisey, M. A. (1997). Beer taxes, workers' compensation, and industrial injury. *Review of Economics and Statistics*, 79(1):155–160.

Sterke, C. S., Verhagen, A. P., Van Beeck, E. F., and van der Cammen, T. J. (2008). The influence of drug use on fall incidents among nursing home residents: a systematic review. *International Psychogeriatrics*, 20(5):890–910.

Viscusi, W. K. and Aldy, J. E. (2003). The value of a statistical life: a critical review of market estimates throughout the world. *Journal of risk and uncertainty*, 27(1):5–76.

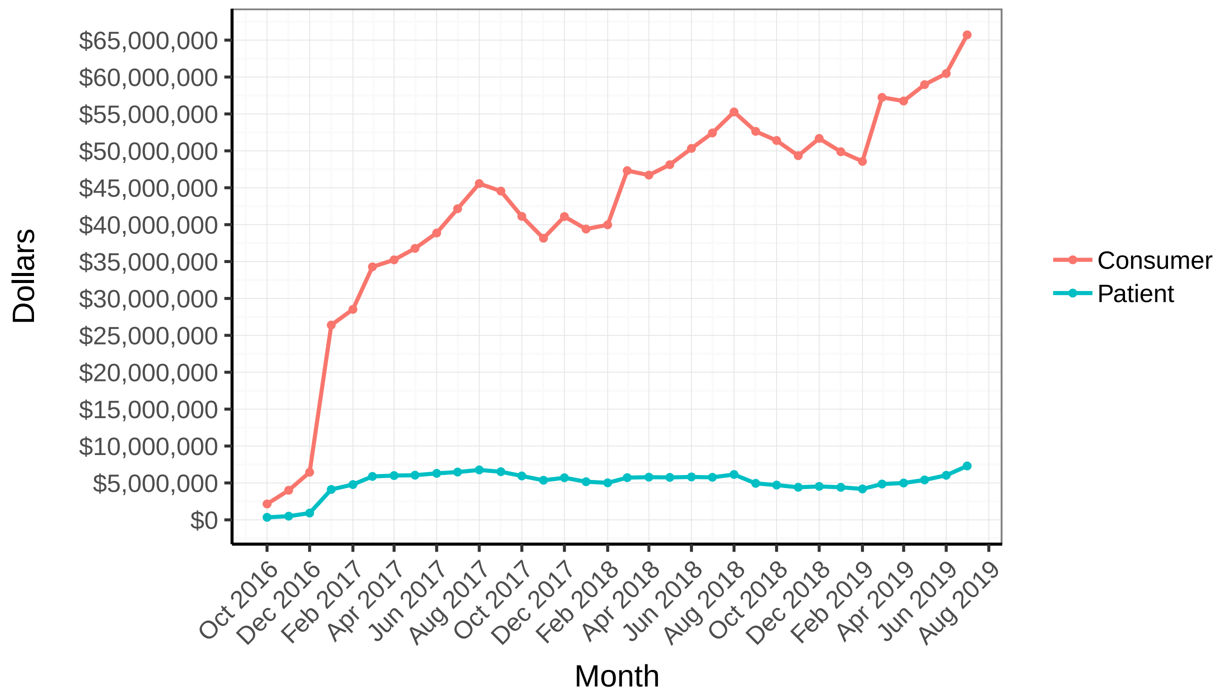
Volkow, N. D., Baler, R. D., Compton, W. M., and Weiss, S. R. (2014). Adverse health effects of marijuana use. *New England Journal of Medicine*, 370(23):2219–2227.

Waehrer, G. M., Dong, X. S., Miller, T., Haile, E., and Men, Y. (2007). Costs of occupational injuries in construction in the united states. *Accident Analysis & Prevention*, 39(6):1258–1266.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

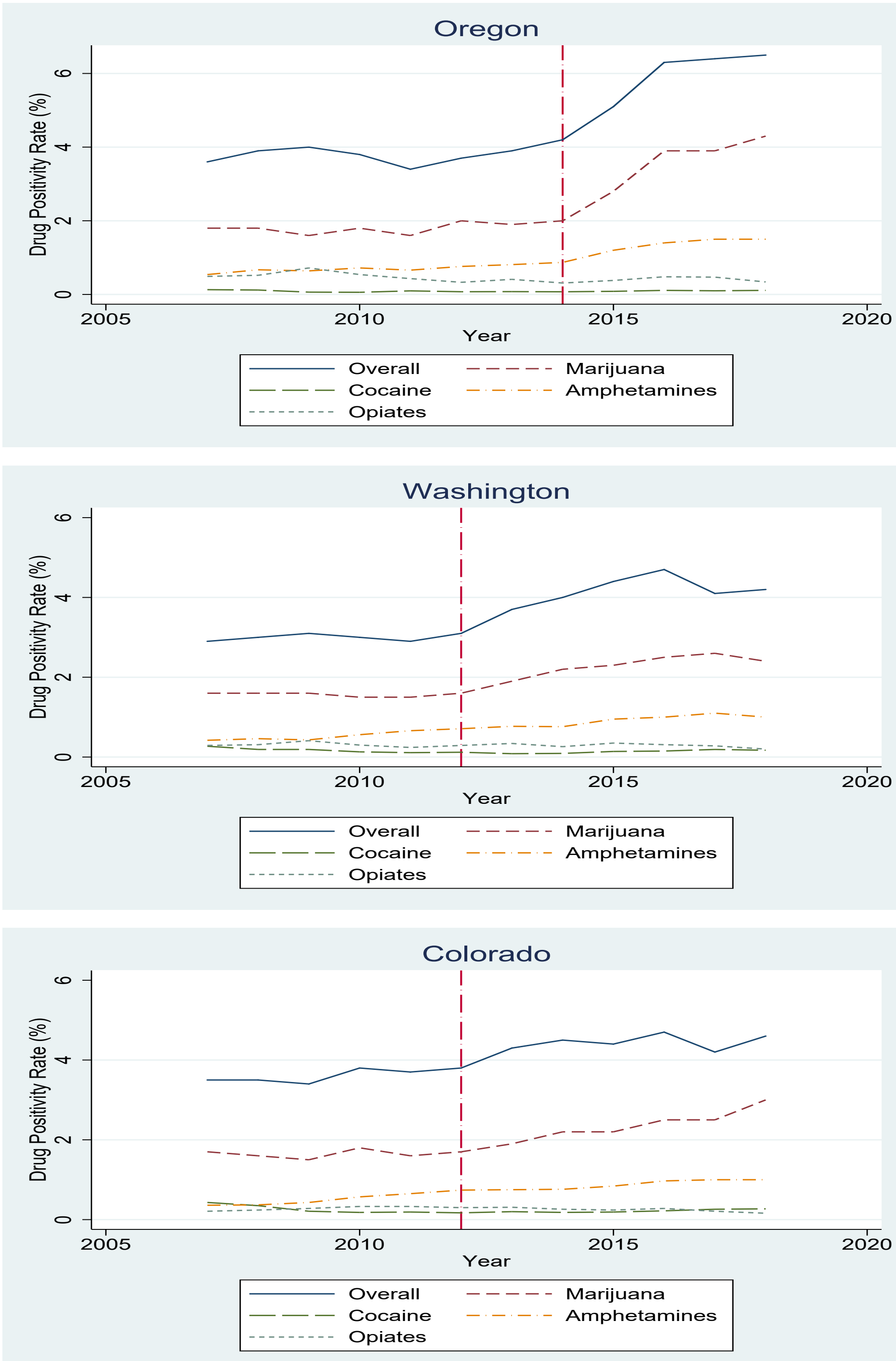
Figures

Figure 1: Total Marijuana Sales by Month and Customer Type.



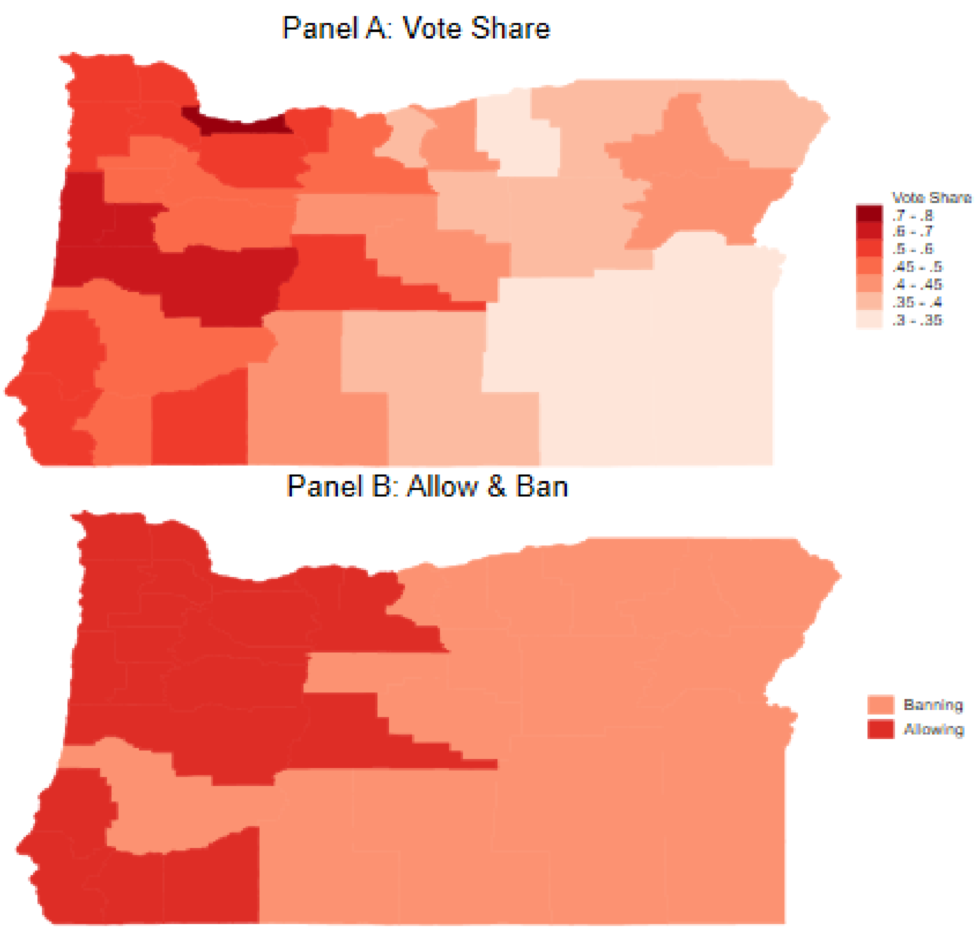
Note: Author using data from Oregon Liquor Control Commission. The red line indicates recreational marijuana sales, and blue line represents medical marijuana sales.

Figure 2: Workplace Drug Positivity Rate by Drug Category and State.



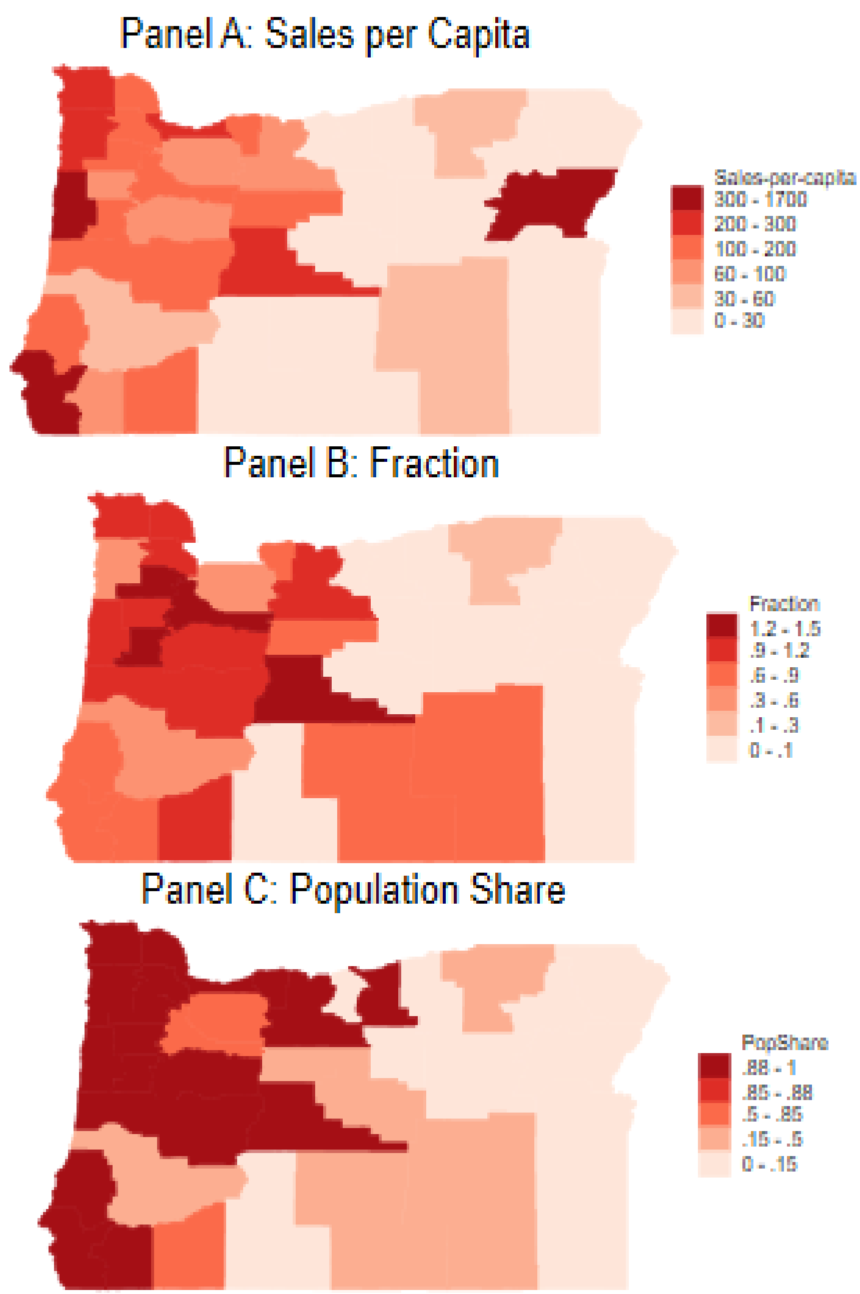
Note: Author using data from Quest Diagnostics. The vertical dashed red line is the recreational marijuana law passage year for each state.

Figure 3: Heatmap of Vote Share in favor of RML



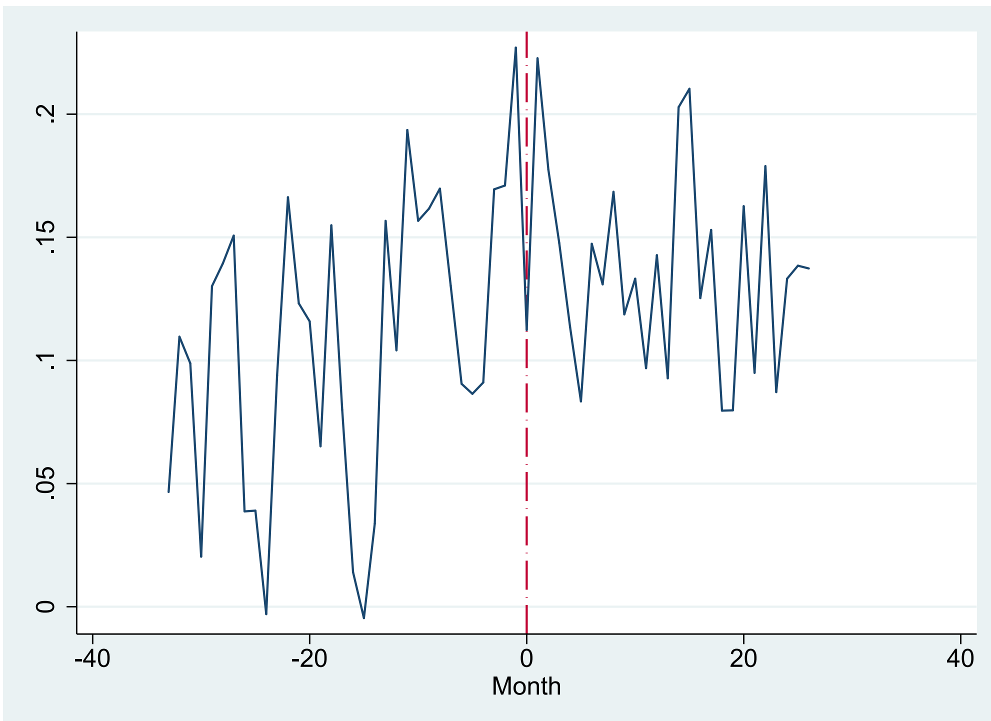
Note: The figure shows the geographic variation in the vote share in favor of recreational marijuana sales legalization (panel A) and the vote share cutoff 46% with allowing and banning sales counties (panel B).

Figure 4: Heatmap of Recreational Marijuana Exposure Measures



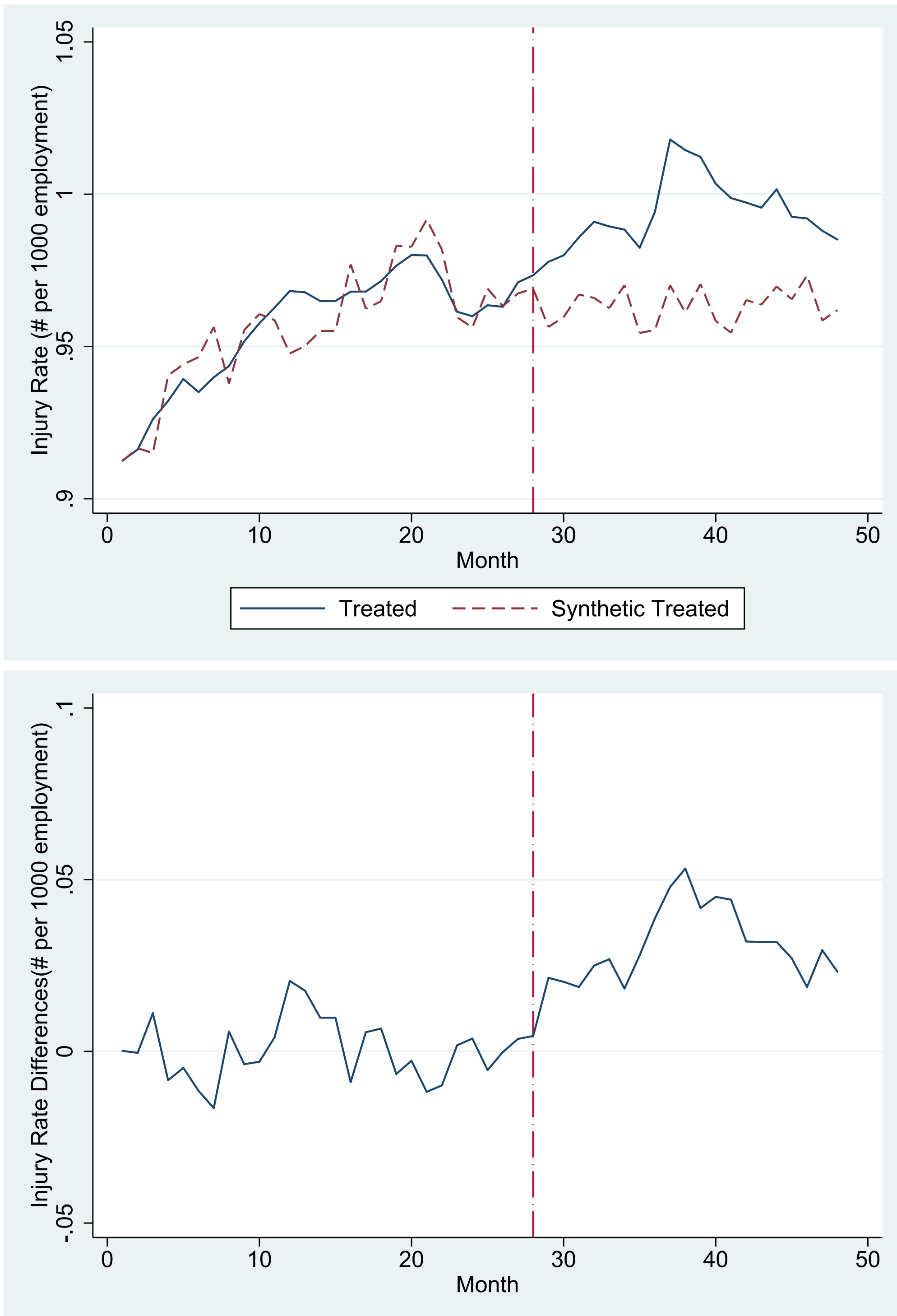
Note: The figure shows the geographic variation in the recreational marijuana measures in Oregon. Panel A-C are sales-per-capita, fraction and population share exposure measure, respectively.

Figure 5: Difference in Injury Rate by Treatment Status



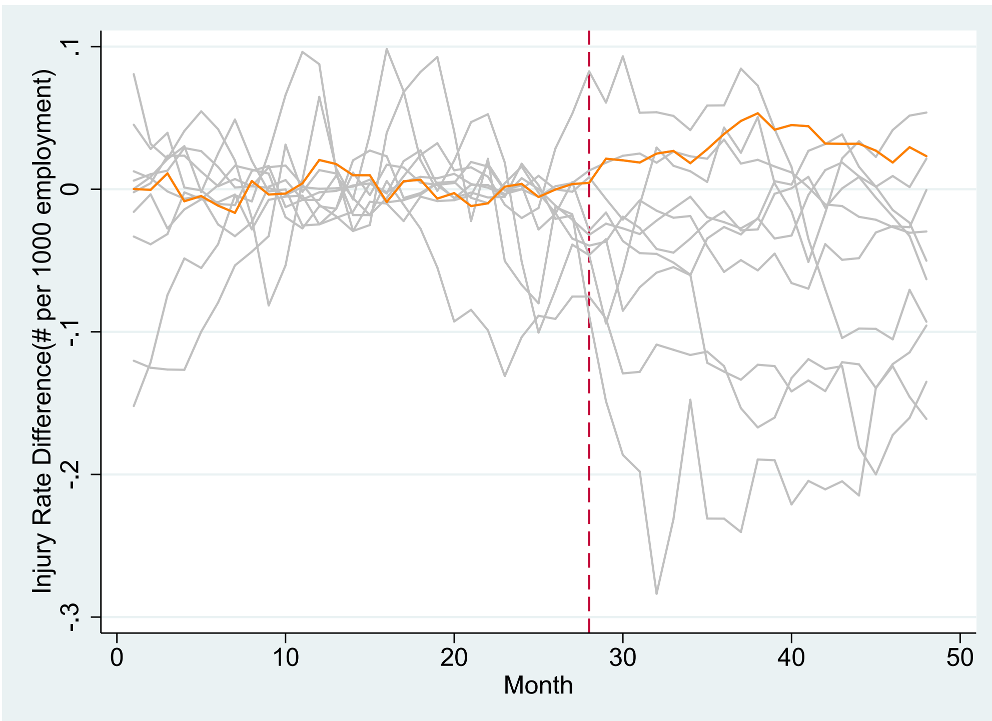
Note: The figure shows the DiD model's mean difference of the injury rate between treatment and control counties for each month from 2013 to 2017. Zero in the horizontal axis (and vertical dashed red line) denoted as October 2015, when recreational marijuana sales market opens.

Figure 6: Synthetic Control Graphs



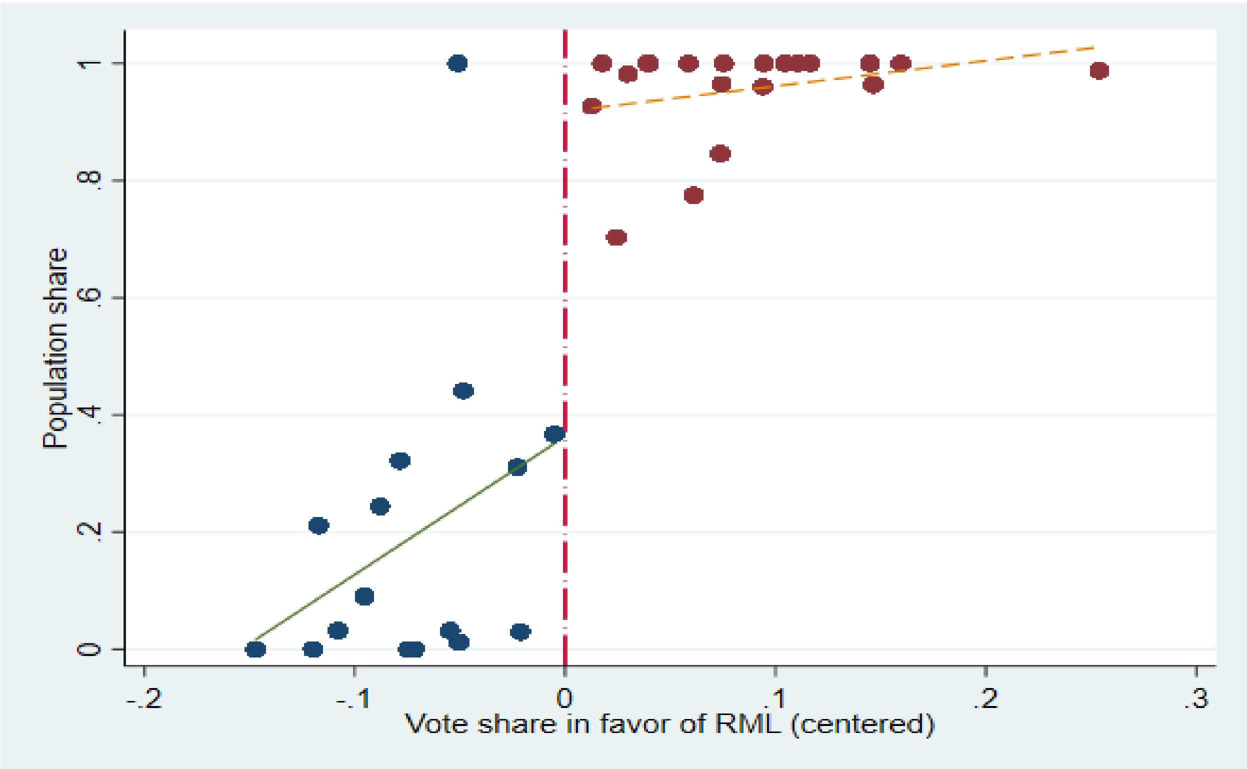
Note: Top panel shows injury rate for treated and synthetic treated. Bottom panel presents the difference in injury rate for the same two units. The vertical dashed red line in both panels indicates the RML Passage, October 2015.

Figure 7: Permutation Test



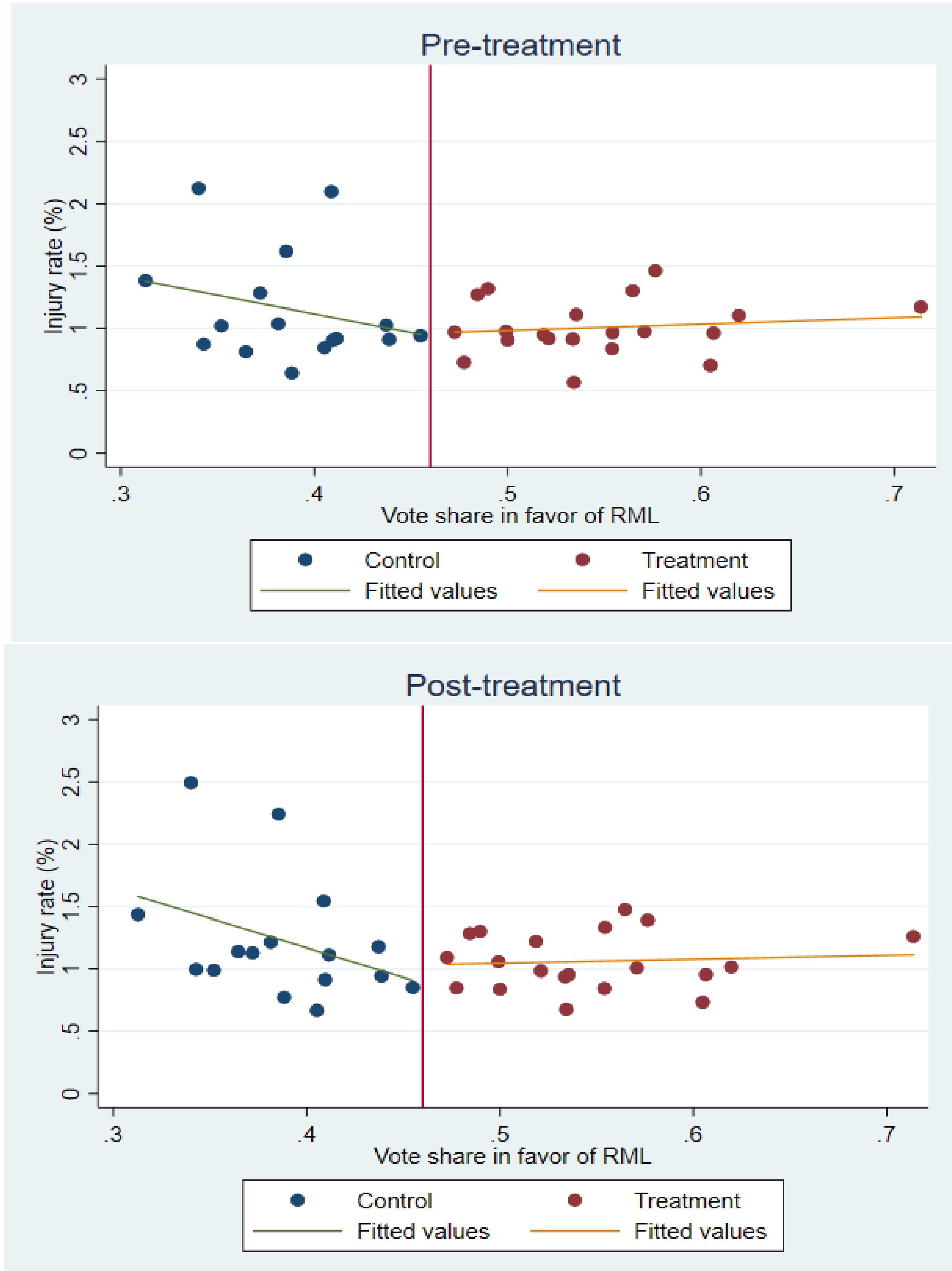
Note: The figure shows the differences in injury rate between the treated vs the synthetic treated (orange line) compared with the distribution of placebo differences (grey lines). The vertical dashed red line indicates the RML Passage, October 2015.

Figure 8: Discontinuities in the Treatment



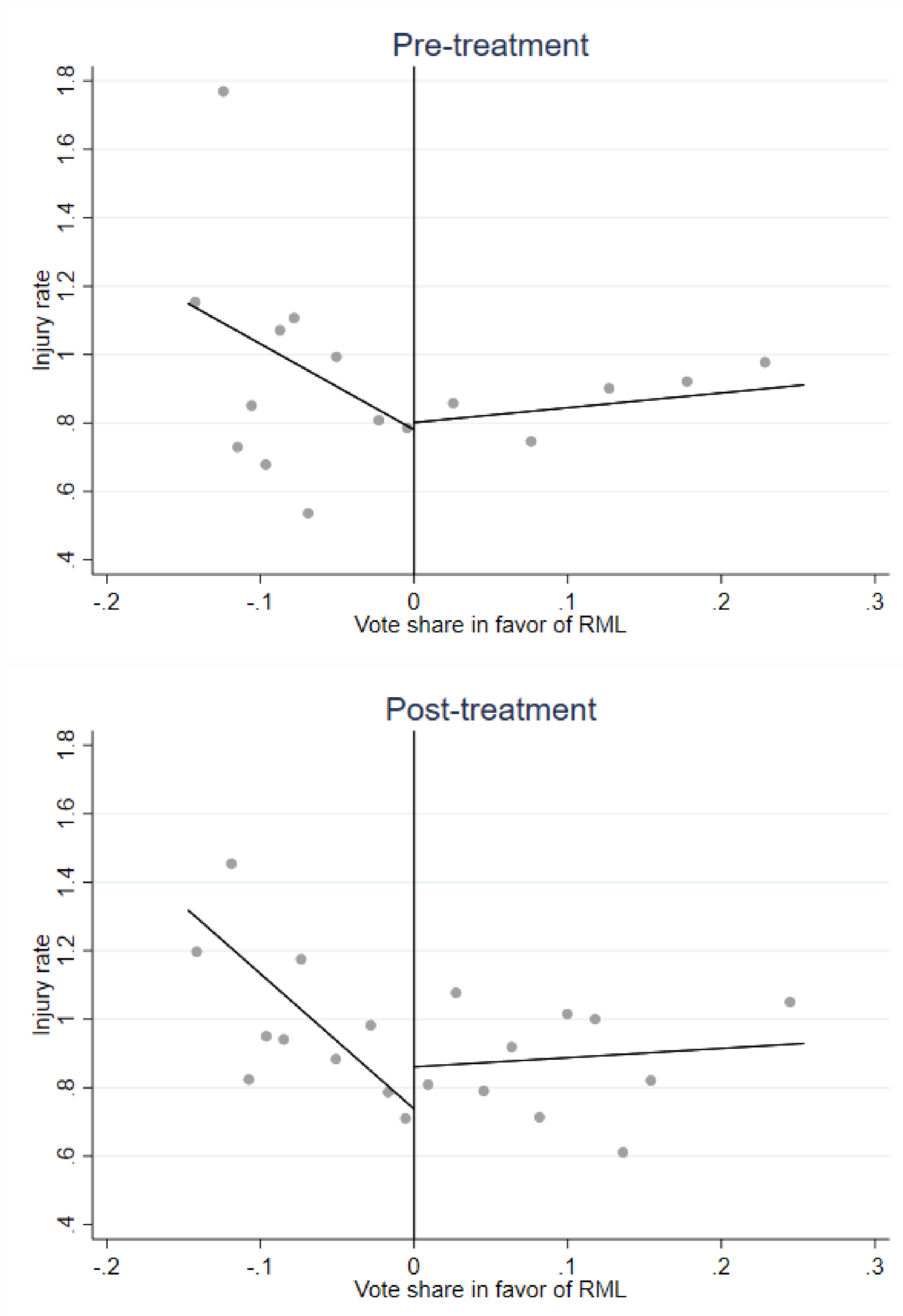
Note: The figure shows the discontinuities in the population share lives in the jurisdictions that allows recreational marijuana sales at the county-specific vote shares. Vote share is centered at 46%. The scatters are overlaid linear fitted lines.

Figure 9: Raw Yearly Injury Rate by Vote Share



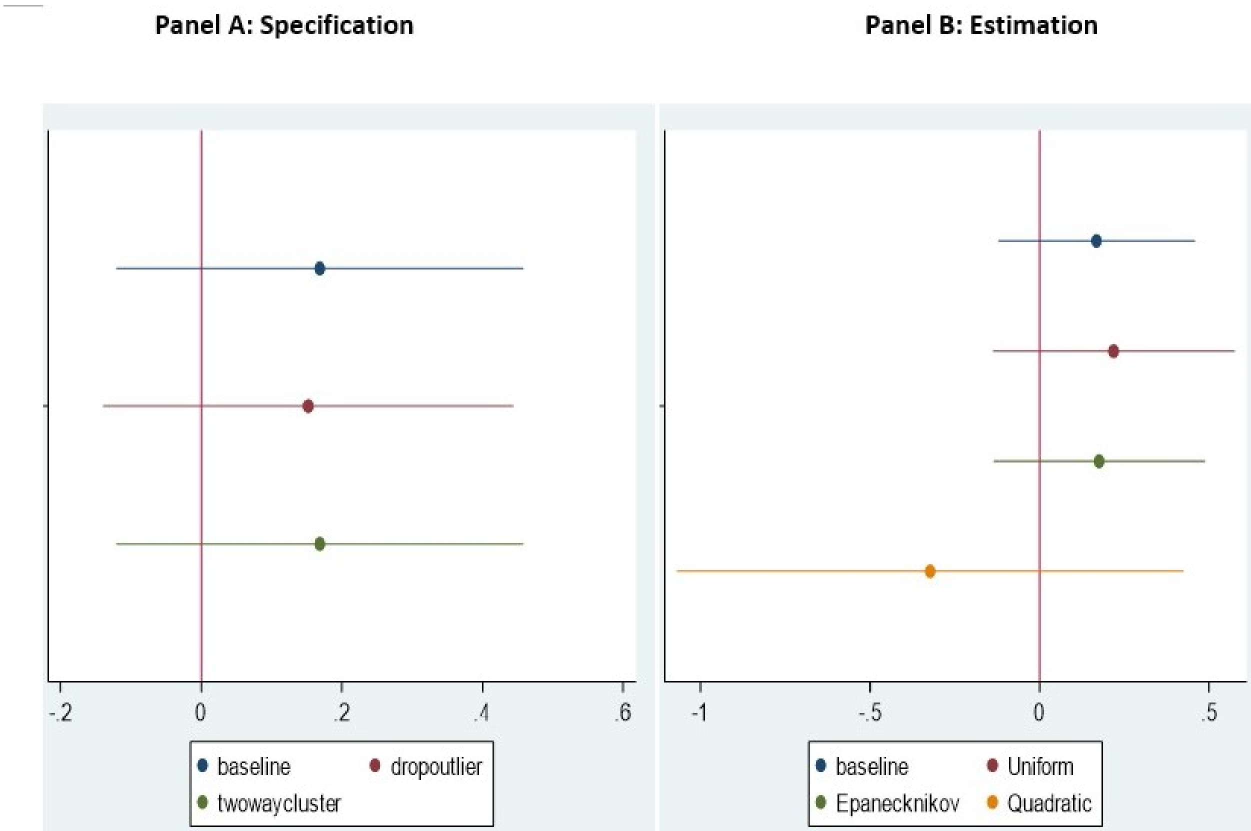
Note: The figure shows the raw yearly injury rate in the sample one year before (top panel) and one year after RML (bottom panel) by county-specific vote share with overlaid linear fitted lines. The vertical red line is the vote share cutoff-46 percent in favor of RML.

Figure 10: RD Graphs



Note: The figure shows the effect of RML on workplace injuries one year before treatment (2013) and one year after treatment (2017). The vertical line is the vote share cutoff-46 percent in favor of RML. Injury rates are number of claims per 1000 employment. The points show the average value of the outcome within 0.011 vote share distance bin. The lines are fitted values from local linear regression.

Figure 11: RD Robustness Checks



Note: The graphs show the robustness of the main results. Panel A shows robustness to different alternative specifications. Panel B shows that the results are robust to using different estimation techniques. The graphs report RD estimates on injury rate, together with 95% confidence intervals, for the sample of post-treatment. All coefficients are estimated using locally linear regression and a triangular kernel for a 0.054 bandwidth unless otherwise specified. Standard errors are clustered at the county level.

Tables

Table 1: Workers' Compensation Claims Characteristics

	# of claims
All observation	92,181
By year	
2013	17,840
2014	18,412
2015	18,285
2016	18,996
2017	18,648
By age range	
Under 18	619
18-24	10,201
25-34	19,986
35-44	20,181
45-54	21,690
55-64	16,435
Over 65	3,069
By industry	
Agriculture, Forestry, Fishing and Hunting	5,232
Mining	109
Construction	7,990
Manufacturing	12,526
Transportation, Information and Utility	8,949
Wholesale	4,225
Retail Trade	10,998
Finance, Insurance and Real Estate	1,693
Services	35,630
Public Administration	4,829
By gender	
Male	57,947
Female	34,232

Source: Oregon Department of Consumer and Business Services, Workers' Compensation Division.

Table 2: Summary Statistics of Main Analysis Sample

	Mean	S.D
Monthly injury counts	42.68	1.59
Employment	48,981	78,412
Monthly injury rate (per 1000 employment)	0.91	0.53
N	2,160	.
Recreational Marijuana Exposure Measures		
Sales-per-capita (in dollars)	119.65	179.89
Fraction	0.34	0.27
Population Share	0.62	0.42

Note: Top panel consistes of monthly injury rate data between 2013-2017, for all 36 counties in the Oregon. Bottom panel reports the recreational marijuana exposure measures at the county level.

Table 3: The Reduced Form Effect

	(1) injuryrate
$\mathbb{1}(X_j \geq 0.46) * After$	0.019 (0.030)
DV Mean	0.913
R-Squared	0.548
N	2160

Note: The table reports the "naive" DiD estimates from Eq.(3). Legalization is defined as above 46% of county vote share. After=1 if injury rate is after Octorber 2015. Standard error is clustered at county level in parentheses.

Table 4: The Effect of Recreational Marijuana Sales Legalization

	(1) injuryrate	(2) injuryrate	(3) injuryrate
Legalization*After	0.034**	0.034**	0.050*
	(0.016)	(0.016)	(0.025)
DV Mean	0.913	0.913	0.913
R-squared	0.549	0.565	0.572
N	2160	2160	2160
Month	Yes	Yes	Yes
Year	Yes	Yes	Yes
County	Yes	Yes	Yes
Year-by-Month FE		Yes	Yes
County specific trend			Yes

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as sales-per-capita that are above 75th percentile. After=1 if injury rate is after October 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 5: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Age Range

	(1) Under 18	(2) 18-24	(3) 25-34	(4) 35-44	(5) 45-54	(6) 55-64	(7) Above 65
Legalization*After	0.000	0.006**	0.010*	0.005	0.007	0.001	0.005**
	(0.001)	(0.003)	(0.005)	(0.004)	(0.006)	(0.007)	(0.002)
DV Mean	0.008	0.096	0.182	0.192	0.221	0.179	0.035
R-squared	0.089	0.184	0.297	0.271	0.233	0.211	0.069
N	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4) by age range. Legalization is defined as sales-per-capita that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Gender

	(1) Female	(2) Male
Legalization*After	0.010	0.024*
	(0.009)	(0.013)
DV Mean	0.301	0.612
R-squared	0.417	0.468
N	2160	2160

Note: The table reports the DiD estimates from Eq.(4) by gender. Legalization is defined as sales-per-capita that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 7: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Industry

	(1) Agriculture, forestry, fishing, hunting	(2) Mining	(3) Construction	(4) Manufacturing	(5) Transportation, information, utility	(6) Wholesale	(7) Retail Trade	(8) Finance, insurance, real estate	(9) Services	(10) Public Administration
Legalization*After	0.003	0.000	0.008	-0.008	0.002	0.003	0.004	0.002	0.012*	0.009***
	(0.004)	(0.001)	(0.005)	(0.005)	(0.004)	(0.002)	(0.005)	(0.001)	(0.007)	(0.003)
DV Mean	0.128	0.003	0.084	0.130	0.081	0.037	0.091	0.012	0.289	0.058
R-squared	0.491	0.011	0.180	0.397	0.518	0.263	0.210	0.144	0.476	0.350
N	2160	2160	2160	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4) by industry. Legalization is defined as sales-per-capita that are above 75^{th} percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 8: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Occupation

	(1) Management, professional	(2) Social, legal, educational service	(3) Art, entertainment, sports	(4) Healthcare suppport	(5) Food, cleaning, sales	(6) Office, administrative support	(7) Construction, installation extraction	(8) Transportation material moving	(9) Military	(10) Others
Legalization*After	0.002	0.003	0.002*	0.001	0.004	0.005**	0.012**	0.013***	-0.000	-0.009
	(0.004)	(0.002)	(0.001)	(0.006)	(0.005)	(0.002)	(0.005)	(0.004)	(0.000)	(0.006)
DV Mean	0.041	0.025	0.003	0.119	0.147	0.036	0.146	0.155	0.000	0.136
R-squared	0.068	0.233	0.062	0.354	0.343	0.221	0.168	0.430	-0.018	0.257
N	2160	2160	2160	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4) by occupation. Legalization is defined as sales-per-capita that are above 75^{th} percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 9: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Ownership

	(1) Local Government	(2) State Government	(3) Private Ownership
Legalization*After	0.004	0.002	0.028*
	(0.005)	(0.003)	(0.014)
DV Mean	0.091	0.035	0.787
R-squared	0.268	0.707	0.507
N	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4) by ownership. Legalization is defined as sales-per-capita that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 10: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Injury Source

	(1) Overexertion	(2) Fall	(3) Struck	(4) Transportation	(5) Expose	(6) Personal	(7) Animal	(8) Fire	(9) Others
Legalization*After	0.004	0.026***	0.006	-0.002	0.001	0.001	-0.001	-0.000	-0.003
	(0.007)	(0.008)	(0.006)	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
DV Mean	0.353	0.248	0.180	0.044	0.022	0.027	0.010	0.003	0.026
R-squared	0.423	0.336	0.244	0.097	0.063	0.404	0.144	-0.015	0.160
N	2160	2160	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4) by source of injury. Legalization is defined as sales-per-capita that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 11: Synthetic Control Estimation Impact of the Recreational Marijuana Sales Legalization

	Average Difference relative to comparison pre-intervention	Average Difference relative to comparison Post-intervention	Difference-in-Difference Analysis	
			Change, Post-Pre	P-Value from one-tailed test
Synthetic Control (base case)	0.00	0.03	0.03***	0.002
Synthetic Control (for each treated county)	0.00	0.036	0.036***	0.001
Standard DiD (base case)	.	.	0.034**	.

Note: The table shows the estimated effects across methods. Average differences pre- and post-intervention are estimates of the difference in the injury rates in treated county relative to the matched synthetic comparison group. The SCMs' p-value are based on the one-tailed test of the significance of the difference -in-differences estimates employs the empirical distribution of the placebo effect estimates of RML for comparison counties. The DiD's p-value is clustered at county level. * p<0.10, ** p<0.05, *** p<0.01.

Table 12: The RD Estimation Impact of the Recreational Marijuana Sales Legalization

	Pre-treatment				Post-treatment			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Injuryrate	-0.035	-0.003	-0.048	-0.122	0.236	0.242	0.195	0.182
	[0.211]	[0.197]	[0.193]	[0.148]	[0.197]	[0.191]	[0.169]	[0.178]
N	432	432	432	432	432	432	432	432
Bandwidth	0.09	0.079	0.068	0.045	0.108	0.095	0.081	0.054
Control Mean	0.916	0.899	0.872	0.803	0.914	0.916	0.899	0.907

Note: The table shows the effect of RML on workplace injury rate using Regression Discontinuity Design. It presents RD estimates for pre-treatment periods (2013) and post-treatment periods (2017). Injury rate are injury claims per 1000 employment. The coefficients are estimated using locally linear regression and a triangular kernel for four different bandwidths: 200%, 175%, 150%, and 100% of optimal bandwidth choice developed by Calonico et al. (2014). Standard errors are clustered at county level are shown in parentheses. The control mean is the mean of the outcome variable for all counties below the thereshold within the respective bandwidth. * p<0.10, ** p<0.05, *** p<0.01.

Appendix

Synthetic Control Method

This section describe how I extend the Synthetic Control Method into multiple treatment units in detail. I first describe the SCM for settings when only a single unit is exposed to treatment. Let $j = (0, ..., J)$ denote counties where the value $j = 0$ corresponds to the county in the treatment group, and the remaining counties are the "donor pool". Outcomes are observed for T periods with the RML starting in $T_0 + 1$. The observed outcome vector of each county is $Y_j = (Y_{j1}...Y_{jT_0}...Y_{jT})'$. The observed outcome can be written as the sum of a untreated potential outcome (Y_{jt}^N), and the effect of the RML (β_{jt}), such that:

$$Y_{jt} = \beta_{jt}D_{jt} + Y_{jt}^N = \beta_{jt}D_{jt} + \delta_t + \theta_t Z_j + \gamma_t \mu_j + \varepsilon_{jt} \tag{7}$$

where D_{jt} is an indicator variable that takes the value of 1 for the treated unit after T_0, and is 0 otherwise. Variable δ_t is a time fixed effect, Z_j is a vector of time-invariant observed (or measured) predictors with time-varying coefficient vector θ_t, μ_j is a vector of time-invariant unobserved predictor variables with time-varing coefficients γ_t. ε_{jt} are unobserved transitory shocks with zero mean. Under the assumption that the relationship between the outcome and the predictors is linear, the SCM generalizes the DiD model by allowing the effects of γ_t of the unobserved predictors μ_j to differ over time while the DiD constrains these effects to be constant.

Before the intervention, the untreated potential outcome Y_{jt}^N corresponds to the observed outcome, for both the treated and control counties. For period after T_0, the untreated counter-factual for the treated county is Y_{0t}^N, which is not observable. In order to estimate the treatment effect for the post-intervention periods, the SCM estimates the unobserved Y_{0t}^N by creating a "synthetic control unit" as follows. Define F_0 as a $K \times 1$ vector with elements equal to the injury rate in the treatment county pre-RML (January 2013 to September 2015), plus additional covariates predictive of the presence of injury rate.[37] Similarly, define a $K \times J$ matrix, F_J, in which row j is the sequence of values for the same variables and time relative to county j in the "donor pool".

The SCM then identifies the vector of non-negative weights $W^* = (\omega_1, ..., \omega_J)$ that create a convex combination of variables for counties in the donor pool, (F_J), that best approximates the pre-RML injury rate of the treated county. The product $F_J W^*$ then gives a weighted average of the pre-RML vectors for all counties (omitting the treatment county). The difference between treatment county and this average given by $F_0 - F_J W^*$. In other words, SCM minimizes the difference between $F_0 - F_J W^*$:

$$\begin{aligned} W^* &= argmin(F_0 - F_J W)^{'}(F_0 - F_J W) \\ s.t. &\sum_{j=1}^{J} \omega_j = 1, \omega_j \geq 0 \end{aligned} \tag{8}$$

Once an optimal weight vector W^* is chosen, both the pre-RML path and the post-RML values for the injury rate in "synthetic county" can be tabulated by calculating the corresponding weighted average for each month using the donor counties with positive weights. The post-RML values for the synthetic control group serve as the counter-factual outcomes for the treatment county.

Hence, the estimator of the counter-factual is constructed as the linear combination of the observed outcomes of the potential control counties: $\hat{Y_{0t}^N} = \sum_{j=1}^{J} \omega_j Y_{jt}$. The estimated treatment effect for the treated county in each time period after T_0 can then be obtained as $\hat{\beta_{jt}} = Y_{0t} - \hat{Y_{0t}^N}$.

Intuitively, the principal estimate of the impact of RML on workplace injury rate uses the synthetic control group to calculate a simple DiD estimate. Specifically, defined as:

$$\begin{aligned} DD = &(Outcome_{Post}^{HighExposure} - Outcome_{Post}^{synth}) \\ &- (Outcome_{Pre}^{HighExposure} - Outcome_{Pre}^{synth}) \end{aligned} \tag{9}$$

Where $Outcome_{Post}^{HighExposure}$ and $Outcome_{Post}^{synth}$ are the injury rate for treated county and synthetic

[37] The other covariates include unemployment rate, proportion of county population in each of four educational attainment categories: less than high school, high school graduate, some college and college graduate. These additional covariates are measured at the baseline year 2010.

county after RML-October 2015, respectively. And $Outcome_{Pre}^{HighExposure}$ and $Outcome_{Pre}^{synth}$ are the injury rate for treated county and synthetic county pre-RML period for each group.[38]

To test the significance of any observed relative increase in treated county's injury rate, I apply permutation test suggested by Abadie et al. (2010) to the DiD estimator displayed in equation (3).[39] Specifically, for each county in the donor pool, I identify synthetic comparison groups based on the solution to the minimization problem in equation (2). I then estimate the DiD in equation (3) for each county as if these counties had the treatment. The distribution of these "placebo" DiD estimates then provides the equivalent of a sampling distribution for the estimate DD. To be specific, if the cumulative density function of the complete set of DiD estimates is given by $F(.)$, the p-value from a one-tail test of the hypothesis that $DD > 0$ is given by $F(DD)$.

Multiple Treatment Units

Since in my context, the treated counties are more than one,[40] I modify Abadie et al. (2010) approach and construct the synthetic control county by weighting the control counties to match the weighted average pre-RML injury rate of the treated counties.

Let $j = (0, ..., J)$ denote counties and t be the time period. With $j = 0$ to M_1 counties are treated, while the remaining $M_1 + 1$ to $M_1 + M_2$ counties are controls. As before, the observed outcome of a county in each month can be written as $Y_{jt} = \beta_{jt} D_{jt} + Y_{jt}^N$. The aggregate outcome for the treated counties can be defined as the previous expression weighted by the population of each county, such that

$$\bar{Y}_t = \bar{\beta}_t D_t + \bar{Y_t^N} \tag{10}$$

[38]Since my data structure is high-frequency monthly injury rate, I use moving average lagged six months to smooth out the seasonality.

[39]Buchmueller et al. (2011) use a similar permutation test to test for an impact of Hawaii's employer mandate to provide health insurance benefits to employees on benefits coverage, health care costs, wages, and employment. Bohn et al. (2014) use similar method to text for an effect of Arizona's 2007 Legal Arizona Workers Act on the proportion of the state's noncitizen Hispanic population.

[40]Recall I define counties with sales-per-capita above the 75 percentile as treatment counties. Hence, there are several treated counties above treatment cutoff.

where $\bar{Y}_t = \dfrac{\sum_{j=0}^{M_1} Y_{jt} p_j}{\sum_{j=0}^{M_1} p_j}$, $\bar{\beta}_t = \dfrac{\sum_{j=0}^{M_1} \beta_{jt} p_j}{\sum_{j=0}^{M_1} p_j}$, $\bar{Y_t^N} = \dfrac{\sum_{j=0}^{M_1} Y_{jt}^N p_j}{\sum_{j=0}^{M_1} p_j}$, D_t is the treatment indicator, and p_j denotes population in each treated county. As with DiD estimation, the SCM with multiple treated units identifies the Average Treatment on the Treated (ATT) parameter. I calculate the ATT by averaging the estimated treatment effects $\bar{\beta}_t$ over the post-RML period, weighted by the population in the treated counties. I then propose to use the SCM to estimate the counter-factual outcome for the treated region ($\bar{Y_t^N}$), by re-weighting the outcomes of the control counties: $\hat{\bar{Y_t^N}} = \sum_{j=K_1+1}^{K_1+K_2} \omega_j Y_{jt}$, where the weight vector W minimizes the pre-RML injury rate of the treated county and the synthetic control county.[41] The estimated treatment effect for the treated county for each time period after T_0 can then be obtained as $\hat{\beta}_t = Y_t - \hat{Y_t^N}$.

To test the significance of any observed relative increase in treated county's injury rate, I apply permutation test suggested by Abadie et al. (2010) to the DiD estimator displayed in equation (3). The main idea is to simulate a distribution of difference between each county in the donor pool and its synthetic control and examine whether treated county shows a post-RML difference from its synthetic control, relative to its pre-RML difference, that is large vis-a-vis the whole distribution. The distribution of these "placebo" DiD estimates then provides the equivalent of a sampling distribution for the estimate DD. To be specific, if the cumulative density function of the complete set of DiD estimates is given by $F(.)$, the p-value from a one-tail test of the hypothesis that $DD > 0$ is given by $F(DD)$.[42]

[41] In this context, I drop 25 percentile and below population counties that are in the donor pool.

[42] Acemoglu et al. (2016) use the SCM to construct the untreated potential outcome for each multiple treated unit and weights the estimated unit-level treatment effects based on the closeness of the synthetic control. Their inference procedure is similar to the one developed here, in that they re-sample placebo-treated units from the control pool. Dube and Zipperer (2015)) pool multiple estimates of treatment effects to generalise inference for a setting with multiple treated units and policies. Xu (2017) proposes a generalisation for the SCM for multiple treated units with a factor model that predicts counterfactual outcomes. Lastly, Kreif et al. (2016) extend the SCM to a setting of a evaluation of a health policy where there are multiple treated units, and test for the null hypothesis that the ATT for the treated region is zero, using a representation of the distribution of the ATT under the null, which is somewhat similar to this paper.

Regression Discontinuity Design

This section describes the alternative estimation technique in the fuzzy RD setting, that is using parametric approach that includes all observations and each of which carries the same weight regardless of how far away from the cutoff it is. This serves as a robustness checks for the RD results. Formally, I utilize Two-Stage Least Squares (2SLS) in the following parametric equations:

$$InjuryRate_{jt} = \theta + \theta_1 \tilde{X}_j (+\theta_2 \tilde{X}_j^2) + \kappa_0 D_j + \pi_1 D_j \tilde{X}_j (+\pi_2 D_j \tilde{X}_j^2) + \xi_{jt}. \tag{11}$$

$$D_j = \gamma + \gamma_1 \tilde{X}_j (+\gamma_2 \tilde{X}_j^2) + \pi_0 T_j + \pi_1 T_j \tilde{X}_j (+\pi_2 T_j \tilde{X}_j^2) + \mu_j. \tag{12}$$

Where $InjuryRate_{jt}$ is injury rate for county j and time (month, year) t. D_j is a high exposure indicator that takes a value of 1 when counties who have marijuana exposure measures certain threshold, zero otherwise. $\tilde{X}_j = (X_j - 0.46)$ is vote share centered at the 46%. I instrument high exposure indicator using the RML vote share cutoff $T_j = \mathbb{1}(\tilde{X}_j \geq 0)$ which equals one if county vote share exceed the official 46% cutoff for marijuana sales legalization. In order to assess more flexible functional forms, the polynomials and interaction terms in parentheses can be added to the model. Finally, ε_{jt}, μ_j are idiosyncratic error terms.

Eq.(11) shows that the treatment effect of RML captured by the parameter κ_0. Eq.(12) is classical first-stage equation linking the endogenous treatment variable D_j to the respective set of exogenous variables and the instruments. The specifications also provide adequate representation of the functional form of the relationship between high exposure and injury rate and the vote share cutoff, as I allow for different slopes on both sides of the cutoff through the inclusion of the interactions of county vote share and the instrument. As in any standard IV framework, the estimated treatment effects have to be interpreted as local average treatment effects (LATE). That is, I estimate the average treatment effect for those counties who allows marijuana sales due to county vote share cutoff in Measure 91 (“compliers”).

Instrumental variable estimates for the effect of RML on workplace injuries are reported in

Table A8. Column (3) through column (6) assess the robustness of my results to picking an smaller vote share window around the 46% cutoff. For instance, column (3) focuses on observations who is between 0.26 and 0.66 vote share counties. In essence, this creates a 0.2 vote share window to both side of the cutoff. As shown in the table, the point estimates remain stable compare with the local linear regression approach, but the standard errors inflate as the sample size shrinks. As mentioned in the main context, a big caveat is that my point estimates in the RD setting lost a lot of precision due to lack of power around the cutoff. Nonetheless, the lack of power make the results somewhat less reliable.

Appendix Tables

Table A1: The Effect of Recreational Marijuana Sales Legalization

	(1) injuryrate	(2) injuryrate	(3) injuryrate
Legalization*After	0.030**	0.030**	0.049*
	(0.012)	(0.012)	(0.028)
DV Mean	0.913	0.913	0.913
R-squared	0.549	0.565	0.573
N	2160	2160	2160
Month	Yes	Yes	Yes
Year	Yes	Yes	Yes
County	Yes	Yes	Yes
Year-by-Month FE		Yes	Yes
County specific trend			Yes

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A2: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Age Range

	(1) Under 18	(2) 18-24	(3) 25-34	(4) 35-44	(5) 45-54	(6) 55-64	(7) Above 65
Legalization*After	-0.001	0.006**	0.013**	0.007	0.008	-0.005	0.003
	(0.001)	(0.003)	(0.006)	(0.006)	(0.005)	(0.004)	(0.002)
DV Mean	0.008	0.096	0.182	0.192	0.221	0.179	0.035
R-squared	0.089	0.185	0.298	0.271	0.233	0.211	0.068
N	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A3: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Gender

	(1) Female	(2) Male
Legalization*After	0.018***	0.012
	(0.006)	(0.009)
DV Mean	0.301	0.612
R-squared	0.418	0.467
N	2160	2160

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75th percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A4: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Industry

	(1) Agriculture, forestry, fishing, hunting	(2) Mining	(3) Construction	(4) Manufacturing	(5) Transportation, information, utility	(6) Wholesale	(7) Retail Trade	(8) Finance, insurance, real estate	(9) Services	(10) Public Administration
Legalization*After	0.006	-0.000	0.010**	0.000	-0.002	0.002	-0.004	0.001	0.014**	0.004**
	(0.004)	(0.000)	(0.004)	(0.007)	(0.003)	(0.002)	(0.005)	(0.001)	(0.006)	(0.002)
DV Mean	0.128	0.003	0.084	0.130	0.081	0.037	0.091	0.012	0.289	0.058
R-squared	0.492	0.011	0.182	0.396	0.518	0.263	0.210	0.144	0.476	0.349
N	2160	2160	2160	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75^{th} percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A5: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Occupation

	(1) Management, professional	(2) Social, legal, educational service	(3) Art, entertainment, sports	(4) Healthcare suppport	(5) Food, cleaning, sales	(6) Office, administrative support	(7) Construction, installation extraction	(8) Transportation material moving	(9) Military	(10) Others
Legalization*After	-0.004**	-0.000	0.002	0.014**	-0.004	0.005***	0.009**	0.002	-0.000	-0.002
	(0.002)	(0.002)	(0.001)	(0.007)	(0.005)	(0.001)	(0.004)	(0.005)	(0.000)	(0.008)
DV Mean	0.041	0.025	0.003	0.119	0.147	0.036	0.146	0.155	0.000	0.136
R-squared	0.069	0.232	0.061	0.357	0.343	0.222	0.168	0.429	-0.018	0.256
N	2160	2160	2160	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75^{th} percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A6: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Ownership

	(1) Local Government	(2) State Government	(3) Private Ownership
Legalization*After	0.004	0.002	0.028*
	(0.005)	(0.003)	(0.014)
DV Mean	0.091	0.035	0.787
R-squared	0.268	0.707	0.507
N	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75^{th} percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A7: The Heterogeneous Effect of Recreational Marijuana Sales Legalization, by Injury Source

	(1) Overexertion	(2) Fall	(3) Struck	(4) Transportation	(5) Expose	(6) Personal	(7) Animal	(8) Fire	(9) Others
Legalization*After	-0.004	0.016**	0.008*	0.001	0.001	0.005*	0.001	0.001	0.001
	(0.006)	(0.007)	(0.004)	(0.003)	(0.001)	(0.003)	(0.001)	(0.000)	(0.002)
DV Mean	0.353	0.248	0.180	0.044	0.022	0.027	0.010	0.003	0.026
R-squared	0.423	0.335	0.245	0.097	0.063	0.406	0.144	-0.015	0.159
N	2160	2160	2160	2160	2160	2160	2160	2160	2160

Note: The table reports the DiD estimates from Eq.(4). Legalization is defined as fraction measure that are above 75^{th} percentile. After=1 if injury rate is after Octorber 2015. Standard errors are clustered at county level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

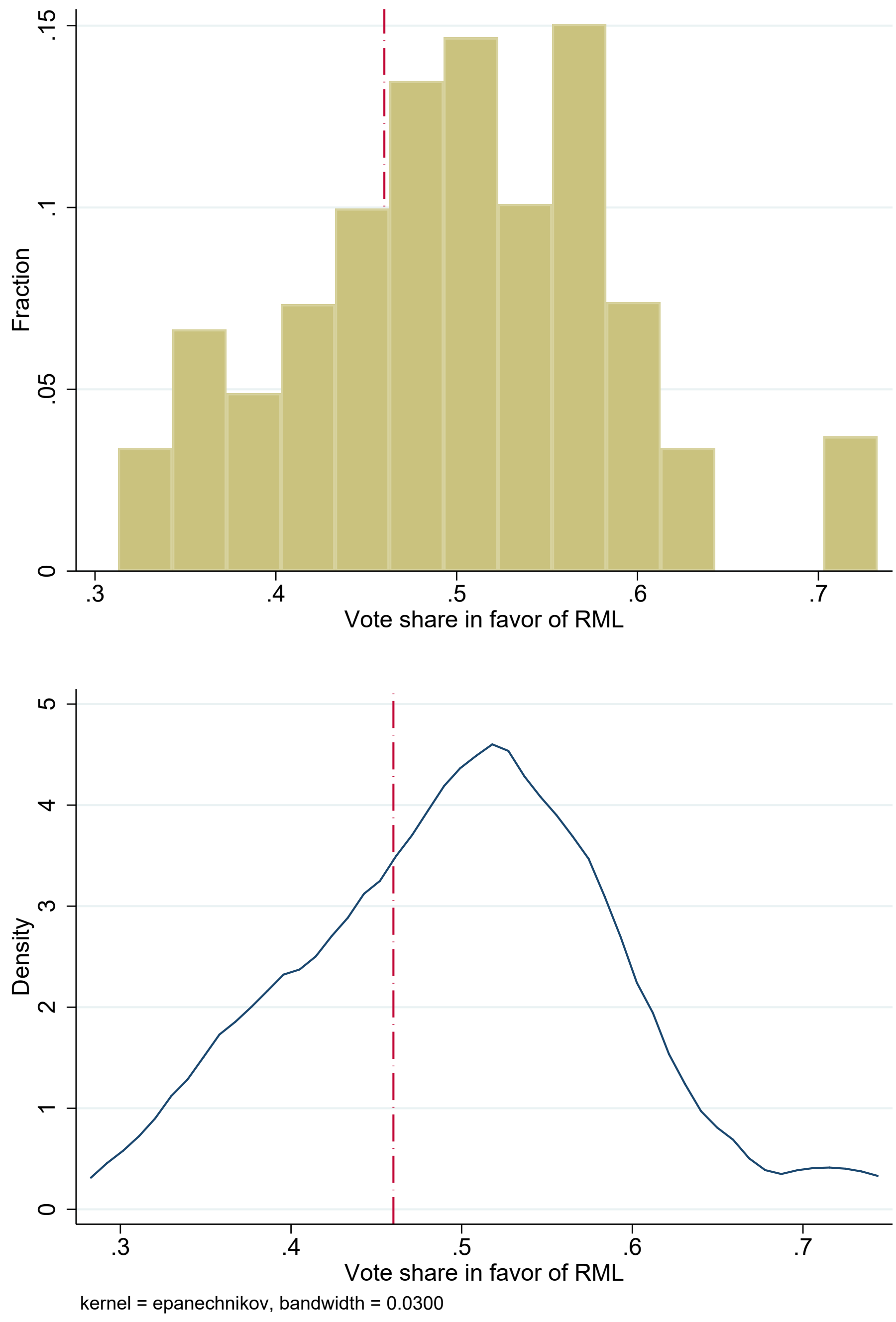
Table A8: The 2SLS Results by Specification and Vote Share Window

	(1) Full Sample	(2) ±0.2 Vote Share	(3) ±0.15 Vote Share	(4) ±0.1 Vote Share
Linear	0.160 (0.263)	0.226 (0.291)	0.239 (0.299)	0.217 (0.256)
Squared	0.257 (0.331)	0.117 (0.329)	0.060 (0.319)	0.085 (0.302)
DV Mean	0.913	0.909	0.909	0.874
N	2160	2100	2040	1500

Note: Each column reports coefficients and standard errors from the second stage of a 2SLS instrumental variable regression. The outcome variable is injury rate defined as injury claims per 1000 employment. Coefficients are displayed for the main explanatory variable, a dummy indicating high exposure county. The dummy was instrumented using a variable for whether vote share passing 46% cutoff. Standard error is clustered at county level in parentheses.

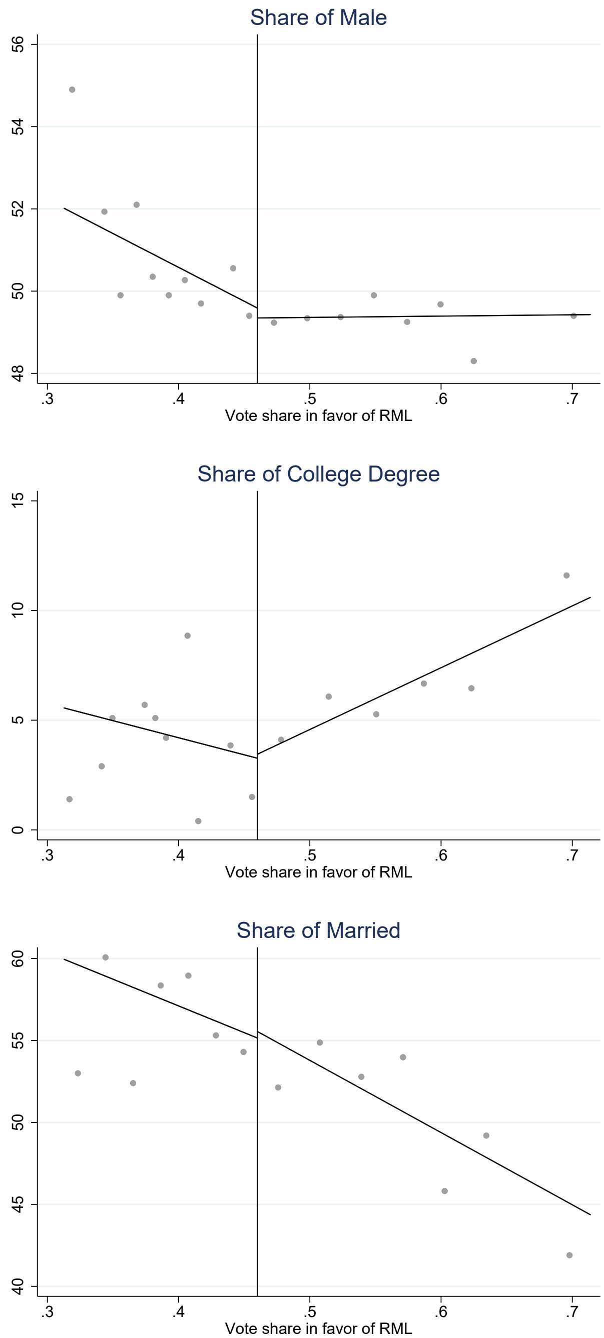
Appendix Figures

Figure A1: RD Validity Check I: Density of Vote Share



Note: The figure shows the distribution of county vote share around the official RML cutoff. The vertical dashed red line is the vote share cutoff-46 percent in favor of RML. Top panel is the distribution figure and the bottom panel is the kernel density graph.

Figure A2: RD Validity Check II: Baseline Covariates



Note: The figure shows the reduced-form effects for the predetermined covariates education, gender and marital status around the cutoff. The vertical line is the vote share cutoff-46 percent in favor of RML. The scatters are overlaid with local polynomial smooths.