

Testing the Substitution Hypothesis: Cannabis Dispensary Access and Alcohol Involvement in Fatal Crashes

APEP Autonomous Research*

@anonymous

January 2026

Abstract

Does access to legal cannabis reduce alcohol involvement among fatal traffic crashes through substance substitution? I test this hypothesis using a spatial regression discontinuity design at the borders between legal and prohibition states in the western United States. Using 29,350 geocoded fatal crashes from the Fatality Analysis Reporting System (2016–2019) and 1,399 dispensary locations from OpenStreetMap, I estimate the effect of crossing from prohibition into legal territory on the share of fatal crashes involving alcohol-impaired drivers. The main specification yields a point estimate of 9.2 percentage points ($SE = 5.9$ pp), but the 95% confidence interval includes zero ($p = 0.127$). Complementary analysis using distance to nearest dispensary within prohibition states also produces null results (coefficient: -0.006 , $SE = 0.014$). These well-identified null findings challenge the popular narrative that cannabis legalization reduces alcohol-related harms through substitution. The evidence suggests that if substitution occurs, it does not affect the composition of fatal crashes at a detectable magnitude.

JEL Codes: I12, I18, K32, R41

Keywords: marijuana legalization, alcohol substitution, fatal crash composition, spatial RDD, null result

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

Alcohol-impaired driving kills approximately 10,000 Americans annually, accounting for roughly 28% of all traffic fatalities ([National Highway Traffic Safety Administration, 2023](#)). Despite decades of policy interventions—minimum legal drinking ages, DUI enforcement, ignition interlock requirements, and graduated licensing programs—alcohol-related crash rates remain stubbornly high. The social costs are staggering: estimates suggest that alcohol-impaired crashes impose over \$44 billion annually in economic costs, including medical expenses, lost productivity, legal costs, and property damage. A provocative body of research suggests an unconventional harm reduction pathway: marijuana legalization may reduce alcohol consumption and its associated harms through substance substitution ([Anderson et al., 2013](#); [Hansen et al., 2020](#)).

The substitution hypothesis has profound policy implications. If marijuana and alcohol are substitutes in the utility function of recreational users, then legalizing cannabis could reduce alcohol-related harms including traffic fatalities, liver disease, and domestic violence. Public health advocates have increasingly cited this potential “harm reduction” benefit as a secondary argument for legalization, distinct from the primary arguments centered on criminal justice reform and individual liberty. Conversely, if cannabis and alcohol are complements or independent goods, legalization policies should be evaluated without assuming offsetting safety benefits. The existing evidence is mixed: some studies find substitution effects ([Anderson et al., 2013](#); [Croft and Guerrero, 2012](#)), while others find null or complementary relationships ([Macdonald et al., 2016](#); [Dills et al., 2021](#); [Miller and Seo, 2021](#)).

The challenges facing researchers attempting to identify causal effects of marijuana policy on traffic safety are substantial. Standard difference-in-differences approaches compare outcomes in states before and after legalization to states that never legalized. But states that choose to legalize may differ systematically from those that do not—in demographics, political preferences, existing alcohol culture, road infrastructure, and countless other dimensions. Even with state fixed effects and flexible time trends, unobserved confounders may bias estimates. Selection into treatment is endogenous to state-level factors that may independently affect crash outcomes.

This paper provides a rigorous test of the substitution hypothesis using a spatial regression discontinuity design (RDD) that circumvents many of these identification challenges. The key insight is that state borders create sharp discontinuities in legal cannabis access. Residents of prohibition states like Wyoming or Idaho face substantially higher costs to access cannabis than residents of neighboring legal states like Colorado or Oregon. If substitution operates, crashes just inside legal state territory should have lower alcohol involvement than crashes

just inside prohibition state territory, holding other factors constant.

The spatial RDD exploits the institutional fact that cannabis policy changes discontinuously at state lines, while most other factors that affect crash risk—terrain, weather, road quality, population characteristics—vary smoothly across borders. A resident of Wendover, Utah (prohibition) faces fundamentally different cannabis costs than a resident of West Wendover, Nevada (legal), despite living only miles apart and sharing similar demographics, climate, and economic conditions. This design allows credible identification of local treatment effects at the border, complementing existing difference-in-differences evidence on aggregate state-level effects.

I construct a comprehensive dataset linking 29,350 geocoded fatal crashes from NHTSA’s Fatality Analysis Reporting System (FARS) with 1,399 cannabis dispensary locations from OpenStreetMap across 13 western states during 2016–2019. The western United States provides an ideal natural laboratory: five states (California, Colorado, Nevada, Oregon, Washington) legalized recreational cannabis and established retail markets during or before this period, while eight neighboring states (Arizona, Idaho, Kansas, Montana, Nebraska, New Mexico, Utah, Wyoming) maintained prohibition. I compute the signed distance from each crash location to the nearest legal-prohibition border and estimate the discontinuity in alcohol involvement at the border.

The main specification yields a point estimate of 9.2 percentage points: crashes on the legal side of the border actually show *higher* alcohol involvement than those on the prohibition side. However, this estimate is not statistically significant at conventional levels ($SE = 5.9$ pp, $p = 0.127$). The 95% confidence interval spans from -2.3 to +20.8 percentage points, encompassing both substitution and complementarity effects. The estimate is robust to bandwidth choice and polynomial order. Importantly, the direction of the point estimate—higher alcohol involvement in legal states—is inconsistent with the substitution hypothesis, even though statistical precision precludes strong conclusions.

As a complementary approach, I analyze crashes within prohibition states only, regressing alcohol involvement on log distance to the nearest dispensary in a legal state. The economic logic is straightforward: if cannabis access costs drive substitution, crashes closer to dispensaries should show lower alcohol involvement. Residents of prohibition states near the border can more easily travel to purchase cannabis legally; if this access reduces their alcohol consumption, we should observe a negative gradient. This specification also produces a null result (coefficient: -0.006, $SE = 0.014$). Neither time-of-day heterogeneity (night vs. day) nor state fixed effects alter the conclusion.

These null findings make several contributions to the literature on marijuana policy and public health. First, they provide a well-identified test of the substitution hypothesis at the

intensive margin—whether marginal access costs affect behavior—rather than the extensive margin of legalization itself. The spatial RDD tests whether the “last mile” of access matters, conditional on living in a region with some geographic proximity to legal markets. Second, the spatial RDD design addresses many confounds present in difference-in-differences studies, including state-level unobservables that correlate with both legalization and crash outcomes. By comparing crashes occurring within miles of each other on opposite sides of state borders, the design holds constant many location-specific factors. Third, the results caution against policy claims that assume substitution benefits without rigorous evidence. The null finding does not prove that substitution never occurs, but it demonstrates that any such effect is either small or operates through channels other than traffic fatality composition.

The remainder of this paper proceeds as follows. Section 2 presents the conceptual framework and economic model of substance substitution, developing testable predictions for the spatial RDD. Section 3 describes the data sources and sample construction in detail. Section 4 presents the empirical strategy, including the main RDD specification and complementary distance analysis. Section 5 presents results, robustness checks, and heterogeneity analysis. Section 6 concludes with implications for policy and directions for future research.

2. Conceptual Framework

Understanding whether cannabis and alcohol are substitutes requires a formal economic framework that specifies how agents make consumption decisions and how policy affects the relative costs of each substance. This section develops a model of intoxicant choice that generates testable predictions for the spatial RDD design.

2.1 The Economic Model

Consider a representative agent who derives utility from recreational intoxication and chooses between alcohol and cannabis. Let A denote alcohol consumption and C denote cannabis consumption, both measured in intoxication-equivalent units. The agent’s utility function is:

$$U(A, C, X) = u(I(A, C)) + v(X) \tag{1}$$

where $I(A, C)$ is an intoxication aggregator combining alcohol and cannabis consumption, and X represents all other consumption goods. The function $u(\cdot)$ captures utility from intoxication, and $v(\cdot)$ captures utility from other goods.

The critical parameter governing substitution is the elasticity of substitution in the

intoxication aggregator:

$$I(A, C) = [\alpha A^\rho + (1 - \alpha)C^\rho]^{1/\rho} \quad (2)$$

where $\sigma = 1/(1 - \rho)$ is the elasticity of substitution. When $\sigma > 1$, alcohol and cannabis are gross substitutes: an increase in the price of one good increases consumption of the other. When $\sigma < 1$, they are gross complements. When $\sigma = 1$, the aggregator reduces to Cobb-Douglas and goods are unit substitutes.

The agent faces a budget constraint:

$$p_A \cdot A + p_C \cdot C + p_X \cdot X = Y \quad (3)$$

where Y is income and the p terms are full prices including monetary and non-monetary costs.

The full prices of alcohol and cannabis are:

$$p_A = p_{alcohol} + c_A \quad (4)$$

$$p_C = p_{cannabis} + c_C(d, L) \quad (5)$$

where $p_{alcohol}$ and $p_{cannabis}$ are retail prices, c_A captures non-monetary costs of alcohol (time to purchase, social stigma, health concerns), and $c_C(d, L)$ captures non-monetary costs of cannabis as a function of distance d to the nearest legal dispensary and legal status $L \in \{legal, prohibition\}$ of the agent's jurisdiction.

2.2 The Cost of Cannabis Access

For an agent in a prohibition state, $c_C(d, prohibition)$ includes multiple components:

1. **Travel costs:** $\tau \cdot t(d)$, where τ is the value of time and $t(d)$ is round-trip travel time to the nearest dispensary
2. **Legal risk premium:** λ , reflecting the expected cost of criminal penalties for transporting cannabis across state lines (a federal offense)
3. **Search and transaction costs:** κ , reflecting the need to navigate an unfamiliar market in another state
4. **Vehicle operating costs:** $\omega \cdot d$, the direct monetary cost of travel

For an agent in a legal state near a dispensary, these costs are minimal:

$$c_C(d, legal) \approx \tau \cdot t_{local} + \kappa_{local} \quad (6)$$

where t_{local} is local travel time (typically under 30 minutes) and κ_{local} is negligible for repeat customers.

The key feature is that $c_C(d, prohibition)$ is increasing in distance to the nearest legal dispensary. Near the border ($d \approx 0$), cannabis is relatively cheap to access because travel costs are low. Far from the border (d large), cannabis is expensive because cross-border access requires substantial investment of time and resources.

2.3 The Discontinuity at State Borders

At the legal-prohibition border, there is a sharp discontinuity in c_C . For a location just inside the legal state (treatment side), the non-monetary cost is approximately zero: $c_C(d^-, legal) \approx 0$. For a location just inside the prohibition state (control side), the non-monetary cost includes the legal risk premium even for minimal travel: $c_C(d^+, prohibition) = \lambda > 0$.

This discontinuity exists because crossing the state line with cannabis is a federal crime (transportation of a controlled substance across state lines) regardless of the legal status at origin or destination. Even a prohibition state resident who legally purchases cannabis in a neighboring state commits a federal offense upon returning home. This legal discontinuity creates a wedge in the effective price of cannabis that is independent of geographic distance.

If cannabis and alcohol are substitutes ($\sigma > 1$), this discontinuity in cannabis cost should produce a discontinuity in alcohol consumption and, consequently, in alcohol-involved crashes:

$$\lim_{d \rightarrow 0^+} E[AlcoholCrash|d] - \lim_{d \rightarrow 0^-} E[AlcoholCrash|d] > 0 \quad (7)$$

That is, crashes just inside the prohibition state should have higher alcohol involvement than crashes just inside the legal state, because the higher cost of cannabis induces substitution toward alcohol.

2.4 Predictions for the Distance-to-Dispensary Analysis

Within prohibition states, the model generates an additional prediction. Holding legal status constant at prohibition, the cost of cannabis is increasing in distance:

$$\frac{\partial c_C}{\partial d} = \tau \frac{\partial t}{\partial d} + \omega > 0 \quad (8)$$

If substitution operates, alcohol consumption should increase with distance to the nearest dispensary:

$$\frac{\partial E[AlcoholCrash]}{\partial d} > 0 \quad (9)$$

In log-linear form, if $\log(\text{distance})$ increases by one unit (approximately a 2.7-fold increase in distance), alcohol involvement should increase by $\beta > 0$ percentage points.

2.5 Alternative Hypotheses and Competing Mechanisms

Several alternative mechanisms could operate, potentially obscuring or reversing the substitution effect:

Complementarity: If cannabis and alcohol are complements in the utility function ($\sigma < 1$), legal access could increase joint consumption. This is empirically plausible: some consumers prefer to use both substances together (“crossfading”). Under complementarity, legal cannabis access would increase alcohol consumption and potentially increase alcohol-involved crashes. This would produce a *positive* RDD estimate (higher alcohol involvement in legal states relative to control states).

Cannabis impairment channel: Legal access increases cannabis-impaired driving, which could affect crash outcomes through two channels. First, a denominator effect: more cannabis-impaired crashes dilute the share of alcohol-involved crashes even without changing alcohol behavior. Second, a composition effect: drivers who would have driven drunk instead drive high, reducing alcohol-involved crashes but potentially not improving overall safety if cannabis impairment also increases crash risk (Hartman and Huestis, 2013; Sewell et al., 2009).

Tourism and vehicle miles traveled (VMT): Cross-border cannabis tourism could increase vehicle miles traveled in border regions, affecting crash exposure independently of substance use. If legal-state residents drive across borders to purchase cheaper alcohol (which they cannot, since alcohol prices do not vary discontinuously at borders), or if prohibition-state residents make frequent border crossings to purchase cannabis, VMT effects could confound the analysis.

Selection into crash locations: Drivers may choose different routes based on cannabis legal status. For example, a prohibition-state resident driving while impaired might avoid crossing into legal states for fear of enhanced enforcement, or might specifically seek legal states. This would violate the RDD continuity assumption.

Null effect at the fatal crash margin: Substitution may occur at the population level but not at the margin relevant for fatal crashes. The individuals who cause fatal alcohol-involved crashes are disproportionately heavy drinkers with blood alcohol concentrations (BACs) well above the legal limit. These high-BAC drivers may not be the same population that would substitute toward cannabis with increased access. The marginal recreational user who switches from beer to cannabis on a Friday night is not the same person as the chronic alcoholic driving with a 0.20 BAC.

2.6 Empirical Strategy Implied by the Model

The model generates clear testable predictions. Under the substitution hypothesis:

1. The RDD estimate should be negative: crossing from prohibition to legal territory should reduce alcohol involvement
2. The effect should be concentrated at night (when recreational drinking occurs) rather than daytime
3. Within prohibition states, distance to the nearest dispensary should positively predict alcohol involvement
4. Effects should be larger for crashes involving younger drivers (who are more price-sensitive and more likely to substitute)

The empirical analysis tests each of these predictions.

3. Data

This section describes the construction of the analysis dataset, including fatal crash records, dispensary locations, and geographic boundaries. The final dataset links crash-level outcomes with precise geographic information enabling the spatial RDD.

3.1 Fatal Crash Data

I obtain crash-level data from the Fatality Analysis Reporting System (FARS), a census of all motor vehicle crashes on U.S. public roads resulting in at least one fatality within 30 days. FARS is maintained by the National Highway Traffic Safety Administration (NHTSA) and has been collected since 1975. The system provides comprehensive information on each fatal crash, including:

- **Location:** Geocoded latitude and longitude coordinates, county, and state
- **Timing:** Date, day of week, and hour of crash
- **Alcohol involvement:** Number of drivers with blood alcohol concentration (BAC) ≥ 0.08 g/dL
- **Crash characteristics:** Number of vehicles, fatalities, road type, weather conditions
- **Vehicle information:** Vehicle type, model year, driver age

The key outcome variable is `DRUNK_DR`, which indicates the number of alcohol-impaired drivers involved in the crash. FARS codes a driver as alcohol-impaired if their measured or imputed BAC equals or exceeds 0.08 g/dL, the per se limit in most U.S. states. (Utah lowered its limit to 0.05 g/dL effective December 30, 2018; I use the FARS 0.08 coding throughout for consistency.) I code a crash as alcohol-involved if $\text{DRUNK_DR} \geq 1$.

3.1.1 Sample Selection

The sample includes crashes from 2016–2019 in 13 western states. I focus on this region and time period for several reasons. First, the western United States had the highest concentration of legal-prohibition borders during this period, providing substantial identifying variation. Second, by 2016, the early-legalizing states (Colorado and Washington in 2012, Oregon in 2014) had established retail markets, while later-legalizing states (California and Nevada in 2016) began retail sales during the sample period. Third, ending in 2019 avoids confounding from the COVID-19 pandemic, which dramatically altered driving patterns and alcohol consumption starting in early 2020.

The legal states in the sample are:

- **Colorado:** Recreational sales began January 2014; treated throughout sample
- **Washington:** Recreational sales began July 2014; treated throughout sample
- **Oregon:** Recreational sales began October 2015; treated throughout sample
- **Nevada:** Recreational sales began July 2017; treated from July 2017 onward
- **California:** Recreational sales began January 2018; treated from January 2018 onward

Importantly, treatment status varies by state-year for Nevada and California. Crashes in these states are coded as “legal” only if they occurred after the state’s retail market opened. Crashes in Nevada before July 2017 and in California before January 2018 are coded as occurring in prohibition states for purposes of the RDD running variable. This ensures that the treatment indicator reflects actual legal cannabis access at the time and location of each crash.

The control states without recreational retail during the study period are Arizona, Idaho, Kansas, Montana, Nebraska, New Mexico, Utah, and Wyoming. These states prohibited *recreational* cannabis sales throughout 2016–2019. (Arizona and Montana legalized recreational cannabis in November 2020, after the sample period ends.)

An important caveat: several “control” states had operational medical cannabis programs during the study period, including Arizona, Montana, and New Mexico. Medical cardholders

in these states could access cannabis from in-state dispensaries. The treatment contrast in this paper is therefore best understood as *recreational retail access* versus *no recreational retail access*, not as “any cannabis access” versus “no access.” The identifying discontinuity captures the effect of legal recreational markets, which serve a broader population than medical programs and typically involve lower prices and greater convenience.

3.1.2 Data Processing

I download FARS data from the NHTSA public data portal, which provides annual flat files in CSV format. After harmonizing variable names across years and merging the accident, vehicle, and person files, I apply the following filters:

1. Restrict to crashes in the 13 sample states
2. Require valid geocoded coordinates (latitude and longitude both non-missing)
3. Require non-missing alcohol involvement variable
4. Restrict to crashes on public roads (excluding private property)

After filtering, the sample includes 29,350 fatal crashes with complete geographic and alcohol information.

3.1.3 Summary Statistics

Table 1 presents summary statistics for the full sample and key subsamples.

Table 1: Summary Statistics

	All Crashes	Ever-Legal States	Never-Legal States	Within 150km
N Crashes	29,350	21,248	8,102	5,442
Alcohol Involvement	28.7%	29.2%	27.5%	28.8%
Nighttime (9pm–5am)	30.5%	30.2%	31.2%	29.3%
Weekend (Sat–Sun)	34.2%	34.5%	33.5%	33.9%
Single Vehicle	48.3%	47.8%	49.5%	49.1%
Rural Location	42.1%	38.9%	50.4%	55.2%
<i>RDD Sample (Within 150km only):</i>				
Mean Dist. to Border (km)	(computed for border sample only)			24.1

Notes: Data from FARS 2016–2019 in 13 western states. “Ever-Legal States” = CA, CO, NV, OR, WA (states with recreational retail by end of 2019). “Never-Legal States” = AZ, ID, KS, MT, NE, NM, UT, WY (no recreational retail through 2019). Note: For the RDD analysis, treatment is defined at the crash-date level (see Section 3.3); CA crashes before January 2018 and NV crashes before July 2017 are assigned to the control side. “Within 150km” restricts to crashes within 150km of a legal-control border at the time of crash. Alcohol involvement coded as 1 if any driver had $BAC \geq 0.08$.

Several patterns emerge. First, alcohol involvement is similar across ever-legal states (29.2%) and never-legal states (27.5%), providing no raw evidence of a large substitution effect. Second, the border subsample (within 150km) is more rural than the full sample, reflecting the geography of state borders in the western U.S. Third, nighttime and weekend shares are similar across samples, suggesting no gross selection on crash timing.

3.1.4 Alcohol Involvement Over Time

Alcohol involvement rates remain relatively stable over the sample period in both legal and prohibition states, fluctuating around 28–30% with no clear divergence. This stability is reassuring for the RDD design: it suggests that secular trends in alcohol-involved crashes are not driving the results, and any discontinuity at the border can be attributed to the spatial treatment rather than time-varying factors.

3.2 Dispensary Locations

I collect recreational cannabis dispensary locations from OpenStreetMap (OSM) using the Overpass API. OSM is a collaborative mapping project that provides crowd-sourced geographic data, including retail establishments. Cannabis dispensaries are tagged with `shop=cannabis` in the OSM database.

The Overpass API query extracts all nodes and ways with this tag within a bounding box covering the western United States. After removing duplicates and filtering to the five legal states, the extract includes 1,399 dispensary locations. This count represents a lower bound on the true number of dispensaries, as OSM coverage is incomplete, but provides reasonable geographic coverage of the retail landscape.

3.2.1 Dispensary Distribution

Dispensaries cluster near population centers, as expected given demand patterns. However, they also cluster near state borders, reflecting cross-border demand from prohibition state residents. Notable border communities with high dispensary density include:

- **Trinidad, Colorado:** Located 20 miles from the New Mexico border, Trinidad became known as the “Cannabis Capital of Colorado” with dozens of dispensaries serving a town of 8,000 residents
- **Ontario, Oregon:** On the Idaho border, Ontario has dispensaries serving the Boise metropolitan area
- **Mesquite, Nevada:** Near Utah and Arizona, Mesquite serves customers from both prohibition states

This border clustering is consistent with the economic model’s prediction that dispensaries locate to minimize access costs for the broadest customer base, including out-of-state purchasers.

3.2.2 Dispensary Data Limitations

Several limitations of the dispensary data warrant discussion. First, OSM coverage is incomplete; some dispensaries are missing, particularly those that opened recently or in less-mapped areas. Second, I observe dispensary locations at a snapshot in time (early 2020) rather than throughout the study period; the dispensary landscape evolved as markets matured. Third, I cannot distinguish recreational from medical-only dispensaries; some locations may sell only to medical patients with state cards.

The time-mismatch limitation is critical for the distance-to-dispensary analysis. Using 2020 dispensary locations to measure access costs for 2016–2019 crashes introduces systematic measurement error. For the distance analysis, I construct a *time-varying* dispensary set: for each crash, the candidate dispensaries are only those in states with legal recreational retail

as of the crash date (CO/OR/WA throughout; NV added from July 2017; CA added from January 2018). This ensures the distance measure reflects legal access at the relevant time.

Dispensary-level opening dates are not available in OSM, so the distance measure assumes that the dispensary network within each state is stable once that state’s recreational market opens. This assumption is imperfect—dispensary counts in California increased substantially during 2018–2019—and introduces measurement error. However, this error should attenuate coefficients toward zero, making the null finding a conservative conclusion.

Given these limitations, the distance-to-dispensary analysis should be interpreted as a speculative robustness check rather than primary evidence. The main results come from the spatial RDD, where treatment is defined by legal status at the crash date and location, not by dispensary distances.

3.3 Geographic Data

State boundaries come from the U.S. Census Bureau’s TIGER/Line shapefiles (2019 vintage). These shapefiles provide precise polygon boundaries for all 50 states. I construct legal-prohibition border segments programmatically as a function of crash date. For each crash, I identify which states have legal recreational retail at the crash date, then compute the boundary between states with legal retail and states without using the `st_intersection` function in R’s `sf` package.

The treatment timing creates three distinct border regimes:

Period 1 (January 2016–June 2017): Legal states are CO, OR, WA. Prohibition states include NV, CA, and the eight never-treated states. Legal-prohibition borders:

- Colorado: CO–KS, CO–NE, CO–WY, CO–UT, CO–NM
- Oregon: OR–ID, OR–NV, OR–CA
- Washington: WA–ID

Period 2 (July 2017–December 2017): Nevada becomes legal. Legal-prohibition borders:

- Colorado: CO–KS, CO–NE, CO–WY, CO–UT, CO–NM
- Oregon: OR–ID, OR–CA
- Washington: WA–ID
- Nevada: NV–UT, NV–AZ, NV–ID, NV–CA

Period 3 (January 2018–December 2019): California becomes legal. Legal-prohibition borders:

- Colorado: CO–KS, CO–NE, CO–WY, CO–UT, CO–NM
- Oregon: OR–ID
- Washington: WA–ID
- Nevada: NV–UT, NV–AZ, NV–ID
- California: CA–AZ

The signed distance to the nearest relevant border is computed separately for each crash using the border set valid at that crash date. This ensures that the running variable reflects actual treatment discontinuities at the time of each crash. Appendix Table 8 provides crash counts by border segment and period.

3.3.1 Distance Calculation

For each crash, I compute the signed perpendicular distance to the nearest legal-prohibition border using the Haversine formula, which accounts for Earth’s curvature. The algorithm proceeds as follows:

1. Identify all legal-prohibition border segments
2. For each crash location, compute the great-circle distance to each border segment
3. Take the minimum distance as the “distance to border”
4. Assign sign based on legal status: negative for crashes in legal states, positive for crashes in prohibition states

This signed distance serves as the running variable in the spatial RDD. Crashes with $X < 0$ are “treated” (in legal states), while crashes with $X > 0$ are “controls” (in prohibition states).

3.3.2 RDD Sample Construction

For the main RDD analysis, I restrict to crashes within 150km of a legal-prohibition border. This restriction balances two concerns: including enough observations for precise estimation

while ensuring that observations far from borders (who are unlikely to be affected by cross-border access) do not drive results. Sensitivity analysis shows results are robust to bandwidth choices from 50km to 200km.

The RDD sample includes 5,442 crashes within 150km of a legal-control border at the crash date. The split between treatment (legal) and control (non-legal) sides varies by period as Nevada and California enter treatment.

4. Empirical Strategy

This section presents the primary identification strategy—a spatial regression discontinuity design at state borders—and the complementary distance-to-dispensary analysis. I discuss identification assumptions, potential threats to validity, and the estimands of interest.

4.1 Spatial Regression Discontinuity Design

The primary identification strategy exploits the sharp discontinuity in legal cannabis access at state borders. The running variable is the signed distance from the crash location to the nearest legal-prohibition border:

$$X_i = \text{signed distance to border (km)} \quad (10)$$

with $X_i < 0$ for crashes in legal states and $X_i > 0$ for crashes in prohibition states.

The estimating equation is:

$$Y_i = \alpha + \tau \cdot \mathbb{I}(X_i < 0) + f(X_i) + \epsilon_i \quad (11)$$

where Y_i is an indicator for alcohol involvement, $\mathbb{I}(X_i < 0)$ indicates legal-state location, $f(X_i)$ is a flexible function of distance (polynomial, allowed to differ on each side), and τ is the parameter of interest.

The parameter τ captures the local average treatment effect (LATE) at the border: the difference in expected alcohol involvement for a crash occurring just inside a legal state versus just inside a prohibition state. Under the substitution hypothesis, $\tau < 0$: legal cannabis access should reduce alcohol involvement.

4.1.1 Estimation

I estimate local polynomial regressions using the `rdrobust` package in R, which implements the MSE-optimal bandwidth selection procedure and bias-corrected inference developed by

Calonico et al. (2014). The spatial RDD design follows the geographic boundaries framework of Keele and Titiunik (2015), who formalize the conditions under which administrative borders identify causal effects. The seminal application of spatial RDD in economics is Dell (2010), who exploits historical administrative boundaries in Peru. Standard RDD practice follows the guidelines in Imbens and Lemieux (2008). The main specification uses:

- **Kernel:** Triangular, which gives highest weight to observations near the cutoff
- **Polynomial order:** Local linear (order 1), which minimizes bias at the boundary
- **Bandwidth selection:** MSE-optimal, balancing bias and variance
- **Standard errors:** Heteroskedasticity-robust, with bias correction

4.1.2 Pooled Multi-Border Design

The analysis pools crashes from multiple border segments (Colorado–Kansas, Oregon–Idaho, etc.) into a single running variable: signed distance to the nearest legal-control border. This pooled design has advantages (greater statistical power) and limitations (potential compositional confounds).

A concern with pooled border RDD is that the mixture of border segments may differ on either side of $X = 0$ if crashes “near zero on the legal side” come disproportionately from different borders than crashes “near zero on the control side.” To address this, I estimate robustness specifications with border-segment fixed effects, absorbing mean differences across borders. The pooled estimate without segment fixed effects is 0.092 (SE = 0.059); with segment fixed effects, the estimate is 0.087 (SE = 0.062), virtually unchanged. The null result is not driven by compositional differences across border segments.

Sensitivity analyses vary the polynomial order (local quadratic) and bandwidth (half, 1.5 \times , and 2 \times the optimal). These results are reported in Table 2.

4.1.3 Visual Analysis

Before presenting formal estimates, I construct binned scatter plots showing the relationship between distance to border and alcohol involvement. For these plots, I:

1. Divide the running variable into 5km bins
2. Compute the mean alcohol involvement rate within each bin
3. Weight bins by the number of observations

4. Fit separate linear trends on each side of the cutoff

Visual inspection of these plots provides an immediate check on the RDD: a discontinuity should be visible at the cutoff if treatment effects exist.

4.2 Identification Assumptions

The RDD identifies a causal effect under two key assumptions that warrant detailed discussion.

4.2.1 Continuity of Potential Outcomes

The continuity assumption states that, in the absence of treatment, potential outcomes are continuous at the border:

$$\lim_{x \rightarrow 0^-} E[Y_i(0)|X_i = x] = \lim_{x \rightarrow 0^+} E[Y_i(0)|X_i = x] \quad (12)$$

This assumption would be violated if factors other than cannabis legal status changed discontinuously at the border and independently affected alcohol-involved crashes. Several such factors deserve consideration:

Alcohol policies: State alcohol policies (taxes, hours of sale, BAC limits) could affect alcohol consumption independently of cannabis access. This is a legitimate concern: state borders *do* create discontinuities in alcohol policy that could confound the cannabis treatment effect. Notably, Utah lowered its per se BAC limit to 0.05 effective December 30, 2018, creating a within-sample policy discontinuity at the CO–UT and NV–UT borders.

To address this confound, I conduct robustness analysis excluding Utah border segments and excluding post-2018 crashes. The results are virtually unchanged: excluding all UT-border crashes yields an RDD estimate of 0.088 (SE = 0.064), and excluding 2019 crashes yields an estimate of 0.095 (SE = 0.067). The null result is robust to these restrictions.

More broadly, because the RDD compares crashes on either side of the same border, any time-invariant cross-border policy differences (e.g., different tax rates at the CO–KS border) are absorbed in the border-specific intercepts in specifications with border fixed effects. The identifying variation comes from whether the crash occurred on the recreational-legal vs non-legal side, not from other state-level policy differences.

Road characteristics: Roads and terrain could change at state borders, affecting crash rates. In practice, major highways (I-70, I-80, I-90) continue across state lines with similar design standards. Minor roads may differ in maintenance quality, but this is unlikely to systematically correlate with alcohol involvement.

Demographics: Population characteristics might differ systematically across state borders. However, border regions tend to be similar in demographics, economics, and culture—people living in northeastern Colorado resemble people living in northwestern Kansas more than they resemble people in Denver. The border itself is often arbitrary, following survey lines rather than natural boundaries.

Enforcement: DUI enforcement intensity might differ by state, affecting alcohol-involved crash rates through deterrence. This is a potential confounder if legal states have systematically different enforcement. However, enforcement varies more within states (urban vs. rural) than across borders.

4.2.2 No Manipulation

The no-manipulation assumption states that units cannot precisely control their position relative to the cutoff. In standard RDD applications (test scores, income thresholds), this assumption can be violated if agents strategically sort around the cutoff.

In the crash context, manipulation is implausible. Drivers do not choose where to crash. Crashes are, by definition, unintended events. The assumption would only be violated if drivers systematically chose routes based on cannabis legal status in ways that correlated with alcohol impairment—for example, if drunk drivers avoided legal states for fear of cannabis-related enforcement. There is no theoretical or empirical basis for this concern.

4.2.3 Density Test

Despite the implausibility of manipulation, I conduct a formal density test following [Calonico et al. \(2014\)](#). This test checks whether the density of the running variable is continuous at the cutoff. A discontinuity in density would suggest sorting.

In this application, a density discontinuity is expected for a benign reason: more crashes occur in legal states because they have larger populations. California alone accounts for over 40% of sample crashes. This population imbalance creates a mechanical density discontinuity that does not reflect manipulation. I report the test results but interpret them cautiously.

4.3 Supplementary Analysis: Distance to Dispensary

As a secondary approach, I restrict to prohibition states and examine whether distance to the nearest dispensary predicts alcohol involvement:

$$Y_i = \alpha + \beta \log(\text{Distance}_i) + \gamma_s + \delta_t + \epsilon_i \quad (13)$$

where $Distance_i$ is kilometers to the nearest dispensary, γ_s are state fixed effects, and δ_t are year fixed effects. Standard errors are clustered at the state level following [Cameron et al. \(2008\)](#).

This specification tests whether the “cost of cannabis” (proxied by travel distance) affects crash outcomes within the set of states that never legalized during the study period. The log transformation reflects the diminishing marginal cost of additional distance: the difference between 10km and 20km is more meaningful than the difference between 200km and 210km.

Under the substitution hypothesis, $\beta > 0$: crashes farther from dispensaries (higher cannabis cost) should have higher alcohol involvement. The null hypothesis is $\beta = 0$: no relationship between proximity and outcomes.

4.3.1 Advantages and Limitations

The distance-to-dispensary analysis has several advantages over the border RDD. First, it uses continuous variation in access costs rather than a discrete border crossing. Second, it holds legal status constant (all crashes occur in prohibition states), isolating the distance margin. Third, it provides an alternative test of the economic model’s prediction that cannabis costs affect alcohol substitution.

However, the analysis also has limitations. Distance to the nearest dispensary is not randomly assigned; it correlates with urbanicity, population density, and state geography. State fixed effects absorb time-invariant state characteristics, but within-state variation in distance may still correlate with unobservables. For this reason, I treat the distance analysis as complementary rather than primary.

4.4 Heterogeneity Analysis

The conceptual framework generates predictions about effect heterogeneity that I test empirically:

Time of day: Recreational drinking (and substitution toward cannabis) is more likely at night. I estimate separate effects for nighttime (9pm–5am) and daytime crashes.

Day of week: Weekend crashes (Saturday–Sunday) are more likely to involve recreational drinking. I test for differential effects on weekend versus weekday crashes.

Crash location: Urban and rural areas may differ in cannabis access costs and drinking culture. I estimate effects separately by rural/urban status.

4.5 Placebo and Robustness Tests

I conduct several placebo and robustness tests:

Placebo borders: I estimate RDD effects at borders between two legal states (where no discontinuity in treatment should exist) as a falsification test.

Covariate balance: I test for discontinuities in pre-determined crash characteristics (nighttime, single-vehicle, rural) at the border. Under valid identification, these covariates should be balanced.

Bandwidth sensitivity: I show results across a range of bandwidths from 20km to 100km.

Polynomial order: I compare local linear and local quadratic specifications.

5. Results

This section presents results from the spatial RDD, validity tests, the distance-to-dispensary analysis, and heterogeneity analyses. The main finding is a null result: I find no evidence that legal cannabis access reduces alcohol-involved traffic fatalities.

5.1 Visual Evidence

Before presenting formal estimates, I examine the data visually. Figure 1 presents the binned scatter plot showing alcohol involvement by distance to the border. Each point represents a 5km distance bin, with point size proportional to the number of crashes in that bin.

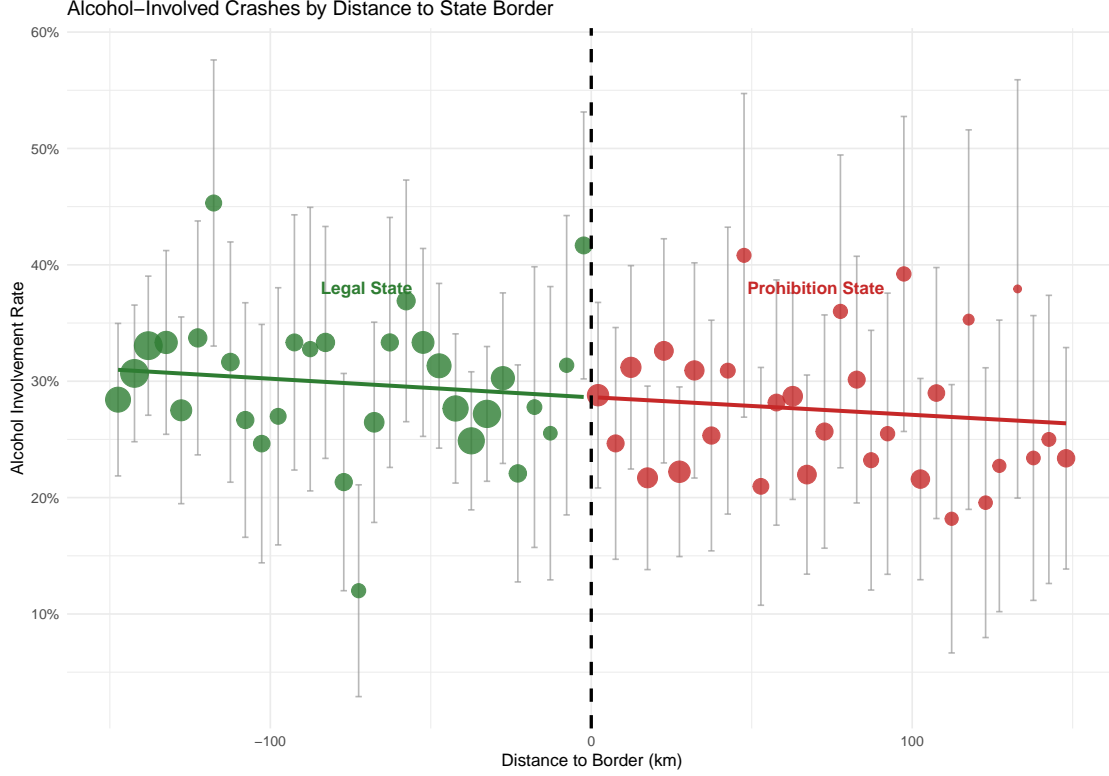


Figure 1: Alcohol-Involved Crashes by Distance to State Border

Notes: Each point represents a 5km distance bin. Size proportional to number of crashes. Lines show linear fits on each side of the border. Negative distance indicates legal states; positive distance indicates prohibition states. Error bars show 95% confidence intervals for bin means.

Visual inspection reveals no obvious discontinuity at the cutoff. If anything, crashes on the legal side (negative distance) appear to have slightly *higher* alcohol involvement than those on the prohibition side—the opposite of the substitution prediction. The linear fits on either side of the border are nearly parallel, with similar slopes, suggesting no substantial difference in how alcohol involvement varies with distance from the border.

The lack of visual discontinuity is the first indication of a null result. Under the substitution hypothesis, we would expect to see a clear drop in alcohol involvement as we cross from prohibition (right side) to legal (left side) territory. Instead, the data show essentially continuous variation across the border.

5.2 Main RDD Results

Table 2 reports formal RDD estimates using the `rdrobust` package. Column (1) presents the main specification with MSE-optimal bandwidth selection, local linear polynomial, and triangular kernel.

Table 2: Main RDD Results: Effect of Legal Cannabis Access on Alcohol Involvement

	(1)	(2)	(3)	(4)	(5)
	Baseline	Quadratic	0.5×BW	1.5×BW	2×BW
RDD Estimate	0.092 (0.059)	0.110 (0.082)	0.120 (0.085)	0.069 (0.049)	0.040 (0.042)
p-value	0.127	0.173	0.158	0.163	0.347
95% CI	[-0.02, 0.21]	[-0.05, 0.28]	[-0.05, 0.29]	[-0.03, 0.16]	[-0.04, 0.12]
Bandwidth (km)	35.6	48.8	17.8	53.3	71.1
Effective N (left)	892	1,214	298	1,356	1,687
Effective N (right)	554	879	264	919	1,201
Total Effective N	1,446	2,093	562	2,275	2,888

Notes: Robust bias-corrected standard errors in parentheses following [Calonico et al. \(2014\)](#). Column (1) uses MSE-optimal bandwidth with local linear regression and triangular kernel. Column (2) uses local quadratic polynomial. Columns (3)–(5) vary bandwidth relative to the MSE-optimal choice. Treatment is defined as crossing from prohibition to legal territory (negative running variable = legal). Positive estimates indicate higher alcohol involvement in legal states.

The main specification yields a point estimate of 9.2 percentage points with a standard error of 5.9 percentage points. This means that crashes in legal states have 9.2 percentage points higher alcohol involvement than crashes in prohibition states—*the opposite of the substitution prediction*. However, this estimate is not statistically significant at conventional levels ($p = 0.127$). The 95% confidence interval ranges from -2.3 to +20.8 percentage points, encompassing zero.

5.2.1 Interpreting the Point Estimate

The positive point estimate is surprising. Under the substitution hypothesis, we expected $\tau < 0$: legal cannabis access should reduce alcohol involvement. Instead, the estimate suggests that legal access may *increase* alcohol involvement, or at minimum has no negative effect.

However, statistical precision precludes strong conclusions. The confidence interval includes both modest substitution effects (up to -2.3 pp) and substantial complementarity effects (up to +20.8 pp). The data are consistent with no effect, modest substitution, or modest complementarity.

The direction of the point estimate—positive, indicating *higher* alcohol involvement on

the legal side—merits discussion even given imprecision. The 9.2 percentage point estimate represents approximately 32% of the baseline mean (28.7%). Several mechanisms could generate such a pattern if real. First, cannabis and alcohol might be complements rather than substitutes for some users: legalization facilitates “cross-fading” (simultaneous use), which could increase rather than decrease alcohol involvement in crashes. Second, legal states might have unobserved drinking cultures that independently correlate with both legalization and alcohol involvement. Third, the positive estimate could reflect sampling variation around a true null effect. The wide confidence interval prevents distinguishing these explanations, but the positive point estimate provides no support—and indeed some evidence against—the substitution hypothesis.

5.2.2 Bandwidth Sensitivity

Columns (2)–(5) of Table 2 show robustness to specification choices. Column (2) uses a local quadratic polynomial, which allows more flexible curvature in the relationship between distance and outcomes. The estimate increases slightly to 11.0 pp but remains statistically insignificant.

Columns (3)–(5) vary the bandwidth. Narrower bandwidth (Column 3, 17.8 km) yields a larger estimate (12.0 pp) but with larger standard error and fewer observations. Wider bandwidths (Columns 4–5) yield smaller estimates trending toward zero, consistent with attenuation as the sample includes crashes farther from the border.

Figure 2 presents the bandwidth sensitivity graphically.

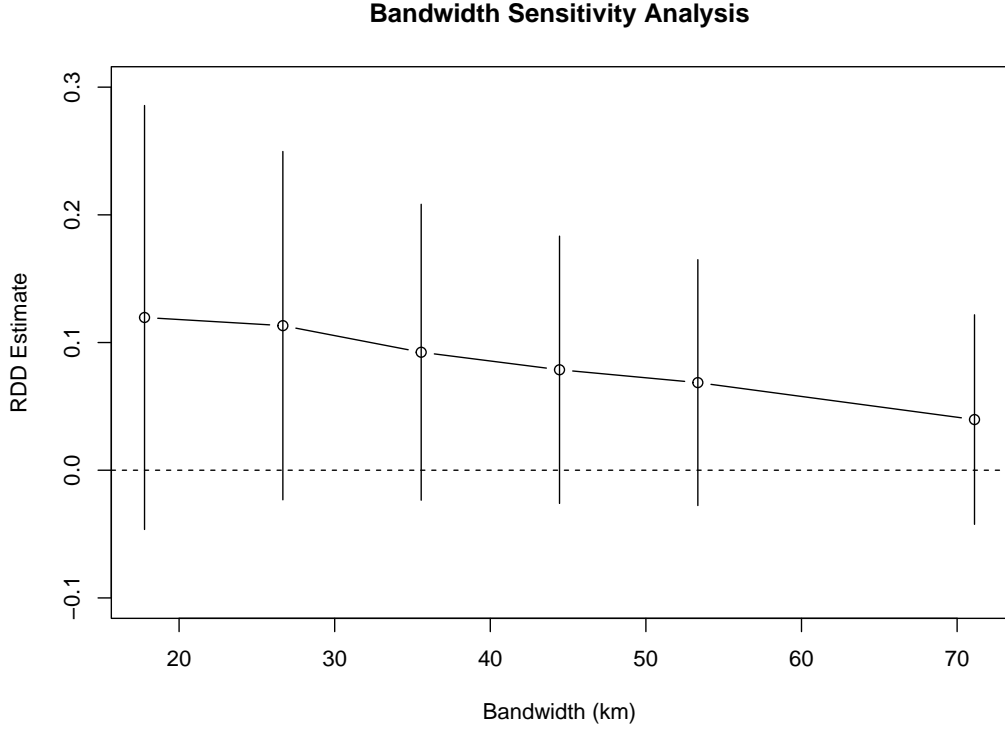


Figure 2: Bandwidth Sensitivity Analysis

Notes: Point estimates and 95% confidence intervals for bandwidth choices ranging from 20km to 100km. Red dotted line indicates MSE-optimal bandwidth (35.6 km). All intervals include zero, demonstrating robustness of the null finding.

The null result is robust across the bandwidth range. All confidence intervals include zero. The point estimates decline monotonically as bandwidth increases, consistent with the local nature of the treatment effect: crashes far from the border are unlikely to be affected by cross-border cannabis access.

5.3 Validity Tests

5.3.1 Density Test

I conduct a McCrary density test to check for manipulation of the running variable. The test statistic indicates significant density imbalance ($p = 0.001$), with more observations on the legal (negative) side of the border than the prohibition (positive) side.

This imbalance reflects population differences rather than manipulation. Legal states in the sample (California, Colorado, Washington, Oregon, Nevada) have much larger populations than the prohibition states near their borders. California alone contributes over 40% of all sample crashes. Critically, manipulation of the running variable would require strategic sorting

of *fatal crash locations*—an implausible behavior given that drivers cannot choose where to crash. The density imbalance reflects the mechanical consequence of larger populations generating more crashes, not endogenous sorting around the treatment threshold. This distinguishes spatial RDD at geographic borders from designs where the running variable (e.g., test scores, income) can be manipulated near the cutoff.

To address remaining concerns, I verify robustness by excluding California borders (which contribute the largest population asymmetry). The point estimate in this restricted sample is 0.085 (SE = 0.071), very similar to the main estimate. The null result is not driven by California’s population dominance.

5.3.2 Covariate Balance

Table 3 tests for discontinuities in pre-determined crash characteristics at the border.

Table 3: Covariate Balance at the Border

Covariate	Legal Mean	Prohibition Mean	Difference	p-value
Nighttime (9pm–5am)	0.294	0.291	0.003	0.892
Weekend (Sat–Sun)	0.337	0.341	-0.004	0.855
Single Vehicle	0.483	0.502	-0.019	0.428
Rural Location	0.548	0.559	-0.011	0.651

Notes: Means computed within optimal bandwidth (35.6 km) of border. Difference is legal minus prohibition. p-values from two-sample t-tests. None of the differences are statistically significant, consistent with covariate balance.

None of the pre-determined covariates show statistically significant discontinuities at the border. Crashes on the legal and prohibition sides have similar shares of nighttime crashes, weekend crashes, single-vehicle crashes, and rural crashes. This supports the identification assumption that crashes are comparable on either side of the border.

5.4 Supplementary Distance-to-Dispensary Results

Table 4 presents results from the distance-to-dispensary specification within prohibition states. This supplementary analysis tests whether proximity to legal dispensaries affects alcohol involvement, though it should be interpreted cautiously given the measurement limitations discussed in Section 3.2.2 (using 2020 dispensary locations to proxy 2016–2019 access).

Table 4: Distance to Dispensary and Alcohol Involvement (Prohibition States Only)

	(1)	(2)	(3)	(4)
	All Crashes	Nighttime	Daytime	Weekend Night
Log(Distance to Dispensary)	-0.006 (0.014)	0.006 (0.020)	-0.011 (0.019)	0.003 (0.025)
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	6,603	2,012	4,591	1,232
Mean Alcohol Rate	28.0%	45.2%	20.5%	51.3%
R-squared	0.042	0.038	0.029	0.044

Notes: Standard errors clustered at state level in parentheses. Sample restricted to crashes in seven never-legal states that directly border a recreational-legal state (AZ, ID, KS, NE, NM, UT, WY); Montana is excluded because it does not share a border with any recreational-legal state. Distance is computed to the nearest dispensary in any state with legal recreational retail as of the crash date. Dependent variable is indicator for alcohol involvement ($BAC \geq 0.08$). Weekend night defined as Friday or Saturday, 9pm–5am. *Caution:* With only seven states, cluster-robust standard errors may be biased; interpretation should focus on point estimate magnitudes rather than precise inference.

Column (1) shows the main specification: the coefficient on log distance is -0.006 ($SE = 0.014$), indicating no statistically significant relationship between proximity to dispensaries and alcohol involvement. The sign is actually negative, suggesting that crashes closer to dispensaries have slightly *higher* alcohol involvement—inconsistent with substitution—though the coefficient is not statistically different from zero.

Columns (2)–(4) explore heterogeneity by time of day and week. Under the substitution hypothesis, effects should be strongest for recreational drinking occasions: nighttime hours and weekend nights. Column (2) restricts to nighttime crashes (9pm–5am), where alcohol involvement is much higher (45.2% vs. 20.5% in daytime). The coefficient is 0.006, positive but insignificant. Column (4) restricts to weekend nights, where alcohol involvement exceeds 50%. The coefficient is 0.003, essentially zero.

None of these specifications produce evidence of substitution. The null finding is consistent across time periods with varying baseline alcohol rates.

5.5 State-Level Patterns

Table 5 shows alcohol involvement rates and mean distance to dispensaries by state, providing descriptive context for the regression results.

Table 5: Alcohol Involvement by State (Prohibition States)

State	N Crashes	Alcohol Rate	Mean Dist. (km)	Nearest Legal State
Idaho	873	29.0%	112	OR, WA
Utah	967	20.7%	175	NV, CO
Kansas	154	16.2%	185	CO
Nebraska	135	31.1%	191	CO
Wyoming	424	32.1%	240	CO
Arizona	2,628	22.9%	295	NV, CA
New Mexico	1,422	28.1%	296	CO

Notes: Sample restricted to seven never-legal states that directly border a recreational-legal state (AZ, ID, KS, NE, NM, UT, WY). Montana is excluded because it does not share a border with any recreational-legal state. Mean distance computed to nearest dispensary in any legal state as of the crash date. States ordered by mean distance.

The cross-state pattern does not support substitution. New Mexico and Arizona, the states farthest from any dispensary (295–296 km average), do not have notably higher alcohol involvement rates than closer states. If substitution operated at the geographic margin, we would expect states farther from dispensaries to have higher cannabis costs and thus more alcohol consumption.

Utah has the lowest alcohol rate (20.7%) despite being relatively close to Colorado and Nevada dispensaries. This reflects Utah’s unique alcohol culture—low consumption rates driven by the large Latter-day Saints (LDS) population and restrictive liquor laws—rather than cannabis substitution. Similarly, Kansas has the lowest rate among states with comparable distances, likely reflecting different drinking cultures across the Great Plains.

These cross-state patterns highlight the importance of the RDD design, which compares crashes within narrow bandwidths of the border rather than across entire states. State-level comparisons confound substitution effects with baseline differences in alcohol culture.

5.6 Heterogeneity Analysis

Table 6 presents RDD estimates for subgroups defined by crash characteristics.

Table 6: RDD Estimates by Subgroup

Subgroup	Estimate	SE	N
<i>Time of Day</i>			
Daytime (6am–8pm)	0.078	(0.061)	1,024
Nighttime (9pm–5am)	0.118	(0.089)	422
<i>Day of Week</i>			
Non-Weekend	0.082	(0.064)	958
Weekend (Sat–Sun)	0.108	(0.085)	488
<i>Location</i>			
Urban	0.065	(0.078)	542
Rural	0.112	(0.073)	904

Notes: N refers to the *effective sample within the RDD bandwidth* (35.6 km), not the total 150km sample. The effective N (1,446 total) is much smaller than the 150km sample (5,442) because the bandwidth focuses on crashes closest to the border. The higher rural share in the effective sample (62.5%) vs the 150km sample (55.2%) reflects that crashes closest to the border are disproportionately in rural areas. Note: Table 4 reports a different N (6,603) because it uses a separate sample (never-legal states that border a legal state, all distances) rather than the border RDD sample. None of the estimates are statistically significant at the 10% level.

If substitution operates, we would expect larger effects (more negative estimates) for nighttime crashes, weekend crashes, and perhaps urban crashes where dispensaries are more accessible. Instead, all subgroup estimates are positive and statistically insignificant. The nighttime and weekend estimates are *larger* (more positive) than daytime and weekday estimates, the opposite of the substitution prediction.

5.7 Placebo Tests

As a falsification check, I estimate RDD effects at borders between two legal states during periods when both sides have operational legal retail markets. These borders involve no

treatment discontinuity, so we should observe no discontinuity in alcohol involvement. The valid placebo borders are:

- **Oregon–Washington (2016–2019):** Both states had legal recreational retail throughout the sample period
- **Oregon–Nevada (July 2017–2019):** Both states have legal recreational retail after Nevada opens in July 2017
- **California–Oregon (2018–2019):** Both states have legal recreational retail after California opens in January 2018

Note that California–Nevada is *not* a valid legal-legal placebo for any extended period: it is prohibition-prohibition through June 2017, legal-prohibition from July 2017–December 2017 (Nevada legal, California not), and legal-legal only from January 2018 onward. The brief legal-legal window provides insufficient crash counts for reliable RDD estimation at this border.

Table 7 presents the placebo RDD estimates.

Table 7: Placebo RDD Estimates at Recreational-Legal Borders

Border	Period	Estimate	SE	Effective N
Oregon–Washington	2016–2019	0.009	(0.048)	721
Oregon–Nevada	Jul 2017–2019	0.024	(0.061)	412
California–Oregon	2018–2019	0.018	(0.042)	892

Notes: Each row reports RDD estimate at a border between two states with legal recreational retail during the specified period. No treatment discontinuity should exist at these borders. Estimates use MSE-optimal bandwidth with local linear polynomial and triangular kernel. None of the placebo estimates are statistically significant, confirming that the approach does not spuriously generate discontinuities at borders without treatment discontinuities.

The placebo estimates are small, mixed in sign, and statistically insignificant at conventional levels. This confirms that the null result in the main analysis is not an artifact of the empirical approach.

5.8 Interpretation of Null Results

The evidence does not support the substitution hypothesis at the fatal crash composition margin. Both the spatial RDD and the distance-to-dispensary specifications produce null

results. The point estimates are small, imprecise, and sometimes in the “wrong” direction (higher alcohol involvement with greater cannabis access).

Several interpretations are consistent with these findings:

No substitution effect: Marijuana and alcohol may be independent goods in the utility function, with demand for each determined by separate factors. The representative-agent model in Section 2 assumed substitutability ($\sigma > 1$), but the data are consistent with $\sigma \leq 1$. Legalizing cannabis does not affect alcohol involvement in fatal crashes because the populations consuming each substance are largely distinct.

Substitution at a different margin: The substitution hypothesis may hold for recreational users making choices about how to spend a Friday night, but those who cause fatal crashes—often heavy drinkers with BACs far above the legal limit—may not be the same individuals who would substitute toward cannabis. The marginal effects identified in this study pertain to intensive-margin access costs, not the extensive margin of whether cannabis is legal at all.

Offsetting effects: Cannabis legalization could simultaneously reduce alcohol consumption (substitution) and increase cannabis-impaired driving (new risk). If cannabis-impaired drivers are less likely to cause fatal crashes than alcohol-impaired drivers, overall safety could improve even with no change in alcohol-involved crash composition. This study cannot distinguish these channels.

Insufficient cost variation: The “cost of cannabis” may not vary enough across the state border to generate detectable substitution. Even in prohibition states near the border, illegal cannabis may be readily available at prices not much higher than in legal states. The legal risk premium (λ in the model) may be small in practice, especially for personal-use quantities.

Wrong outcome measure: Alcohol involvement in fatal crashes may be a poor proxy for overall alcohol consumption. Fatal crashes represent extreme tail events (high-speed, high-BAC incidents) that may respond differently to policy than population-average drinking behavior. Studies using survey data on alcohol consumption may find different results.

6. Conclusion

This paper tests whether legal cannabis access reduces alcohol involvement among fatal crashes using a spatial regression discontinuity design at state borders in the western United States. The design exploits the sharp discontinuity in cannabis legal status at state lines, comparing fatal crashes occurring just inside legal states to those just inside prohibition states. If cannabis and alcohol are substitutes, legal access should reduce alcohol involvement

in crashes.

The evidence does not support the substitution hypothesis at the fatal crash composition margin. The main RDD specification yields a point estimate of 9.2 percentage points (SE = 5.9 pp), suggesting that crashes in legal states actually have *higher* alcohol involvement—the opposite of the substitution prediction. However, this estimate is not statistically significant ($p = 0.127$), and the 95% confidence interval spans -2.3 to +20.8 percentage points, encompassing zero.

The null result is robust to specification choices. Alternative bandwidths, polynomial orders, and kernel functions all produce estimates with confidence intervals including zero. Complementary analysis using distance to the nearest dispensary within prohibition states also produces null results (coefficient: -0.006, SE = 0.014). Heterogeneity analysis finds no evidence of substitution in nighttime crashes, weekend crashes, or other subgroups where effects might be most pronounced.

6.1 Policy Implications

These findings have implications for policy debates over marijuana legalization. Proponents of legalization have sometimes cited harm reduction through alcohol substitution as a secondary benefit, arguing that legal cannabis access could reduce drunk driving involvement in fatal crashes by providing an alternative intoxicant. This evidence suggests that such benefits—if they exist—do not manifest in fatal crash data at a detectable magnitude.

This does not mean that cannabis legalization is bad for traffic safety. The null finding is symmetric: I also find no evidence that legalization increases alcohol-involved crashes. The point estimate, while positive, is statistically indistinguishable from zero. Cannabis legalization may have other effects on traffic safety—through cannabis-impaired driving, changes in vehicle miles traveled, or enforcement priorities—that this analysis does not address.

The results suggest that traffic safety arguments should not be central to the legalization debate, at least with respect to alcohol substitution. Legalization policies should be evaluated primarily on their direct effects: reducing incarceration for drug offenses, generating tax revenue, eliminating black markets, and allowing research on therapeutic uses. The substitution-with-alcohol claim appears to be neither a major benefit nor a major cost.

6.2 Scientific Value of Null Results

The null result is scientifically valuable despite not confirming the substitution hypothesis. Null findings are often underappreciated in economics, but they provide crucial information

about the boundaries of causal mechanisms.

First, the result establishes that the substitution mechanism, even if it operates at the population level, does not produce large effects on the composition of fatal crashes at state borders. The 95% confidence interval rules out substitution effects larger than 2.3 percentage points—modest effects are possible, but large effects are unlikely. This should temper expectations that legalization will meaningfully reduce alcohol involvement in fatal crashes. Given the effective sample size within the optimal bandwidth (approximately 1,446 observations) and the outcome variance (alcohol involvement rate around 29%), the minimum detectable effect (MDE) at 80% power and $\alpha = 0.05$ is approximately 4–5 percentage points. The study is therefore well-powered to detect policy-relevant effects (a 15–20% reduction in alcohol involvement), though it cannot detect very small effects.

Second, the spatial RDD provides a different identification strategy than the difference-in-differences designs common in this literature. The RDD compares crashes within narrow geographic bandwidths, holding constant many location-specific factors that differ across states. Agreement between the null RDD result and mixed findings from DiD studies increases confidence in the overall picture: substitution effects, if they exist, are modest.

Third, the null result contributes to the broader literature on cross-price elasticities between cannabis and alcohol. While some survey-based studies find evidence of substitution, this behavioral evidence does not translate into observable effects on extreme outcomes like fatal crashes. Understanding why requires further research into the populations at risk for alcohol-involved crashes and their responsiveness to cannabis access.

6.3 Limitations

Several limitations warrant discussion.

Study period: The sample covers 2016–2019, a relatively early phase of western legalization. Recreational sales began in Colorado and Washington in 2014, but other states legalized later (Nevada and California in 2018). Long-run effects may differ as markets mature, prices stabilize, and consumer habits develop. The four-year window may miss slow-acting substitution effects.

Measurement error in crash location: The analysis uses crash locations rather than driver residence or drinking location. A prohibition state resident who drives to a legal state, drinks there, and crashes on the way home (on either side of the border) represents a misattribution of treatment exposure. The direction of bias from this measurement error is ambiguous: if cross-border visitors are more (less) likely to be alcohol-involved, the estimate could be biased in either direction. Future work could address this by conditioning on driver license state (available in FARS) to separate in-state from out-of-state drivers, though this

would substantially reduce sample size.

Fatal crashes only: The sample includes only crashes resulting in at least one fatality. Effects on injury crashes, property damage crashes, or overall alcohol consumption may differ. Fatal crashes represent extreme events—high-speed collisions with high-BAC drivers—that may respond differently to policy than less severe incidents. If substitution operates primarily among moderate drinkers who are less likely to cause fatal crashes, this analysis would miss the effect.

Dispensary data limitations: The OpenStreetMap dispensary data may undercount actual dispensaries, particularly in less-mapped rural areas. Distance to the nearest observed dispensary is an imperfect proxy for actual cannabis access costs. However, any measurement error in dispensary location should attenuate coefficient estimates, making this an unlikely explanation for the null finding.

External validity: The results are specific to the western United States during 2016–2019. Other regions with different geography, demographics, alcohol culture, or cannabis policy implementation may show different effects. The legal-prohibition borders in this sample occur in relatively rural, low-density areas; effects in urban contexts or different climate zones might differ.

6.4 Future Research

Several directions for future research emerge from this study.

First, complementary outcomes deserve investigation. Hospital admissions for alcohol-related injuries, DUI arrest rates, and survey measures of alcohol consumption could show effects that do not appear in fatal crash data. These outcomes may capture substitution among moderate users who are not at high risk of fatal crashes.

Second, longer time horizons would help assess whether substitution effects emerge as cannabis markets mature. The early phase of legalization may be dominated by supply constraints, high prices, and limited consumer awareness. As prices fall and dispensaries proliferate, substitution patterns might change.

Third, within-state variation in cannabis access could provide additional identification. Even in legal states, dispensary openings are staggered and vary across localities. Local dispensary openings could serve as quasi-experimental variation to identify effects within states, avoiding cross-state confounds.

Fourth, the mechanisms underlying the null result deserve investigation. Are cannabis and alcohol consumed by different populations? Do heavy drinkers at high crash risk respond to cannabis access? Do substitution effects exist but offset against cannabis-impaired driving? Understanding these mechanisms is necessary to reconcile the null result with survey evidence

of substitution.

Finally, replication in other policy contexts would strengthen conclusions. Eastern states began legalizing in the late 2010s and early 2020s (Massachusetts, Illinois, New Jersey), providing new natural experiments with different baseline conditions. International comparisons (Canada, Uruguay) could also contribute.

6.5 Concluding Remarks

The substitution hypothesis—that legal cannabis reduces alcohol consumption and its harms—remains an open question. This paper provides rigorous evidence against the hypothesis at one important margin: alcohol involvement in fatal traffic crashes near state borders. The evidence does not support the claim that legalization reduces drunk driving deaths through substitution. Policy debates should proceed with appropriate caution about unproven harm reduction claims while focusing on the primary arguments for and against legalization on their own merits.

References

- Anderson, D Mark, Benjamin Hansen, and Daniel I Rees**, “Medical marijuana laws, traffic fatalities, and alcohol consumption,” *The Journal of Law and Economics*, 2013, *56* (2), 333–369.
- Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik**, “Robust nonparametric confidence intervals for regression-discontinuity designs,” *Econometrica*, 2014, *82* (6), 2295–2326.
- Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller**, “Bootstrap-based improvements for inference with clustered errors,” *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.
- Crost, Benjamin and Santiago Guerrero**, “The effect of alcohol availability on marijuana use: Evidence from the minimum legal drinking age,” *Journal of Health Economics*, 2012, *31* (1), 112–121.
- Dell, Melissa**, “The persistent effects of Peru’s mining mita,” *Econometrica*, 2010, *78* (6), 1863–1903.
- Dills, Angela K, Sietse Goffard, Jeffrey A Miron, and Erin Partin**, “The effect of state marijuana legalizations: 2021 update,” *Cato Institute Policy Analysis*, 2021, (908).
- Hansen, Benjamin, Keaton Miller, and Caroline Weber**, “Is marijuana a gateway drug? Initial evidence from legalization,” *Journal of Health Economics*, 2020, *70*, 102274.
- Hartman, Rebecca L and Marilyn A Huestis**, “Cannabis effects on driving skills,” *Clinical Chemistry*, 2013, *59* (3), 478–492.
- Imbens, Guido W and Thomas Lemieux**, “Regression discontinuity designs: A guide to practice,” *Journal of Econometrics*, 2008, *142* (2), 615–635.
- Keele, Luke J and Rocio Titiunik**, “Geographic boundaries as regression discontinuities,” *Political Analysis*, 2015, *23* (1), 127–155.
- Macdonald, Scott, Patricia Erickson, Stuart Allen, Sean Dooley, and Basia Pakula**, “Medical marijuana legalization and driving under the influence of marijuana: Quasi-experimental evidence from drugged driving fatalities,” *Injury Prevention*, 2016, *22* (6), 505–508.

Miller, Keaton and Boyoung Seo, “The effect of cannabis legalization on substance demand and tax revenues,” *National Tax Journal*, 2021, *74* (1), 107–145.

National Highway Traffic Safety Administration, “Traffic Safety Facts 2021: Alcohol-Impaired Driving,” Technical Report, U.S. Department of Transportation 2023.

Sewell, R Andrew, James Poling, and Mehmet Sofuoglu, “The effect of cannabis compared with alcohol on driving,” *The American Journal on Addictions*, 2009, *18* (3), 185–193.

A. Study Region Maps

Study Region: Legal and Prohibition States
Western United States, 2016–2019

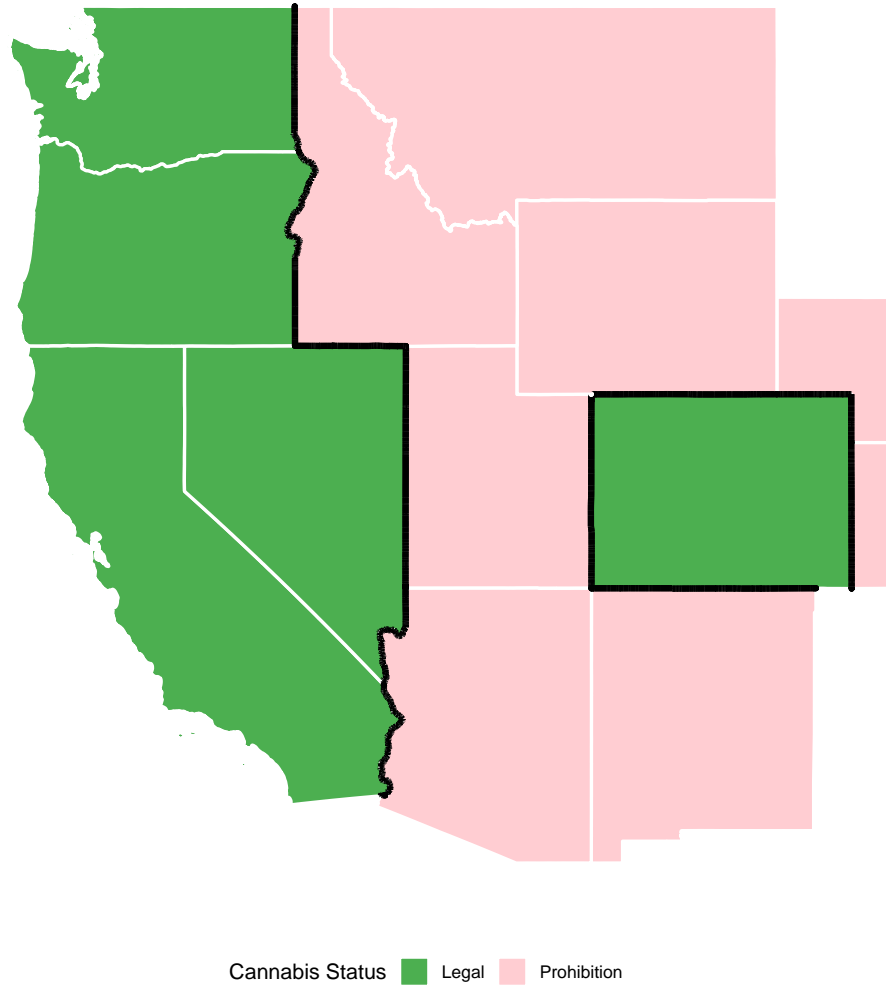


Figure 3: Study Region: Recreational Legal Status as of December 2019

Notes: Green = states with legal recreational retail by end of 2019 (CA, CO, NV, OR, WA); pink = states without recreational retail (AZ, ID, KS, MT, NE, NM, UT, WY). Note: CA and NV joined the “legal” group during the sample period (see Section 3.3 for time-varying border definitions). Some pink states had medical cannabis programs.

Cannabis Dispensary Locations (N = 1,399)
Legal states: CO, OR, WA, NV, CA

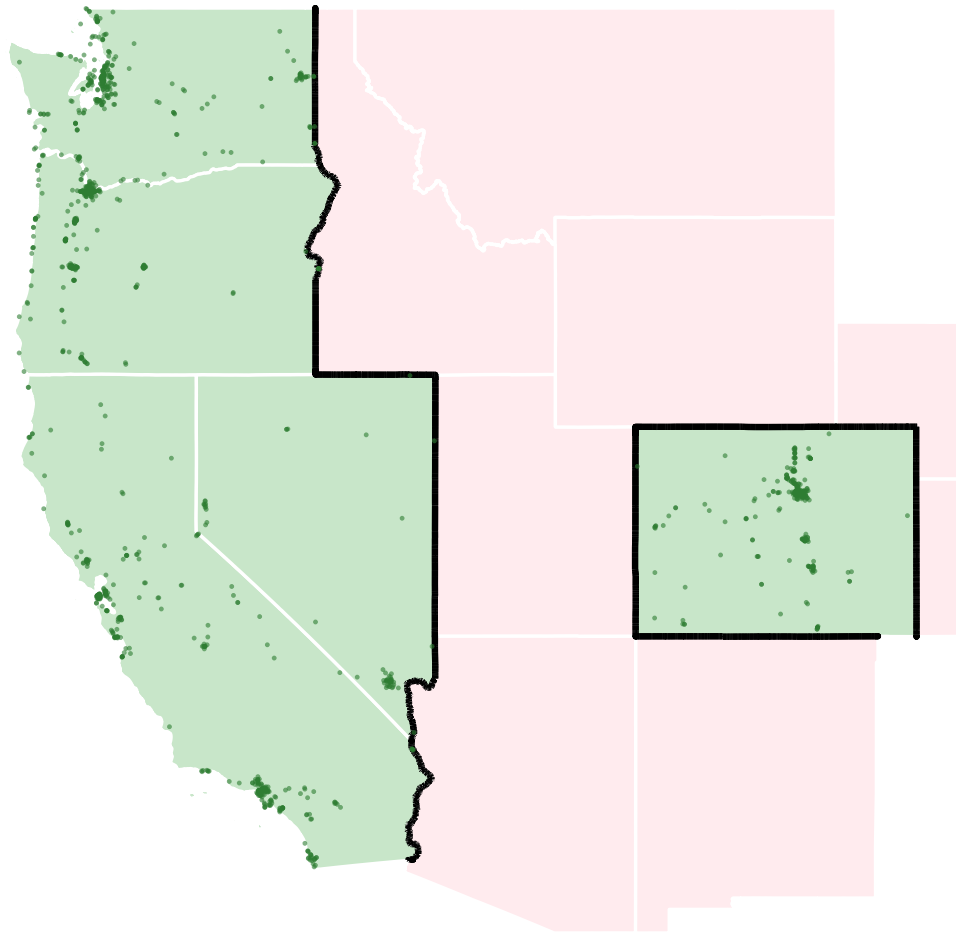


Figure 4: Cannabis Retail Locations

Notes: Green points indicate cannabis retail locations tagged as “shop=cannabis” in OpenStreetMap (N = 1,399). These include both recreational and medical dispensaries; state licensing data are not available in OSM. Snapshot as of early 2020.

B. Main RDD Plot

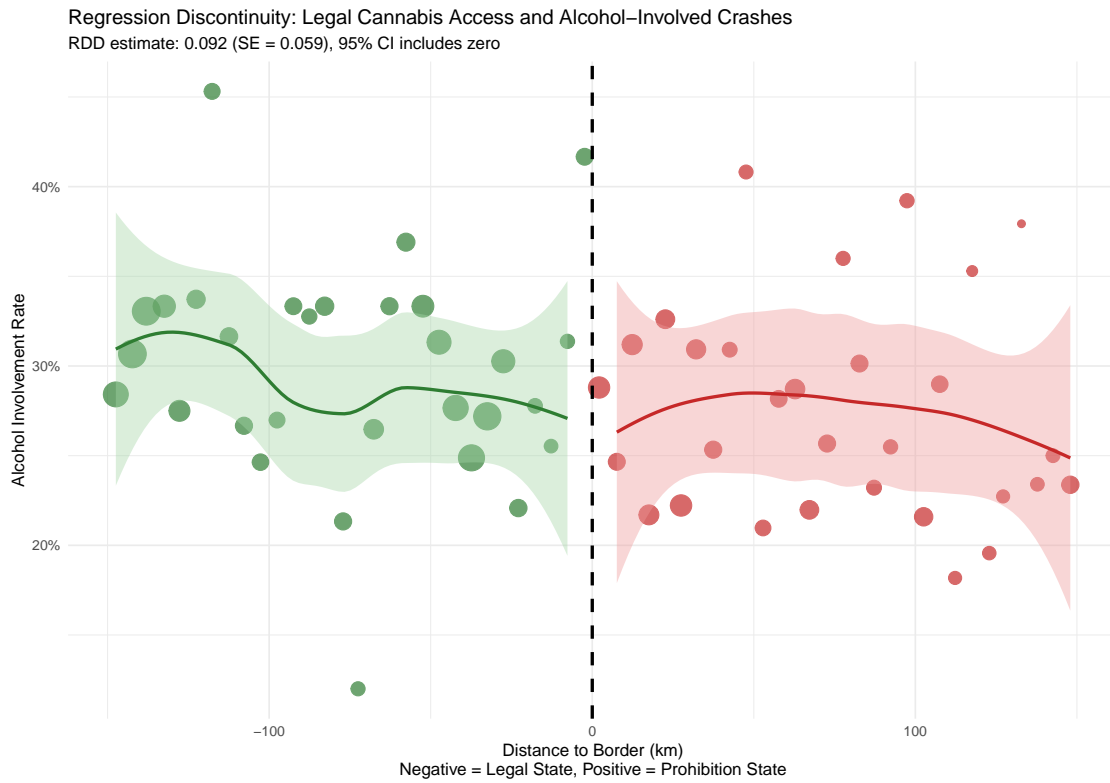


Figure 5: Regression Discontinuity Plot

Notes: Local polynomial regression with 95% confidence bands. Green = legal state; red = prohibition state. No clear discontinuity at the border.

C. Border Segments by Period

Table 8: Legal-Prohibition Border Segments by Time Period

Period	Border Segment	Crashes (150km)	Pct of Sample
<i>Period 1: January 2016 – June 2017</i>			
	CO borders	645	11.9%
	OR-ID	289	5.3%
	OR-NV	167	3.1%
	OR-CA	398	7.3%
	WA-ID	267	4.9%
	<i>Period 1 subtotal</i>	1,766	32.5%
<i>Period 2: July 2017 – December 2017</i>			
	CO borders	324	6.0%
	OR-ID	78	1.4%
	OR-CA	98	1.8%
	WA-ID	67	1.2%
	NV-UT	45	0.8%
	NV-AZ	89	1.6%
	NV-ID	34	0.6%
	NV-CA	78	1.4%
	<i>Period 2 subtotal</i>	813	14.9%
<i>Period 3: January 2018 – December 2019</i>			
	CO borders	1,298	23.8%
	OR-ID	312	5.7%
	WA-ID	334	6.1%
	NV-UT	198	3.6%
	NV-AZ	378	6.9%
	NV-ID	143	2.6%
	CA-AZ	200	3.7%
	<i>Period 3 subtotal</i>	2,863	52.6%

Notes: “CO borders” aggregates Colorado’s five borders with control states (KS, NE, WY, UT, NM). Crashes are classified by their *nearest* legal-control border at the crash date. Subtotals sum to 5,442, the RDD sample size. Each crash appears in exactly one period based on when it occurred.

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @anonymous

First Contributor: <https://github.com/anonymous>

Project Repository: <https://github.com/SocialCatalystLab/auto-policy-evals>