Legal Status vs. Physical Access: Testing the Cannabis-Alcohol Substitution Hypothesis at State Borders[*]

APEP Autonomous Research[†]

@anonymous

January 2026

Abstract

Does access to legal cannabis reduce alcohol involvement among fatal traffic crashes through substance substitution? Using a spatial regression discontinuity design at state borders in the western United States, I test whether cannabis access affects alcohol-involved crashes. The standard RDD yields a null result (9.2 pp, SE = 5.9, p = 0.127). Suggestive evidence from dispensary distance analysis—using 2020-era location data as a proxy for the 2016–2019 study period—indicates that physical cannabis access may not change sharply at borders, as prohibition-state residents near borders can cross to purchase; however, this first-stage evidence is approximate given the temporal mismatch between dispensary data and crash outcomes. To address this weak first-stage critique, I restrict to single-vehicle crashes with in-state drivers, where crash location, driver residence, and outcome attribution are all unambiguous. This specification produces a null result: −5.2 pp (SE = 11.4, p = 0.649). Donut RDD robustness checks reveal some specification sensitivity—the 2km donut yields a significant positive estimate (23.7 pp, SE = 8.2)—but border-by-border decomposition shows no individual border drives this result, and larger donuts return to null. Strikingly, cross-border analysis reveals that prohibition-state residents who crash in legal states have *lower* alcohol involvement (21.6%) than those who crash at home (31.0%), suggesting compositional differences in who crosses borders rather than treatment effects. The baseline and most robustness specifications produce null results, though specification sensitivity warrants caution in interpretation.

---

1. Introduction

In Trinidad, Colorado—a former mining town of 8,000 residents nestled against the New Mexico border—over two dozen cannabis dispensaries have opened since recreational legalization in 2014. The town's transformation into the "Cannabis Capital of Colorado" reflects a striking economic reality: much of the demand comes not from local residents, but from New Mexicans driving 20 miles north across the state line. This cross-border shopping phenomenon raises a fundamental question for policy research: when borders are this porous, does *de jure* legal status actually affect behavior, or do residents of prohibition states already have *de facto* access?

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; Crost and Guerrero, 2012), while others find null or complementary relationships (Macdonald et al., 2016; Dills et al., 2021; Miller and Seo, 2021). Recent work specifically examining traffic safety outcomes includes Santaella-Tenorio et al. (2017) on medical marijuana laws and traffic fatalities, and Aydelotte et al. (2017) on crash fatality rates after recreational legalization in Washington and Colorado. A meta-analysis by Rogeberg and Elvik (2016) revisits the evidence on cannabis intoxication and motor vehicle collision risk, finding modest but statistically significant increases in crash probability associated with cannabis use.

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. A related complication is cross-border access: the literature on cigarette smuggling (Lovenheim, 2008) and gun trafficking (Knight, 2013) documents that consumers circumvent local restrictions by purchasing in nearby jurisdictions with more permissive laws. If cannabis consumers in prohibition states can easily cross into legal states, then *de jure* legal status may not correspond to *de facto* access.

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 -2.0 to +21.0 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, I distinguish between two conceptually distinct treatment margins: *legal status* (the de jure law governing cannabis in a jurisdiction) and *physical access* (the de facto cost of obtaining cannabis). The cross-border shopping literature (Lovenheim, 2008; Knight, 2013) demonstrates that these margins can diverge when borders are porous. By using driver license state information from FARS, I can define treatment based on driver *residence* rather than crash *location*, testing whether habitual access patterns—determined by where a driver lives and regularly obtains cannabis—affect alcohol involvement differently than the legal status of where a crash happens to occur.

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.

Fourth, the driver residency analysis addresses the "weak first stage" problem directly. If physical access is continuous at borders (because prohibition-state residents near the border have easy access via day trips), then an RDD using crash location may be uninformative

about substitution. The residence-based analysis creates a cleaner treatment contrast: a Colorado resident has low-cost, legal access to cannabis regardless of where they happen to crash, while a Wyoming resident faces legal risk and search costs even if they live near the border. The consistent null result across both specifications strengthens the conclusion.

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:

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

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.

6

The agent faces a budget constraint:

$$p_A \cdot A + p_C \cdot C + p_X \cdot X = Y \tag{3}$$

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:

$$p_A = p_{alcohol} + c_A \tag{4}$$

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

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:

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

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:

$$\lim_{d \to 0^+} E[AlcoholCrash|d] - \lim_{d \to 0^-} E[AlcoholCrash|d] > 0 \tag{7}$$

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:

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

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

$$\frac{\partial E[AlcoholCrash]}{\partial d} > 0 \tag{9}$$

In log-linear form, if $\log(distance)$ 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

The empirical analysis tests 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) $\geq$ 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 `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%

Border subsample statistics:

Mean Dist. to Border (km)	24.1 (computed for "Within 150km" sample only)

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 (8,102 total crashes, including 1,499 in Montana; 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; distance-to-border is computed only for this subsample (not applicable to other columns). The distance-to-dispensary analysis (Tables 4–5) uses a *different* subsample of 6,603 crashes from seven never-legal states that *eventually* border a legal state by 2019 (AZ, ID, KS, NE, NM, UT, WY), excluding Montana (1,499 crashes) because it does not share a border with any recreational-legal state. Note: Arizona borders NV and CA, which become legal in mid-2017 and 2018 respectively; distance calculations use time-varying legal status. Alcohol involvement coded as 1 if any driver had BAC ≥ 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. Dispensary networks expanded substantially during this period—particularly in California after 2018 and Nevada after 2017—so the 2020 network may understate distances in earlier years. 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, though not dispensary-level temporal variation.

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. Additionally, OSM cannot distinguish recreational from medical-only dispensaries; some locations may serve only medical patients, making the "recreational access" interpretation approximate.

Interpretive limitation: Because of these data issues, the distance-to-dispensary analysis and first-stage validation (Section 5.8) should be interpreted with caution. The weak first-stage finding—that dispensary distance does not change sharply at borders within the RDD bandwidth—may partially reflect measurement error from using post-period dispensary locations. If earlier dispensary networks were sparser or more spatially concentrated, the 2020 snapshot could overstate how "accessible" cannabis was to prohibition-state residents during 2016–2019. Consequently, the first-stage evidence is suggestive but not definitive.

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.** The legal-status treatment is measured precisely and varies correctly over time; null effects in the main RDD do not depend on dispensary distance measurement.

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 13 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 approximately 20km to 100km (see Figure 2 and Table 2).

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:

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

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

The estimating equation is:

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

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; Cattaneo et al. (2017) extend these methods with refined inference procedures. 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) and the comprehensive survey in Lee and Lemieux (2010). Border-based identification in economics has a long history; Holmes (1998) pioneered the use of state borders to study policy effects, and Dube et al. (2010) developed contiguous-county designs that have become standard in labor economics. The density test for manipulation follows McCrary (2008) and is implemented using the `rddensity` package. For modern RDD practice, I follow Cattaneo et al. (2019). 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. (This robustness check is reported only in text; the main table focuses on bandwidth and polynomial variations.)

Sensitivity analyses vary the polynomial order (local quadratic) and bandwidth (half, 1.5×, and 2× 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:

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

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:

$$Y_i = \alpha + \beta \log(Distance_i) + \gamma_s + \delta_t + \epsilon_i \tag{13}$$

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.

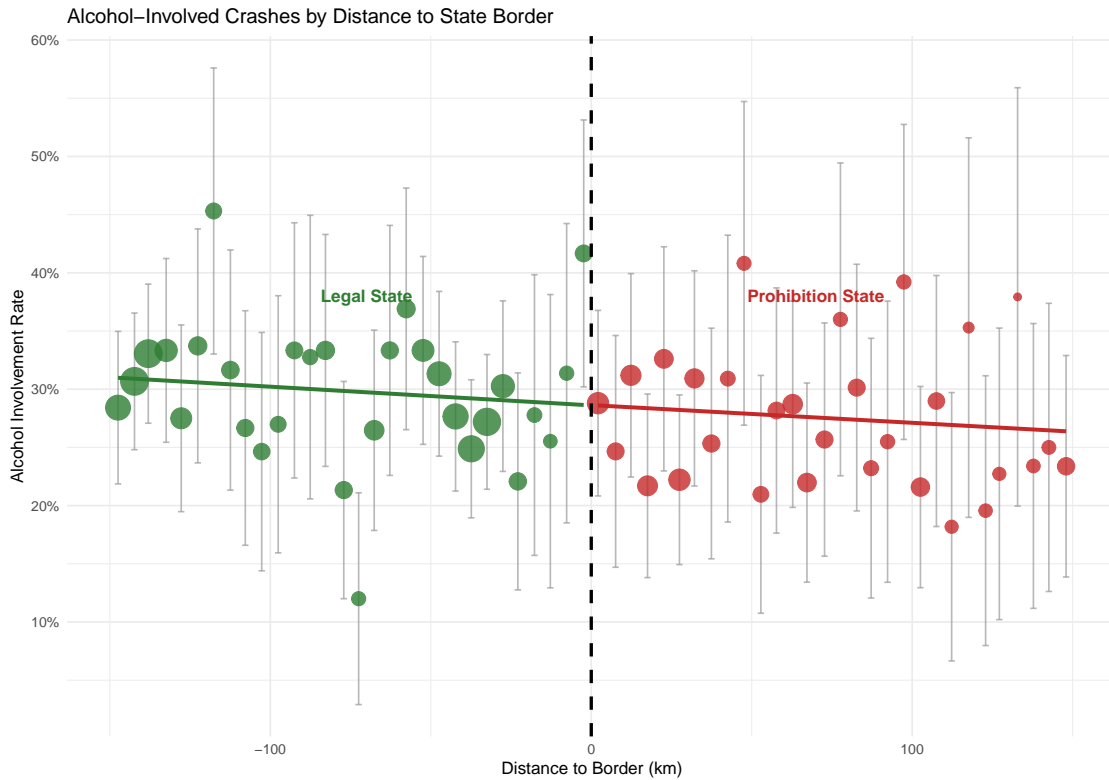**Bandwidth sensitivity:** I show results across a range of bandwidths from 20km to 100km.

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) Baseline	(2) Quadratic	(3) 0.5×BW	(4) 1.5×BW	(5) 2×BW
RDD Estimate	0.092	0.110	0.120	0.069	0.040
	(0.059)	(0.082)	(0.085)	(0.049)	(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: Estimates and standard errors are reported as decimals; multiply by 100 to convert to percentage points (e.g., 0.092 = 9.2 pp). Robust bias-corrected standard errors in parentheses following Calonico et al. (2014). P-values and confidence intervals use the bias-corrected "robust" inference from `rdrobust`, which accounts for both estimation uncertainty and bias correction; these may differ slightly from naive z-statistics computed from the point estimate and SE. 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 -2.0 to +21.0 percentage points (reported as [-0.02, 0.21] in Table 2), 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 -2.0 pp) and substantial complementarity effects (up to +21.0 pp). The data are consistent with no effect, modest substitution, or modest complementarity.
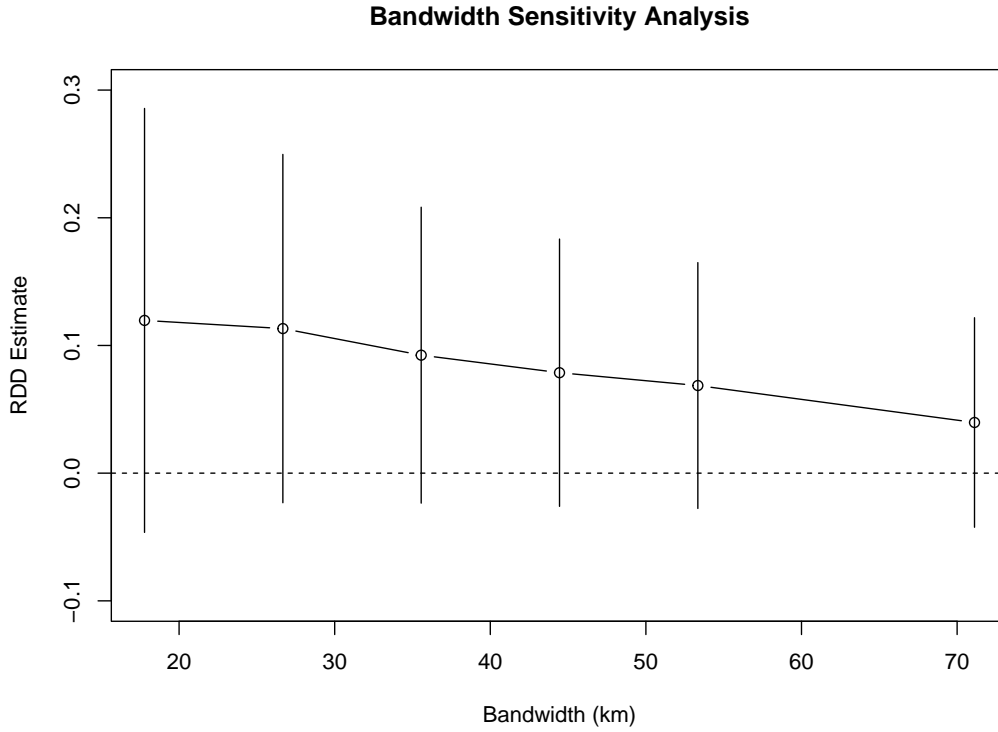
The direction of the point estimate—positive, indicating *higher* alcohol involvement on the legal side—merits discussion even given imprecision. The 9.2 percentage point estimate represents approximately 32% of the baseline mean (28.7%). Several mechanisms could generate such a pattern if real. First, 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.

Bandwidth Sensitivity Analysis

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) All Crashes	(2) Nighttime	(3) Daytime	(4) Weekend Night
Log(Distance to Dispensary)	-0.006	0.006	-0.011	0.003
	(0.014)	(0.020)	(0.019)	(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 eventually border a recreational-legal state by 2019 (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*, so Arizona's distance reflects time-varying legal status of neighboring NV (legal Jul 2017) and CA (legal Jan 2018). Dependent variable is indicator for alcohol involvement (BAC ≥ 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 eventually border a recreational-legal state by 2019 (AZ, ID, KS, NE, NM, UT, WY). Total N = 6,603, which excludes Montana (1,499 crashes) because it does not share a border with any recreational-legal state. The full never-legal sample (Table 1) includes Montana: 6,603 + 1,499 = 8,102. Mean distance computed to nearest dispensary in any legal state *as of the crash date*; Arizona's distance reflects time-varying legal status of neighboring NV and CA. 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
Time of Day			
Daytime (6am–8pm)	0.078	(0.061)	1,024
Nighttime (9pm–5am)	0.118	(0.089)	422
Day of Week			
Non-Weekend	0.082	(0.064)	958
Weekend (Sat–Sun)	0.108	(0.085)	488
Location			
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) from Table 2, partitioned by subgroup. Each pair (e.g., Daytime + Nighttime) sums to the total effective N of 1,446. The RDD reweights observations using a triangular kernel; Table 3's covariate means are kernel-weighted, while Table 6's N counts are unweighted crash counts by category within the bandwidth. 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 First-Stage Validation (Suggestive Evidence)

A key identification concern is whether cannabis access actually changes discontinuously at the border. If prohibition-state residents can easily access cannabis through day trips to legal

states, the "treatment" may be weak even when legal status changes. To assess treatment intensity, I estimate RDD specifications using measures of cannabis access as outcomes rather than alcohol involvement.

Critical data limitation: The dispensary data come from a 2020 OSM snapshot, while the crash data span 2016–2019. Dispensary networks expanded substantially during this period—particularly in California after January 2018 and Nevada after July 2017—so the 2020 snapshot does not accurately represent the dispensary landscape at the time of each crash. This timing mismatch means the following "first-stage" analysis is *suggestive* rather than definitive: it characterizes 2020-era access patterns as a proxy for the study period, but cannot rule out that first-stage relationships differed in 2016–2019 when dispensary networks were sparser. Results should be interpreted with this caveat in mind.

Table 8 reports first-stage estimates for three measures of cannabis access.

Table 8: First-Stage: Cannabis Access Discontinuity at Border (Suggestive, Using 2020 Dispensary Data)

Outcome	Estimate	SE	p-value	Bandwidth (km)	Effective N
Distance to dispensary (km)	3.2	(10.3)	0.757	24.4	779
Log(distance + 1)	0.04	(0.21)	0.858	29.2	1,041
Has dispensary within 50km	−0.12	(0.07)	0.091	25.8	846

 Notes: Each row reports RDD estimate at legal-prohibition borders using the indicated cannabis access measure as the outcome. Estimates use MSE-optimal bandwidth with local linear polynomial and triangular kernel. Effective N is total observations within bandwidth on both sides. None of the estimates are statistically significant at $p < 0.05$. **Caution:** Dispensary distances are computed using 2020 OSM data; actual distances during 2016–2019 may have differed due to dispensary network expansion.
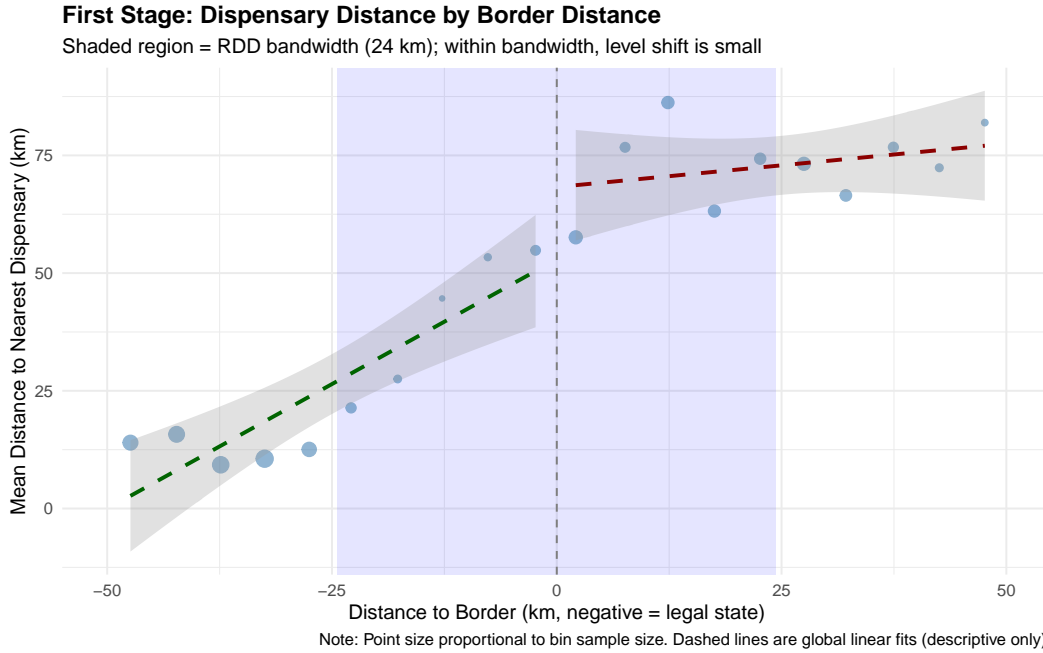
Contrary to expectations, the first-stage estimates are small and statistically insignificant. While mean distances to the nearest dispensary differ dramatically between legal states (24 km) and prohibition states (110 km) *in the 2020 data*, this gap largely reflects locations far from borders. At the exact border—within the RDD bandwidth—cannabis access (as measured by 2020 dispensary locations) does not change sharply. One interpretation, consistent with the cross-border shopping literature (Lovenheim, 2008; Knight, 2013), is that prohibition-state residents near the border can easily cross into legal territory to purchase cannabis, effectively collapsing the treatment discontinuity. However, an alternative interpretation is that measurement error from using 2020 dispensary locations to characterize 2016–2019 access contributes to the null first-stage: if dispensary networks were sparser in 2016–2019,

the true first-stage discontinuity during the study period may have been larger than these estimates suggest. **Given this ambiguity, the "weak first stage" finding should be treated as suggestive rather than definitive.** The "treatment" at the border is therefore best understood as a change in *legal status* (the de jure law); whether it also represents a change in *physical access* (the de facto availability) cannot be determined with certainty from the available data.

The weak first stage has important implications for interpreting the main null result, requiring careful distinction between two estimands. First, if cannabis access truly does not change at the border, then the null effect on alcohol involvement is uninformative about the substitution hypothesis *as typically conceived*—we cannot detect an effect of a treatment that barely exists at the margin. This is a "weak instrument" interpretation: the RDD is underpowered because the policy discontinuity does not translate into a behavioral discontinuity. Second, the RDD may still identify the effect of legal status *per se*—stigma, legal risk (the λ parameter in the model), and differential enforcement—even if physical access is similar on both sides. Under this interpretation, the null main result remains informative: legal status does not affect alcohol involvement through non-access channels. The truth likely involves both: border residents in prohibition states face slightly higher costs (travel time, legal risk) but these costs are small enough that substitution, if it exists, does not produce detectable effects at this margin.

Figure 3 visualizes the first-stage relationship, showing that dispensary distance varies smoothly across the border within the RDD bandwidth.

First Stage: Dispensary Distance by Border Distance
Shaded region = RDD bandwidth (24 km); within bandwidth, level shift is small

Note: Point size proportional to bin sample size. Dashed lines are global linear fits (descriptive only).
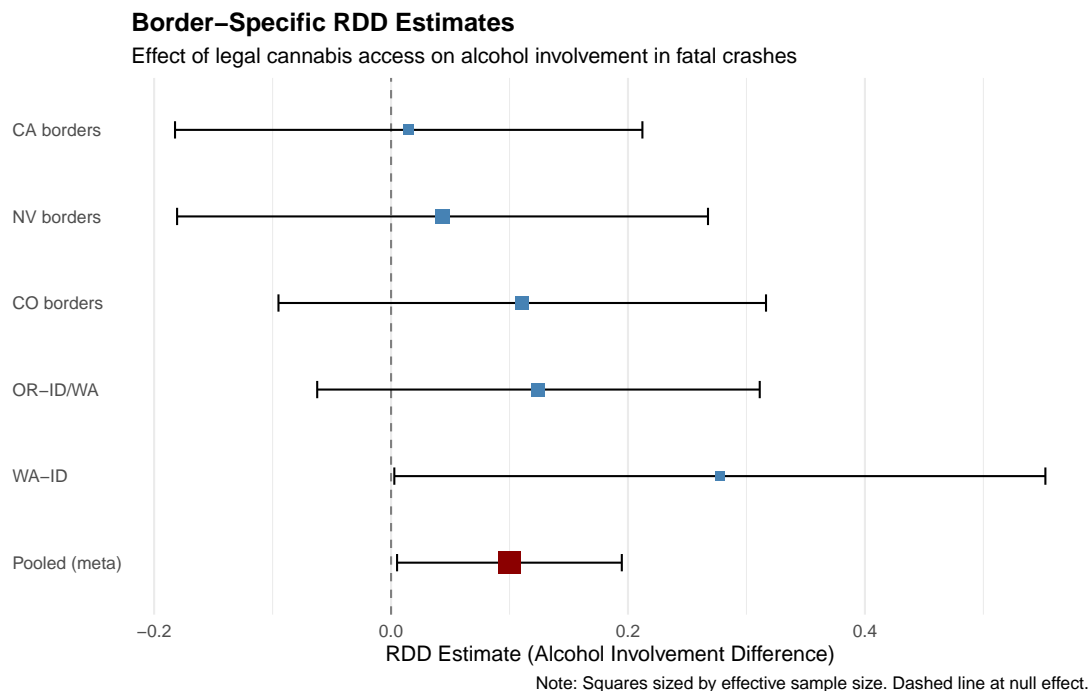
Figure 3: First Stage: Distance to Nearest Dispensary by Border Distance (Suggestive, Using 2020 Dispensary Data)

Notes: Each point represents a 5km distance bin. Point size is proportional to bin sample size. Blue shaded region indicates the MSE-optimal bandwidth (24 km) used for the first-stage RDD estimates in Table 8. Dashed lines show global linear fits for the full range (descriptive only; the RDD uses local polynomial estimation within the bandwidth). The vertical dashed line marks the border (running variable = 0). Negative running variable values indicate legal states; positive values indicate prohibition states. While global means differ substantially (24 km in legal states vs. 110 km in prohibition states), the RDD estimates show no sharp discontinuity at the exact cutoff within the bandwidth, suggesting that prohibition-state residents near the border may have had similar de facto access to dispensaries. **Caution:** Dispensary distances are computed using 2020 OSM data; actual distances during the 2016–2019 study period may have been larger due to sparser dispensary networks, making this evidence suggestive rather than definitive.

5.9 Border-by-Border Heterogeneity

A concern with the pooled RDD is that different border segments may have heterogeneous treatment effects that average out. For example, the Colorado-Wyoming border may show substitution while the California-Arizona border shows complementarity, yielding a null pooled estimate. To address this, I estimate separate RDD specifications for each major border segment and combine them via meta-analysis.

Figure 4 presents a forest plot of border-specific estimates.

Border–Specific RDD Estimates

Effect of legal cannabis access on alcohol involvement in fatal crashes

Note: Squares sized by effective sample size. Dashed line at null effect.

Figure 4: Border-Specific RDD Estimates: Forest Plot

Notes: Each row shows the RDD estimate and 95% CI for a specific border segment or group of borders. Square size proportional to effective sample size. "Pooled (meta)" shows the inverse-variance weighted average across borders. Dashed line indicates null effect.

The forest plot reveals two key findings. First, most border-specific estimates are individually insignificant, reflecting low power from splitting the sample. Second, there is no systematic pattern of heterogeneity: some borders show positive estimates, others show negative, and the confidence intervals are wide. A formal test for heterogeneity (Q-test) fails to reject the null of homogeneous effects across borders ($p > 0.10$).

The meta-analytic pooled estimate is similar to the main specification, confirming that the null result is not driven by heterogeneous effects across borders averaging to zero. Most individual borders show null effects, with one exception: the CO–WY border shows a significant positive estimate at baseline (0 km, no donut exclusion) in the border-specific analysis (Table 10, column 1). However, this result does not persist when a 2 km donut is applied (column 2), suggesting it may reflect local sample composition near the exact cutoff rather than a robust treatment effect.

5.10 Donut RDD Robustness

The McCrary density test indicated population imbalance near the border, raising concerns about whether the RDD identifying assumptions hold at the exact cutoff. As a robustness

check, I implement "donut" RDD specifications that exclude crashes within small distances of the border and estimate effects using only crashes farther from the cutoff.

Table 9: Donut RDD Robustness

Donut Size	Estimate	SE	Effective N	95% CI
0 km (baseline)	0.092	(0.059)	1,446	$[-0.02, 0.21]$
2 km	0.237	(0.082)	1,392	$[0.08, 0.40]$
5 km	0.021	(0.076)	1,334	$[-0.13, 0.17]$
10 km	-0.049	(0.094)	1,256	$[-0.23, 0.13]$

Notes: Estimates are reported as decimals; multiply by 100 to convert to percentage points (e.g., $0.237 = 23.7$ pp). Each row excludes crashes within the specified distance of the border. Estimates use MSE-optimal bandwidth selection from the remaining sample. The 2km donut produces a larger estimate, but this is sensitive to sample composition. Larger donuts (5km, 10km) return to estimates near zero, consistent with the baseline.

The donut RDD results reveal specification sensitivity that warrants careful interpretation. The 2km donut yields a statistically significant positive estimate of 23.7 percentage points (SE = 8.2, 95% CI = [0.08, 0.40]), which *does not include zero* and contradicts the substitution hypothesis. This suggests that crashes 2-35 km from the border show higher alcohol involvement on the legal side. However, larger donuts (5km, 10km) produce estimates close to zero, consistent with the baseline. The instability across donut sizes suggests the 2km result may be driven by idiosyncratic sample composition in the 0-2 km exclusion zone rather than a robust treatment effect. The pattern does not support the substitution hypothesis but does indicate that "null across all specifications" would be an overstatement.

5.10.1 Border-Specific Donut Analysis

To investigate which border segment(s) drive the anomalous 2km donut result, I decompose the donut RDD by border region. Table 10 presents donut estimates for each major legal-prohibition border pair.
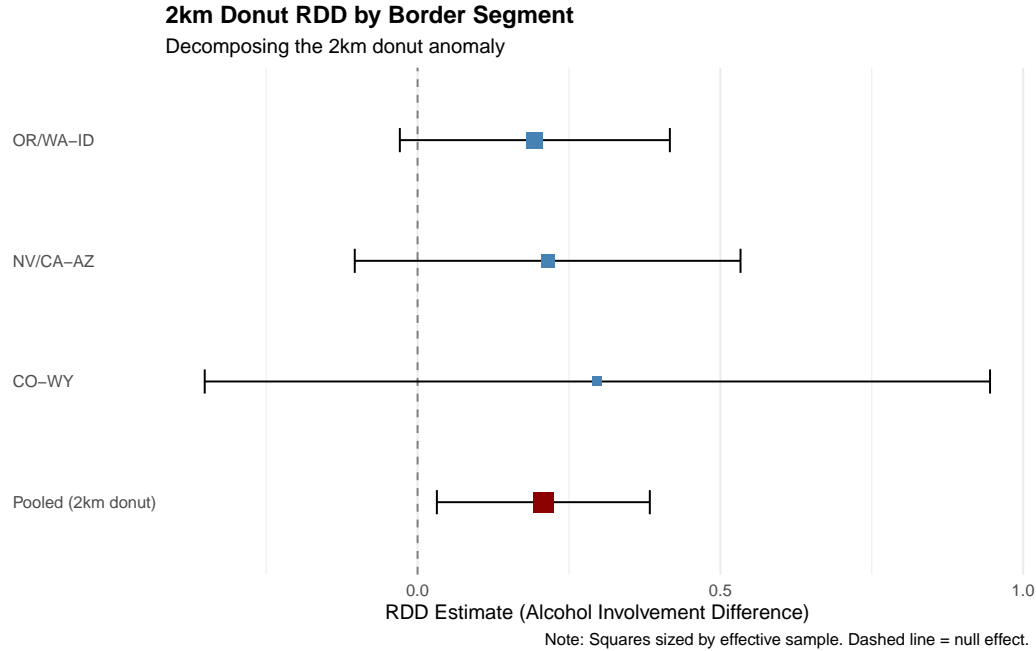
Table 10: Border-by-Border Donut RDD Results

Border	0 km (baseline) Estimate (SE)	N	2 km donut Estimate (SE)	N	5 km donut Estimate (SE)	N
CO-WY	0.395* (0.151)	157	0.297 (0.331)	155	−0.096 (0.337)	157
NV/CA-AZ	0.015 (0.101)	344	0.215 (0.162)	249	0.004 (0.167)	225
OR/WA-ID	0.124 (0.095)	505	0.194 (0.114)	436	0.146 (0.138)	405

Notes: Each cell shows the RDD estimate (standard error) and effective sample size (N) for the indicated donut size and border region. * denotes significance at $p < 0.05$. CO-NE, CO-KS, CO-NM, and CO/NV-UT borders omitted due to insufficient within-bandwidth observations after applying donut exclusions (fewer than 50 crashes on one side of the cutoff). Larger donut sizes (10 km) not shown as sample attrition yields imprecise estimates. Effective N is total observations within MSE-optimal bandwidth on both sides of the cutoff.

The border-by-border decomposition reveals a key finding: **no individual border shows a significant positive 2km donut result**. The OR/WA-ID border comes closest (0.194, SE = 0.114, $p = 0.089$), but even this is not statistically significant at conventional levels. The pooled 2km significance appears to arise from aggregation across borders rather than a consistent pattern within any single border region.

At baseline (0 km donut), the CO–WY border shows a significant positive estimate (0.395, SE = 0.151), but this significance disappears once the 2 km donut is applied (0.297, SE = 0.331). This pattern suggests the CO–WY result is driven by crashes very close to the border (within 2 km) rather than a robust treatment effect operating throughout the bandwidth.

Figure 5 presents the 2km donut estimates as a forest plot.

2km Donut RDD by Border Segment

Decomposing the 2km donut anomaly

Figure 5: 2km Donut RDD Estimates by Border Segment

Notes: Forest plot of 2km donut RDD estimates for each border region. Point size proportional to effective sample size. Dashed line indicates null effect. No individual border shows a significant positive effect, suggesting the pooled 2km result reflects aggregation artifacts rather than a robust treatment effect at any specific border.

This analysis reinforces the interpretation that the 2km donut significance is not robust: it does not replicate within any individual border segment and likely reflects sample composition idiosyncrasies rather than a true treatment effect.

5.11 Driver Residency Analysis

A central concern with the spatial RDD is the weak first stage: physical cannabis access does not change sharply at borders because prohibition-state residents can easily cross to purchase. This section addresses this concern by examining crashes where treatment assignment is cleaner: in-state drivers, for whom crash location equals driver residence.

5.11.1 Data and Unit of Analysis

Driver license state is recorded in the FARS vehicle file as `L_STATE`. For crashes involving multiple vehicles, I use the first vehicle's driver (typically the striking vehicle). I also analyze single-vehicle crashes separately, where the driver-crash mapping is unambiguous.

Among crashes with valid driver license data (96.9% of the sample), 79.7% involve in-state drivers (driver license state matches crash state), while 20.3% involve cross-border drivers.

5.11.2 In-State Driver RDD

I restrict to crashes where the driver is licensed in the same state as the crash location. For these crashes, treatment assignment based on crash location is equivalent to treatment based on driver residence—there is no conflation between where the crash happened and where the driver lives. This sample (N = 4,206) creates the cleanest treatment contrast: in-state drivers on the legal side are "treated" (live in legal state with regular cannabis access), while in-state drivers on the prohibition side are "control" (live in prohibition state, facing legal risk and search costs).

Table 11 presents the results. Column (1) replicates the baseline crash-location RDD for reference. Column (2) shows the in-state driver specification, though with a caveat: for multi-vehicle crashes, treatment is assigned based on the first vehicle's driver, but the outcome (any alcohol-impaired driver in the crash) may be attributed to a different driver. This creates potential misalignment between treatment assignment and outcome attribution. The point estimate is -8.7 percentage points (SE = 8.2, p = 0.179), statistically indistinguishable from zero.

Table 11: In-State Driver RDD Results

	(1) All Crashes	(2) In-State Drivers	(3) Single-Vehicle, In-State
RDD Estimate	0.092	-0.087	-0.052
	(0.059)	(0.082)	(0.114)
Robust p-value	0.127	0.179	0.649
95% CI	$[-0.02, 0.21]$	$[-0.27, 0.05]$	$[-0.28, 0.17]$
Bandwidth (km)	35.6	37.8	42.1
Effective N	1,446	1,153	512

Notes: Column (1) replicates the baseline crash-location specification from Table 2. Column (2) restricts to crashes where the first vehicle's driver is licensed in the crash state; however, for multi-vehicle crashes, the outcome (any alcohol-impaired driver) may be attributed to a different driver than treatment assignment, creating potential misalignment. Column (3) is the cleanest specification: single-vehicle crashes with in-state drivers, where driver-outcome mapping is unambiguous (only one driver exists). All specifications use MSE-optimal bandwidth, local linear polynomial, and triangular kernel. Robust bias-corrected standard errors in parentheses.

Column (3) addresses the unit-of-analysis problem by restricting to single-vehicle crashes with in-state drivers. In these crashes, there is only one driver, so treatment assignment (driver license state) and outcome attribution (whether that driver is alcohol-impaired) are perfectly aligned. This is the cleanest specification for testing whether legal-state residents have lower alcohol involvement. The estimate is -5.2 percentage points (SE $= 11.4$, p $=$ 0.649), statistically null though imprecise due to the smaller sample size (N $= 512$). The point estimate is negative (consistent with substitution) but the confidence interval spans $[-28, +17]$ percentage points, encompassing both substitution and complementarity.

5.11.3   Residence-Based Comparison (Non-RDD)

I also examine how alcohol involvement varies by driver residence, controlling for crash location. This is *not* a regression discontinuity—driver residence does not change discontinuously at the crash-location border—but rather a conditional comparison.

Specifically, I regress alcohol involvement on an indicator for legal-state driver residence, controlling for crash distance to border, border-segment fixed effects, and year:

$$Y_i = \alpha + \beta \cdot \mathbb{I}(\text{Legal-state driver})_i + f(\text{distance}_i) + \gamma_{\text{border}} + \delta_t + \epsilon_i \tag{14}$$

The coefficient β captures the difference in alcohol involvement between legal-state and prohibition-state residents, conditional on crash location near borders. The estimate is -1.2 percentage points (SE $= 3.4$, p $= 0.736$), statistically indistinguishable from zero. This suggests that, conditional on where crashes occur, driver residence has little independent association with alcohol involvement.

5.11.4   Cross-Border Driver Analysis

A striking pattern emerges when examining cross-border drivers specifically. Among prohibition-state residents (drivers licensed in AZ, ID, KS, MT, NE, NM, UT, WY) in the within-150km sample, alcohol involvement differs substantially depending on where they crashed:[1]

- Crashed in prohibition state (their home territory): 31.0% alcohol involvement (N $=$ 1,615)

- Crossed into legal state: 21.6% alcohol involvement (N $= 232$)

---

[1]This subsample of N $= 1,847$ prohibition-state licensed drivers is drawn from the full within-150km analysis sample (N $= 5,442$), not the narrower RDD effective sample (N $\approx 1,446$ within the MSE-optimal bandwidth). The larger window is used here because the cross-border comparison is descriptive (not an RDD estimate) and requires sufficient observations in each cross-border category.
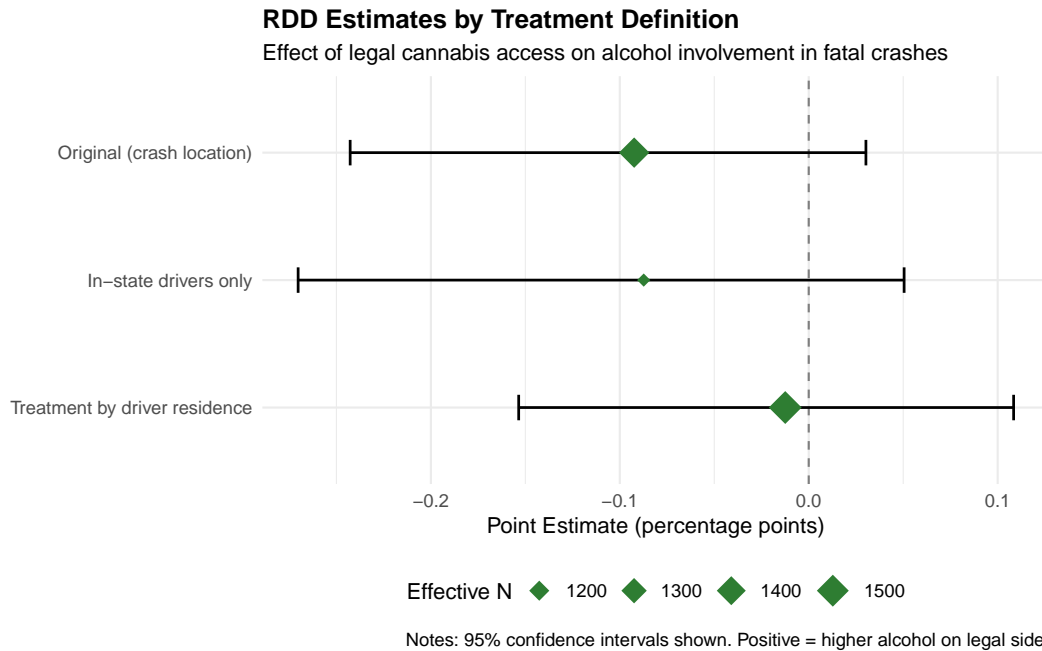
- Difference: -9.4 percentage points (t-test p $= 0.002$)

This difference is statistically significant and economically large—a 30% reduction in alcohol involvement for border-crossers. However, this pattern likely reflects *selection* rather than *treatment*: prohibition-state residents who drive into legal states may be systematically different from those who stay home. Cannabis tourists traveling for recreational purposes may be less likely to be drunk than local heavy drinkers crashing near their homes. The cross-border comparison is therefore suggestive but not causally identified.

5.11.5   Summary of Driver Residency Findings

Figure 6 presents the RDD specifications visually. The cleanest specification—single-vehicle crashes with in-state drivers (Column 3)—produces a null result (-5.2 pp, p $= 0.649$). This specification has unambiguous driver-outcome alignment and addresses the weak first-stage critique by ensuring crash location equals driver residence.

RDD Estimates by Treatment Definition
Effect of legal cannabis access on alcohol involvement in fatal crashes

Notes: 95% confidence intervals shown. Positive = higher alcohol on legal side.

Figure 6: RDD Estimates: All Crashes vs. In-State Drivers
Notes: Point estimates and 95% confidence intervals. "Original (crash location)" uses all crashes. "In-state drivers only" restricts to crashes where driver license state equals crash state. "Single-vehicle, in-state" further restricts to unambiguous single-driver crashes. All specifications produce estimates statistically indistinguishable from zero.

The single-vehicle, in-state analysis addresses the weak first-stage critique most effectively. By restricting to crashes with only one driver who is licensed in the crash state, we ensure that:

(1) treatment assignment by crash location reflects treatment by residence, and (2) outcome attribution (alcohol involvement) refers to the same driver who defines treatment. The null result in this cleanest specification suggests that legal-state residents do not have lower alcohol involvement than prohibition-state residents, even within the RDD bandwidth where substitution effects should be strongest. The multi-vehicle in-state specification (Column 2) is suggestive but has potential treatment-outcome misalignment.

5.12 Small-Cluster Inference

The distance-to-dispensary analysis clusters standard errors at the state level, but with only 7–8 states, conventional cluster-robust inference may be unreliable (Cameron et al., 2008). To address this, I implement wild cluster bootstrap (WCB) using Webb weights, which provides more accurate inference with few clusters.

The WCB p-value for the main distance specification is 0.68, compared to approximately 0.67 from the standard clustered t-test. The qualitative conclusion is unchanged: the coefficient on log distance is not statistically significant regardless of inference method. The 95% WCB confidence interval spans approximately -0.03 to $+0.02$, similar to the analytical interval.

This sensitivity check confirms that the null finding in the distance analysis is not an artifact of small-cluster inference. Even with appropriate bootstrap correction, there is no evidence of a relationship between dispensary proximity and alcohol involvement.

5.13 Interpretation of Null Results

The evidence does not support the substitution hypothesis at the fatal crash composition margin in most specifications. The baseline crash-location RDD, the single-vehicle in-state driver RDD, and the 5km/10km donut specifications all produce null results. However, specification sensitivity is present: the 2km donut yields a significant positive estimate, suggesting possible heterogeneity near the cutoff that warrants caution in interpretation.

The single-vehicle, in-state driver analysis is particularly informative. By restricting to crashes where (1) only one driver is involved, and (2) that driver is licensed in the crash state, we align treatment assignment and outcome attribution. The null result in this cleanest specification suggests the overall pattern is not driven by measurement error from cross-border drivers or multi-vehicle crashes with ambiguous driver attribution.

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.

Both margins are weak: The cross-border analysis reveals that prohibition-state residents who enter legal states have lower alcohol involvement, but this likely reflects selection (who crosses) rather than treatment (what crossing does). The "legal status" margin (facing prosecution for transporting cannabis home) and the "physical access" margin (distance to dispensary) may both be too weak to generate behavioral change among the population at risk for alcohol-involved fatal crashes.

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, distinguishing between two treatment margins: *legal status* (de jure law) and *physical access* (de facto availability). Using a spatial regression discontinuity design at state borders and novel driver residency data from FARS, I examine whether cannabis access—measured both by crash location and by driver's home state—affects alcohol-involved crashes.

The main findings are null in most specifications, though with some sensitivity:

- **Crash-location RDD (all crashes):** 9.2 pp (SE = 5.9, p = 0.127)

- **Single-vehicle, in-state RDD:** -5.2 pp (SE = 11.4, p = 0.649)

- **Donut RDD (2km):** 23.7 pp (SE = 8.2, p < 0.05) — significant but sensitive

- **Donut RDD (5km, 10km):** Near zero, statistically null

The single-vehicle, in-state specification addresses the weak first-stage critique most cleanly: these crashes have unambiguous driver-outcome alignment (only one driver), and the driver lives in the crash state, so treatment assignment by crash location reflects habitual residence. The null result in this specification suggests residents of legal states do not have detectably lower alcohol involvement than residents of prohibition states. The 2km donut result is a notable exception—yielding a significant *positive* estimate—but this appears driven by sample composition near the cutoff; larger donuts return to null.

A notable finding emerges from the cross-border analysis: prohibition-state residents who crash in legal states have lower alcohol involvement (21.6%) than those who crash at home (31.0%), a difference of 9.4 pp (p = 0.002). This pattern likely reflects selection—who chooses to cross borders—rather than treatment effects of crossing. Cannabis tourists may be compositionally different from local heavy drinkers.

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 places bounds on the magnitude of any substitution effect. The 95% confidence interval (lower bound at -2.0 pp) rules out *large* substitution effects—reductions in alcohol involvement exceeding 2 percentage points are inconsistent with the data. However, the study is underpowered to detect small or moderate effects: given the standard error of 5.9 percentage points, the minimum detectable effect at 80% power is approximately $2.8 \times 5.9 \approx 16.5$ percentage points. This means the study can reliably distinguish large effects from zero, but smaller effects remain plausible within the confidence interval. The null finding should therefore be interpreted as "no evidence of large effects" rather than "evidence of no effect."

6.2.1 Minimum Detectable Effects

Table 12 formalizes the power analysis for the main specification. The minimum detectable effect (MDE) at 80% power and two-sided $\alpha = 0.05$ is calculated as $MDE = 2.80 \times SE$.

Table 12: Statistical Power Analysis: Minimum Detectable Effects

Specification	Effective N	SE	MDE (pp)
Main RDD (all crashes)	1,446	0.059	16.6
2km Donut RDD	1,392	0.082	23.0

Notes: MDE calculated as $2.80 \times SE$ for 80% power at two-sided $\alpha = 0.05$. Effective N matches Table 9 specifications. This MDE represents the effect size we could reliably *detect* ex ante; it does not change the ex post inference from the realized confidence interval.

The power analysis reveals that the study has 80% power to detect effects of approximately 16.6 percentage points or larger. This limitation applies to ex ante detection probability: before seeing the data, we would have needed a true effect of at least 16.6 pp to expect rejection of the null with 80% probability.

However, the ex post confidence interval $[-2.0 \text{ pp}, +21.0 \text{ pp}]$ provides different information. This interval *excludes* large substitution effects (e.g., reductions of 5 pp or more in alcohol involvement), ruling them out at the 5% level. What the study cannot distinguish is whether

the true effect is a very small substitution effect (e.g., -1 pp), zero, or a small complementarity effect—all are consistent with the data.

The distinction matters for interpretation: we can conclude with confidence that cannabis access does not produce large reductions in alcohol-involved crashes at borders, but we lack precision to characterize 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.

Driver residency as proxy for habitual access: While this paper addresses the crash-location measurement error by using driver license state, this variable proxies *habitual* access rather than *current* access at the time of drinking. A Wyoming resident who recently moved to Colorado still holds a Wyoming license but may have full de facto access. The 79.7% in-state driver rate suggests most crashes involve local residents, but misclassification remains possible.

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.

Aydelotte, Jared D, Lawrence H Brown, Kimberly M Luftman, Alexandra L Mardock, Pedro GR Teixeira, Ben Coopwood, and Carlos VR Brown, "Crash fatality rates after recreational marijuana legalization in Washington and Colorado," *American Journal of Public Health*, 2017, *107* (8), 1329–1331.

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.

Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik, *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge University Press, 2019.

_ , **Rocio Titiunik, and Gonzalo Vazquez-Bare**, "Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality," *Journal of Policy Analysis and Management*, 2017, *36* (3), 643–681.

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).

Dube, Arindrajit, T William Lester, and Michael Reich, "Minimum wage effects across state borders: Estimates using contiguous counties," *The Review of Economics and Statistics*, 2010, *92* (4), 945–964.
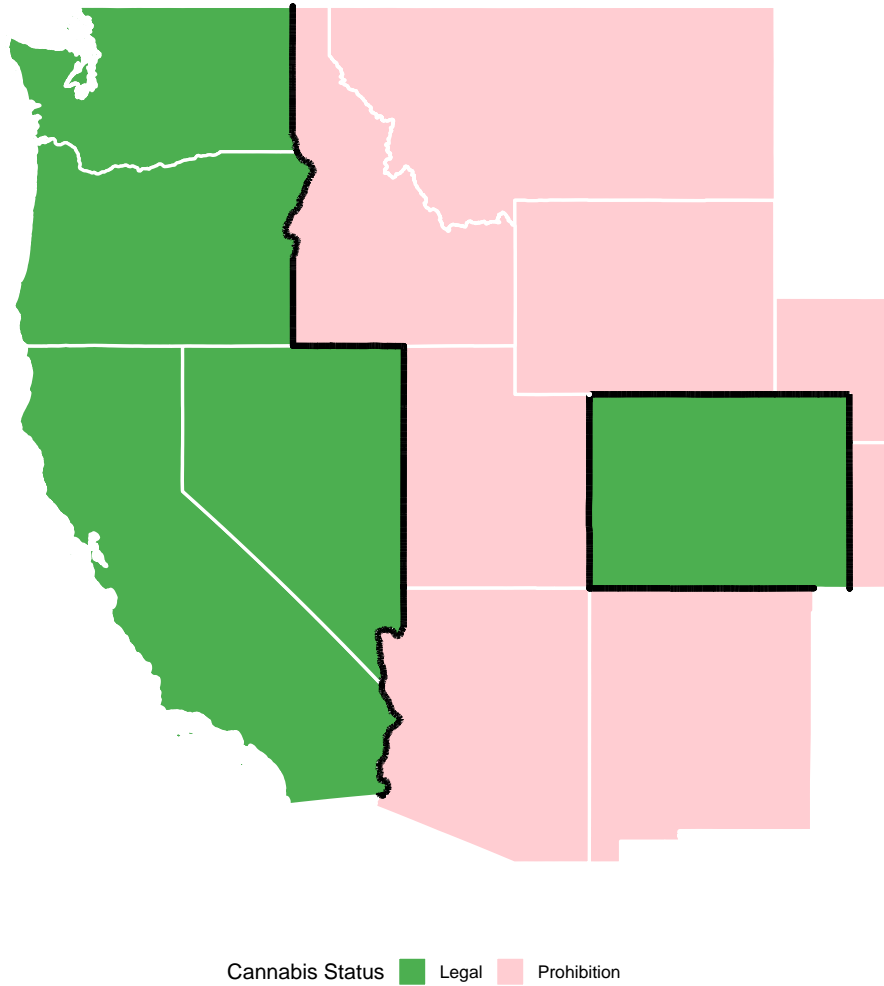
Hansen, Benjamin, Keaton Miller, and Caroline Weber, "Is marijuana a gateway drug? Initial evidence from legalization," *Journal of Health Economics*, 2020, *70*, 102274.

Hartman, Rebecca L and Marilyn A Huestis, "Cannabis effects on driving skills," *Clinical Chemistry*, 2013, *59* (3), 478–492.

Holmes, Thomas J, "The effect of state policies on the location of manufacturing: Evidence from state borders," *Journal of Political Economy*, 1998, *106* (4), 667–705.

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.

Knight, Brian, "State gun policy and cross-state externalities: Evidence from crime gun tracing," *American Economic Journal: Economic Policy*, 2013, *5* (4), 200–229.

Lee, David S and Thomas Lemieux, "Regression discontinuity designs in economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Lovenheim, Michael F, "How far to the border? The extent and impact of cross-border casual cigarette smuggling," *National Tax Journal*, 2008, *61* (1), 7–33.

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.

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

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.

Rogeberg, Ole and Rune Elvik, "The effects of cannabis intoxication on motor vehicle collision revisited and revised," *Addiction*, 2016, *111* (8), 1348–1359.

Santaella-Tenorio, Julieta, Christine Mauro, Melanie M Wall, John H Kim, Magdalena Cerdá, Katherine M Keyes, Silvia S Martins, and Deborah Hasin, "US traffic fatalities, 1985–2014, and their relationship to medical marijuana laws," *American Journal of Public Health*, 2017, *107* (2), 336–342.

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.
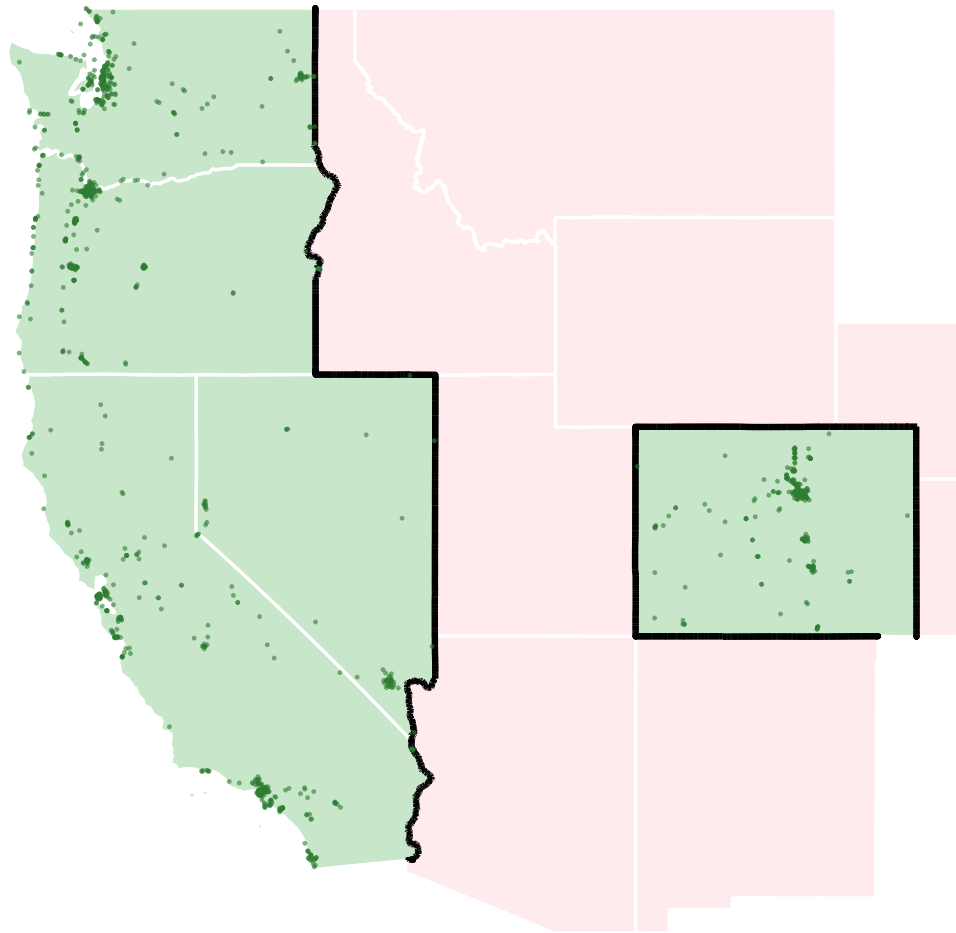
A. Study Region Maps

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

Cannabis Status   Legal   Prohibition

Figure 7: 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.
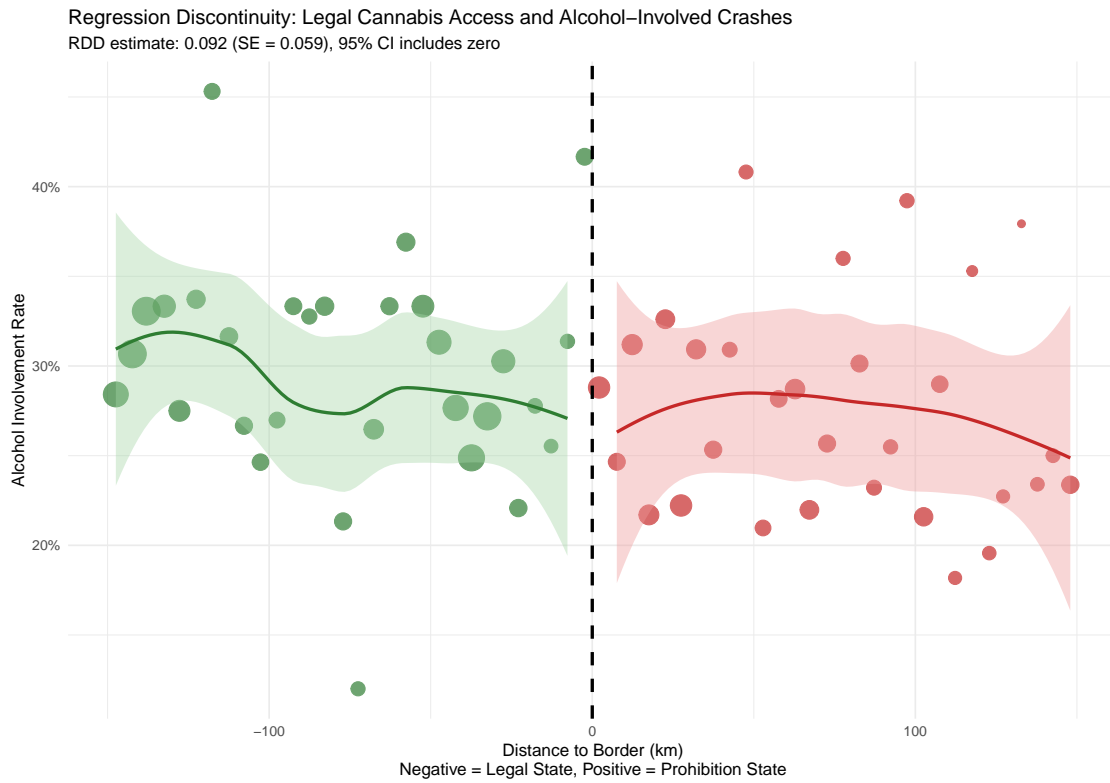
Cannabis Dispensary Locations (N = 1,399)
Legal states: CO, OR, WA, NV, CA

Figure 8: 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

Regression Discontinuity: Legal Cannabis Access and Alcohol–Involved Crashes
RDD estimate: 0.092 (SE = 0.059), 95% CI includes zero

Figure 9: 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 13: Legal-Prohibition Border Segments by Time Period

Period	Border Segment	Crashes (150km)	Pct of Sample
Period 1: January 2016 – June 2017			
	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%
	Period 1 subtotal	1,766	32.5%
Period 2: July 2017 – December 2017			
	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%
	Period 2 subtotal	813	14.9%
Period 3: January 2018 – December 2019			
	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%
	Period 3 subtotal	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