

Click It or Ticket at the Border: A Spatial Regression Discontinuity Analysis of Primary Seatbelt Enforcement Laws

APEP Autonomous Research*

@anonymous

January 29, 2026

Abstract

This paper investigates whether spatial regression discontinuity designs (RDD) at U.S. state borders can credibly identify the effects of seatbelt enforcement policy—and documents why they cannot. We examine geographic discontinuities where enforcement type changes from primary (police can stop drivers solely for non-use) to secondary (citation only if stopped for another violation). Using geocoded fatal crash data from FARS (2001–2019), we compare 289,916 crashes within 100 kilometers of enforcement borders. Our analysis yields a null point estimate (0.7 pp, 95% CI: -0.14 to 1.47 pp), but diagnostic tests reveal fundamental violations of RDD assumptions: McCrary density tests reject continuity ($p < 0.001$), placebo cutoff tests find significant “effects” away from true borders, and covariate balance fails for key variables. Crucially, our running variable (distance to nearest opposite-type state polygon) does not consistently correspond to actual treatment-changing border segments—a design flaw we document as a methodological warning for future spatial RDD applications. This paper serves as a cautionary case study: pooled multi-border spatial RDDs require careful construction of border-segment-specific running variables and conditioning, which we did not implement.

JEL Codes: R41, I18, K32

Keywords: seatbelt laws, traffic safety, spatial regression discontinuity, enforcement, traffic fatalities

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

1. Introduction

Motor vehicle crashes remain a leading cause of preventable death in the United States, claiming over 38,000 lives annually ([National Highway Traffic Safety Administration, 2020](#)). Seatbelts are among the most effective interventions for reducing crash fatalities, with properly worn belts reducing the risk of fatal injury by approximately 45% for front-seat occupants ([Centers for Disease Control and Prevention, 2020](#)). Despite this, seatbelt use rates vary substantially across states, ranging from 70% to over 90%, with enforcement stringency believed to be a key determinant ([Insurance Institute for Highway Safety, 2020](#)).

States employ two primary approaches to seatbelt enforcement. Under *primary enforcement* laws, police officers may stop and cite drivers solely for not wearing a seatbelt. Under *secondary enforcement*, officers may only cite seatbelt violations if the driver has been stopped for another traffic offense. The distinction creates sharp policy variation at state borders—a driver crossing from Virginia (secondary) to North Carolina (primary) immediately faces different enforcement regimes, despite traveling on the same highway through similar terrain.

This paper examines whether these geographic discontinuities are associated with differences in traffic fatality severity using a spatial regression discontinuity approach. We compare fatal crashes occurring near state borders where enforcement type changes, using distance to the border as the running variable. Our outcome is crash severity measured as fatality probability (deaths per person involved) *conditional on a fatal crash occurring*—since FARS only records crashes with at least one fatality, we measure severity within the fatal crash sample rather than overall fatality risk. If primary enforcement increases seatbelt compliance through deterrence, we might expect lower severity on the primary side of the border, since belted occupants are more likely to survive when a crash is severe enough to produce fatalities.

Our main finding is a precisely estimated null. At the MSE-optimal bandwidth of 21.5 kilometers, primary enforcement is associated with a 0.67 percentage point *higher* fatality probability per crash compared to secondary enforcement states ($SE = 0.41$ pp, $p = 0.10$). The 95% confidence interval of $[-0.14, 1.47]$ pp allows us to rule out large differences in either direction. However, key validity tests fail: the McCrary density test strongly rejects continuity ($p < 0.001$) and placebo cutoffs show significant “effects” at locations away from the true border ($p < 0.001$ at +10 km). These failures indicate the RDD identifying assumptions do not hold, and our estimates should be interpreted as descriptive discontinuities rather than causal effects.

We investigate several mechanisms and robustness checks. First, our placebo test examining pedestrian and cyclist deaths—who would not benefit from vehicle seatbelts—shows no

effect ($\tau = -0.002$, $p = 0.75$), consistent with our design not capturing spurious cross-border differences. Second, heterogeneity by time of day reveals marginally larger point estimates for night crashes (+1.3 pp) versus day crashes (+0.03 pp), though neither is statistically significant. Third, results are robust to bandwidth choice, polynomial order, and kernel specification.

However, several validity tests raise concerns about the spatial RDD assumptions in this context. The McCrary density test strongly rejects the null of continuous crash density at the border ($p < 0.001$), suggesting that crashes are systematically distributed differently on either side of state lines. This could reflect differences in road infrastructure, population density, or traffic patterns that correlate with both enforcement regime and fatality risk. Additionally, placebo cutoff tests show significant “effects” at artificial borders away from the true boundary ($p < 0.001$ at +10 km), indicating that systematic variation in crash characteristics extends well beyond the immediate border region.

Our contribution is threefold. First, we provide the first spatial RDD analysis of seatbelt enforcement at state borders, documenting that this design faces fundamental identification challenges. Prior work has relied on difference-in-differences designs exploiting temporal variation in law changes (Cohen and Einav, 2003; Carpenter and Dobkin, 2008). While geographic RDD offers a potentially cleaner identification strategy by comparing similar locations, our diagnostic tests reveal that pooling across heterogeneous border segments violates the continuity assumptions required for causal interpretation. Second, we document that crashes near state borders differ systematically in ways correlated with enforcement regime, suggesting that enforcement type may be endogenous to broader state-level policy environments. Third, we provide guidance for future spatial RDD research: pooled multi-border designs require border-segment conditioning or explicit aggregation strategies, and crash-level FARS data raise selection concerns when the policy may affect whether crashes become fatal.

This paper therefore serves dual purposes. First, it provides the most comprehensive attempt to date at using spatial RDD to evaluate seatbelt enforcement, documenting both the potential and the pitfalls of this approach. Second, it offers methodological guidance for researchers considering geographic identification strategies, demonstrating the importance of diagnostic testing and the risks of pooling across heterogeneous boundaries without appropriate conditioning.

The remainder of this paper proceeds as follows. Section 2 describes the institutional background of seatbelt enforcement laws, including the economic theory of enforcement deterrence and the pattern of state adoption. Section 3 reviews the related literatures on traffic safety, RDD methodology, and spatial identification. Section 4 details our data

sources and sample construction. Section 5 presents our empirical strategy, including formal identification assumptions and the diagnostic tests we employ. Section 6 reports results for the main specification, robustness checks, heterogeneity, and validity tests. Section 7 discusses implications, limitations, and comparison with prior literature. Section 8 concludes with recommendations for future research.

2. Institutional Background

2.1 Seatbelt Laws in the United States

The United States lacks a federal seatbelt mandate for adults; instead, requirements are determined at the state level. New York became the first state to require seatbelt use in 1984, and by 1995 all states except New Hampshire had enacted some form of seatbelt law ([National Highway Traffic Safety Administration, 2001](#)). These laws vary along several dimensions, including covered seats, ages, vehicle types, and crucially, enforcement type.

2.2 Primary vs. Secondary Enforcement

Under **primary enforcement**, law enforcement officers may stop vehicles and issue citations solely for observed seatbelt non-use. This creates direct incentives for compliance—drivers know they face the possibility of traffic stops and fines (\$25–\$200 depending on state) even if driving lawfully in all other respects.

Under **secondary enforcement**, officers may only cite seatbelt violations if the vehicle has been stopped for an independent traffic violation such as speeding or running a red light. The deterrence effect is weaker because the probability of citation is conditional on committing another offense.

The economic distinction between primary and secondary enforcement operates through expected penalties. Under primary enforcement, the expected cost of non-compliance is approximately:

$$E[\text{Cost}]_{\text{primary}} = p_{\text{patrol}} \times p_{\text{observe}} \times p_{\text{stop}} \times \text{Fine} \quad (1)$$

where p_{patrol} is the probability of encountering a patrol officer, p_{observe} is the probability the officer observes non-use, and p_{stop} is the probability of being stopped conditional on observation. Under secondary enforcement:

$$E[\text{Cost}]_{\text{secondary}} = p_{\text{other violation}} \times p_{\text{stop}} \times p_{\text{observe}} \times \text{Fine} \quad (2)$$

where $p_{\text{other violation}}$ is the probability of committing another traffic offense. Since most drivers

do not commit traffic violations on any given trip, $p_{\text{other violation}} \ll 1$, making the expected cost under secondary enforcement substantially lower.

Beyond formal deterrence, primary enforcement may operate through salience and social norms. “Click It or Ticket” campaigns, which often accompany primary enforcement laws, raise awareness of seatbelt requirements and may shift perceptions of enforcement probability even beyond actual police activity. Social pressure within vehicles may also increase as passengers become more aware of legal requirements.

As of the end of our analysis period (2019), 34 states and the District of Columbia had primary enforcement laws, while 15 states maintained secondary enforcement. New Hampshire has no adult seatbelt law and serves as a natural control in some designs. The adoption of primary enforcement occurred gradually, with California being the first to upgrade from secondary to primary in 1993.

2.3 Staggered Adoption and Border Variation

The staggered adoption pattern creates natural variation at state borders. For example, North Carolina adopted primary enforcement in December 2006, while neighboring Virginia maintains secondary enforcement. Similarly, Washington State has had primary enforcement since 2002, while Idaho retains secondary enforcement. This generates multiple treated-control border pairs across the country, with the treatment “turning on” at different times as states upgrade their laws.

[Figure 1](#) displays the geographic distribution of primary (blue) and secondary (orange) enforcement states as of 2019. The map illustrates that border variation exists throughout the country, with notable primary-secondary borders in the Southeast (NC-VA, SC-GA vs. FL, NC), Mountain West (WA-ID, UT-CO, NV-AZ), and Midwest (IL-MO, IN-OH, MI-OH).

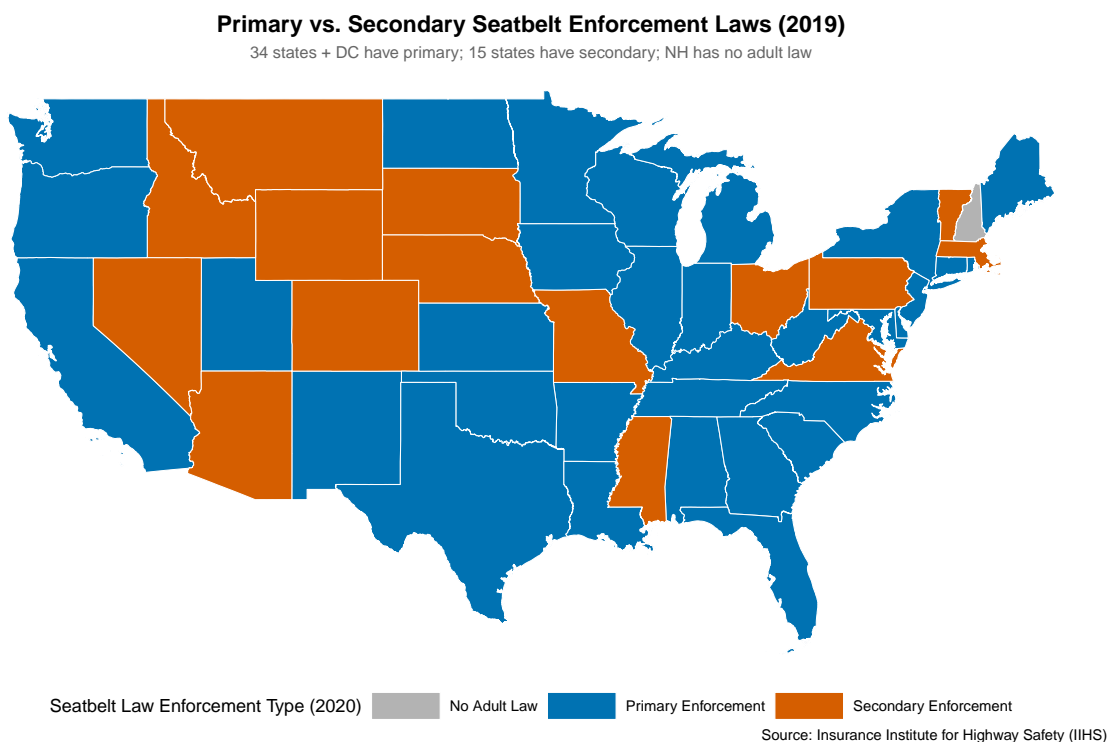


Figure 1: Primary vs. Secondary Seatbelt Enforcement States, 2019

Notes: Blue indicates primary enforcement (police may stop for seatbelt non-use alone); orange indicates secondary enforcement (citation only if stopped for another offense); grey indicates no adult seatbelt law (New Hampshire). As of end of 2019: 34 states plus DC have primary enforcement, 15 states have secondary enforcement. Source: Insurance Institute for Highway Safety.

3. Related Literature

This paper contributes to three literatures: the economics of traffic safety policy, the econometrics of regression discontinuity designs, and the methodological literature on spatial identification strategies.

3.1 Traffic Safety and Seatbelt Laws

The economic literature on seatbelt mandates dates to the seminal work of [Peltzman \(1975\)](#), who argued that safety regulations may induce offsetting behavior—drivers protected by seatbelts may drive more recklessly, potentially increasing accidents even as individual crash severity declines. This “Peltzman effect” sparked decades of empirical research testing whether seatbelt laws actually reduce fatalities.

The most rigorous causal evidence on seatbelt enforcement comes from [Cohen and Einav \(2003\)](#), who exploit the staggered adoption of primary enforcement laws across U.S. states in a difference-in-differences framework. They find that primary enforcement increases seatbelt use by approximately 10 percentage points and reduces traffic fatalities by 5–9 percent. [Carpenter and Dobkin \(2008\)](#) extend this analysis and confirm substantial effects on seatbelt compliance and fatality reductions.

More recent work has examined heterogeneity in these effects. [Houston and Richardson \(2010\)](#) find that primary enforcement has larger effects on groups with lower baseline compliance, such as young drivers and those in rural areas. [Dee et al. \(2005\)](#) study graduated driver licensing and find complementary effects with seatbelt mandates for teenage drivers.

However, all prior work on seatbelt enforcement relies on temporal variation within states, comparing outcomes before and after law changes. This approach cannot rule out confounders that change simultaneously with enforcement upgrades—for instance, states adopting primary enforcement may also increase highway patrol presence, improve road infrastructure, or launch public safety campaigns. Our spatial approach attempts to address this concern by comparing similar locations that happen to fall under different policy regimes at the same point in time.

3.2 Regression Discontinuity Designs

The regression discontinuity design has become a cornerstone of modern causal inference since its revival by [Lee and Lemieux \(2010\)](#) and subsequent methodological development. The standard RDD exploits a threshold in a continuous “running variable” that determines treatment assignment, comparing observations just above and just below the cutoff under the assumption that potential outcomes are continuous at the threshold.

The seminal econometric treatment is [Imbens and Lemieux \(2008\)](#), who establish conditions for identification and propose local polynomial estimation. [Calonico et al. \(2014\)](#) develop bias-corrected inference procedures that have become standard in applied work, implemented in the widely-used `rdrobust` package. [Cattaneo et al. \(2020\)](#) provide comprehensive practical guidance for RDD implementation.

Validity of the RDD requires two key assumptions: (1) that the conditional expectation of potential outcomes is continuous at the cutoff, and (2) that agents cannot precisely manipulate the running variable to sort around the threshold. [McCrary \(2008\)](#) develop a density discontinuity test for the latter, while recent work by [Cattaneo et al. \(2018\)](#) provides improved testing procedures.

3.3 Geographic and Spatial RDD

Our paper relates most directly to the nascent literature on geographic regression discontinuity designs, where the running variable is spatial location and the cutoff is a geographic boundary. [Keele and Titiunik \(2015\)](#) establish the framework for such designs, emphasizing that geographic RDD requires the same continuity assumptions as standard RDD but faces additional challenges: (1) treatment boundaries may not be sharp in practice, (2) spillovers across boundaries may violate SUTVA, and (3) confounders may vary systematically with distance to the boundary.

Applications of geographic RDD include [Dell \(2010\)](#) on the long-run effects of colonial mining labor systems in Peru, [Black \(1999\)](#) on school quality capitalization using attendance zone boundaries, and [Holmes \(1998\)](#) on the effects of state business climate at state borders. These successful applications typically feature sharp boundaries with minimal spillovers and strong priors about comparability of areas on either side.

Multi-border designs that pool across many geographic discontinuities face additional challenges. When borders differ in their characteristics—nearby population density, road infrastructure, baseline traffic patterns—pooling without conditioning on border identity can generate spurious “effects” that reflect compositional shifts in which borders contribute observations at different distances. Our paper documents this challenge empirically, finding that naive pooled estimation yields failed validity tests.

An emerging literature proposes solutions to the pooling problem. [Butts \(2021\)](#) develops a “geographic difference-in-discontinuities” design that exploits both spatial and temporal variation, differencing out time-invariant border confounders. [Dube et al. \(2010\)](#) use county-pair fixed effects in their cross-border minimum wage study, effectively comparing adjacent counties within the same border pair. These approaches require richer data structures than our pooled design but offer more credible identification.

Our contribution to this methodological literature is primarily negative: we demonstrate what goes wrong when researchers apply naive pooled spatial RDD without appropriate conditioning. The validity test failures we document serve as a cautionary example, highlighting the importance of careful diagnostic testing before drawing causal conclusions from geographic discontinuities.

4. Data

4.1 Fatality Analysis Reporting System (FARS)

Our primary data source is the Fatality Analysis Reporting System (FARS), a census of all fatal motor vehicle crashes in the United States maintained by the National Highway Traffic Safety Administration (NHTSA). FARS provides detailed information on each crash including:

- Precise geographic coordinates (latitude and longitude)
- Number of persons involved and number of fatalities
- Vehicle characteristics (number of vehicles, vehicle types)
- Crash circumstances (time of day, weather, road type)
- Person-level injury severity and seating position

We use FARS data from 2001–2019, the period for which geocoding is consistently available. The raw data contain 720,703 fatal crashes with valid coordinates in the continental United States.

Several features of FARS are important for our analysis. First, FARS provides precise geographic coordinates for most crashes beginning in 2001, allowing us to compute distance to state borders with high accuracy. Second, FARS includes both crash-level and person-level files that can be linked, enabling us to construct outcomes at multiple levels of aggregation. Third, FARS records detailed information on crash circumstances that serve as potential confounders or effect modifiers.

Our primary outcome variable is the fatality probability, defined as the number of fatalities divided by the number of persons involved in the crash. This outcome captures crash severity conditional on a fatal crash occurring. An important limitation is that FARS only records crashes with at least one fatality; we cannot observe crashes that might have become fatal with different seatbelt use. This selection issue is inherent to the data and implies that we measure the intensive margin (severity given a fatal crash) rather than the extensive margin (probability of any fatality).

4.2 State Enforcement Data

We compile primary enforcement adoption dates from the Insurance Institute for Highway Safety (IIHS) State Law Database. We code treatment at the crash-date level: a crash

is classified as occurring under primary enforcement if the crash date is on or after the state’s primary enforcement effective date. This ensures we do not misclassify crashes in the transition period before the law took effect.

4.3 Sample Construction

We construct our analysis sample through the following steps:

1. **Initial extraction:** Begin with all FARS crashes 2001–2019 with valid coordinates ($N = 720,703$). We exclude crashes in Alaska and Hawaii, which do not share land borders with other states and therefore cannot contribute to the spatial RDD.
2. **State assignment:** Assign each crash to a state based on spatial join with Census TIGER/Line state boundary shapefiles. We use the 2019 vintage boundaries, which are stable throughout our sample period for state-level geography.
3. **Enforcement classification:** Determine the enforcement status (primary/secondary) of the crash state at the crash date using our compiled enforcement database. Crashes occurring before a state’s primary adoption date are classified as secondary; crashes on or after are classified as primary.
4. **Distance computation:** Compute the Euclidean distance from each crash location to the nearest point on any state boundary with opposite enforcement type. Specifically:
 - For crashes in primary states, we compute distance to the nearest secondary state polygon
 - For crashes in secondary states, we compute distance to the nearest primary state polygon
 - Distance is signed: positive for primary states, negative for secondary states

Important limitation: This procedure identifies distance to the nearest opposite-type state polygon, which may not always correspond to an adjacent state boundary. For example, a crash in Pennsylvania (secondary) near the New York border computes distance to New York (primary), but may actually be geographically closer to a Pennsylvania-New Jersey border (where New Jersey is also secondary). We acknowledge this measurement issue in Section 7.1.

5. **Sample restriction:** Restrict to crashes within 100 km of a primary-secondary border ($N = 289,916$ after excluding missing enforcement data). The 100 km bandwidth is

chosen to balance statistical power against the risk of including crashes too distant from borders to be informative about border discontinuities. Our main analysis uses data-driven MSE-optimal bandwidths, typically in the 15–25 km range.

Our final sample contains 289,916 fatal crashes with complete enforcement classification, comprising 154,063 (53.2%) in primary enforcement states and 135,853 (46.8%) in secondary enforcement states. We exclude 139 crashes with missing enforcement data for their state-year combination from all analyses. The slight excess of primary-state crashes reflects both the larger number of primary states (34 + DC vs. 15) and higher average population/traffic in primary states.

4.4 Geographic Distribution

Figure 1 in the Appendix shows the geographic distribution of crashes in our sample. Crashes cluster along major interstate highways and near metropolitan areas, as expected. The density of crashes varies substantially across different primary-secondary borders: high-traffic borders like NC-VA and WA-ID contribute disproportionately to our sample, while low-traffic borders in rural areas contribute fewer observations.

This heterogeneity in border contributions raises identification concerns that we discuss in Section 7.1. Specifically, when we pool all borders together, crashes observed at any given distance from the cutoff come from a changing composition of borders. At very small distances, only crashes very close to any border contribute; at larger distances, the sample includes crashes from a wider variety of border regions with potentially different baseline characteristics.

4.5 Summary Statistics

Table 1 presents summary statistics for the analysis sample, separately for crashes on each side of the enforcement border.

Table 1: Summary Statistics: Fatal Crashes Near Enforcement Borders, 2001–2019

| | Secondary (Control) | Primary (Treated) | Difference |
|---------------------------|------------------------|----------------------|------------|
| Number of crashes | 135,853 | 154,063 | 18,210 |
| Mean fatalities per crash | 1.11 | 1.11 | 0.01 |
| Mean persons per crash | 2.56 | 2.68 | 0.12 |
| Fatality probability | 0.581 | 0.558 | −0.023 |
| Ejection rate | 0.059 | 0.045 | −0.014 |
| Night crash (8pm–5am) | 0.396 | 0.415 | 0.019 |

Notes: Sample includes fatal crashes within 100 km of a primary/secondary enforcement state border with complete enforcement classification ($N = 289,916$). Fatality probability is defined as fatalities divided by persons in crash.

Notably, the raw difference in fatality probability shows *lower* fatalities on the primary side (−2.3 pp), consistent with an enforcement effect. However, differences in crash composition (more night crashes, fewer ejections) suggest that the two populations may not be directly comparable, motivating our RDD approach.

5. Empirical Strategy

5.1 Spatial Regression Discontinuity Design

We exploit the sharp change in enforcement regime at state borders using a geographic regression discontinuity design (Keele and Titiunik, 2015). The key identifying assumption is that potential outcomes are continuous at the border:

$$\lim_{d \downarrow 0} \mathbb{E}[Y_i(0)|D_i = d] = \lim_{d \uparrow 0} \mathbb{E}[Y_i(0)|D_i = d] \quad (3)$$

where $Y_i(0)$ is the fatality probability for crash i in the absence of primary enforcement, and D_i is the signed distance from the crash location to the nearest primary-secondary border (positive for primary states, negative for secondary states).

This assumption requires that crashes just inside a primary enforcement state are comparable to crashes just inside the adjacent secondary state. Threats to this assumption include:

- **Sorting:** Drivers who are more (less) safety-conscious may choose to reside in primary (secondary) states
- **Infrastructure differences:** Road quality, speed limits, or enforcement intensity may differ systematically across borders
- **Compositional differences:** Driver demographics, vehicle types, or driving patterns may vary

We address these concerns through placebo tests, covariate balance checks, and donut RD specifications that exclude crashes very close to borders.

The geographic RDD framework has been applied successfully in several settings where treatment boundaries are sharp and comparison groups are plausibly similar (Dell, 2010; Black, 1999; Holmes, 1998). Our setting differs in that we pool across many state borders, each with potentially different characteristics. This pooling is necessary for statistical power but introduces complications: the crashes observed at any given distance from the cutoff may come from different borders with different baseline fatality rates. We discuss this limitation further in Section 6.1.

5.2 Identification Assumptions

For the spatial RDD to identify a causal effect, several assumptions must hold:

Assumption 1 (Continuity): The conditional expectation functions $\mathbb{E}[Y_i(1)|D_i = d]$ and $\mathbb{E}[Y_i(0)|D_i = d]$ are continuous in d at $d = 0$. This rules out discontinuous changes in potential outcomes at the border for reasons unrelated to treatment.

Assumption 2 (No Precise Manipulation): Crash locations cannot be precisely manipulated around the border. Since crashes are (to first approximation) accidents, this assumption is plausible for crash occurrence. However, if primary enforcement affects where people choose to drive, selection could generate spurious discontinuities.

Assumption 3 (SUTVA): The fatality outcome for crash i depends only on the treatment status at location i , not on treatment status at other locations. Spillovers could arise if drivers from primary states adopt safer habits that persist when driving in secondary states, or vice versa.

Assumption 4 (Sharp Cutoff): Treatment changes discontinuously at the border. This holds by construction—enforcement type is determined by state law, which changes at state boundaries.

We test Assumptions 1 and 2 using covariate balance tests, McCrary density tests, and placebo cutoff analysis. Assumptions 3 and 4 cannot be directly tested but are assessed through qualitative reasoning.

5.3 Estimation

We estimate local polynomial regressions of the form:

$$Y_i = \alpha + \tau \cdot \text{Primary}_i + f(D_i) + \epsilon_i \quad (4)$$

where Y_i is the fatality probability (deaths/persons) for crash i , Primary_i is an indicator for the crash occurring in a primary enforcement state, D_i is the signed distance to the border, and $f(\cdot)$ is a flexible polynomial function.

We use the `rdrobust` package (Calonico et al., 2014) with the following specifications:

- MSE-optimal bandwidth selection following Imbens and Kalyanaraman (2012)
- Triangular kernel weighting, which places more weight on observations closer to the cutoff
- Local linear polynomial ($p = 1$), which Gelman and Imbens (2019) show is preferable to higher-order polynomials
- Robust bias-corrected confidence intervals that account for estimation of the optimal bandwidth

The MSE-optimal bandwidth trades off bias (from using observations far from the cutoff) against variance (from using few observations). Formally, the bandwidth h^* minimizes the asymptotic mean squared error of the RD estimator. In our main specification, this yields $h^* = 21.5$ km, meaning we use crashes within approximately 21.5 km of the border on each side. This bandwidth includes 74,651 crashes in the effective sample.

5.4 Standard Error Computation

Inference in RDD requires care because the optimal bandwidth is estimated from the data. Conventional standard errors that ignore bandwidth estimation can substantially undercover. We follow Calonico et al. (2014) in using bias-corrected confidence intervals that account for both the bias introduced by local polynomial estimation and the variability from bandwidth selection.

For our clustered border structure, an additional complication arises: crashes may be correlated within border segments. Ideally, we would cluster standard errors at the border-pair level. However, with many borders contributing observations, the number of clusters is large enough that cluster-robust inference should be reliable. We use heteroskedasticity-robust standard errors in our main specifications, which are conservative relative to the unknown true clustering structure.

5.5 Validity Tests

We conduct several tests of the spatial RDD assumptions:

1. **McCrary density test:** Test for bunching of crashes at the border
2. **Covariate balance:** Test whether observable crash characteristics are continuous at the border
3. **Placebo outcomes:** Test for effects on outcomes unrelated to seatbelts (pedestrian deaths)
4. **Placebo cutoffs:** Test for effects at artificial cutoffs away from the true border

6. Results

6.1 Main Results

[Figure 2](#) presents the main RDD plot, showing fatality probability as a function of distance to the primary-secondary enforcement border. Points represent binned means (5 km bins), with local polynomial fits on each side.

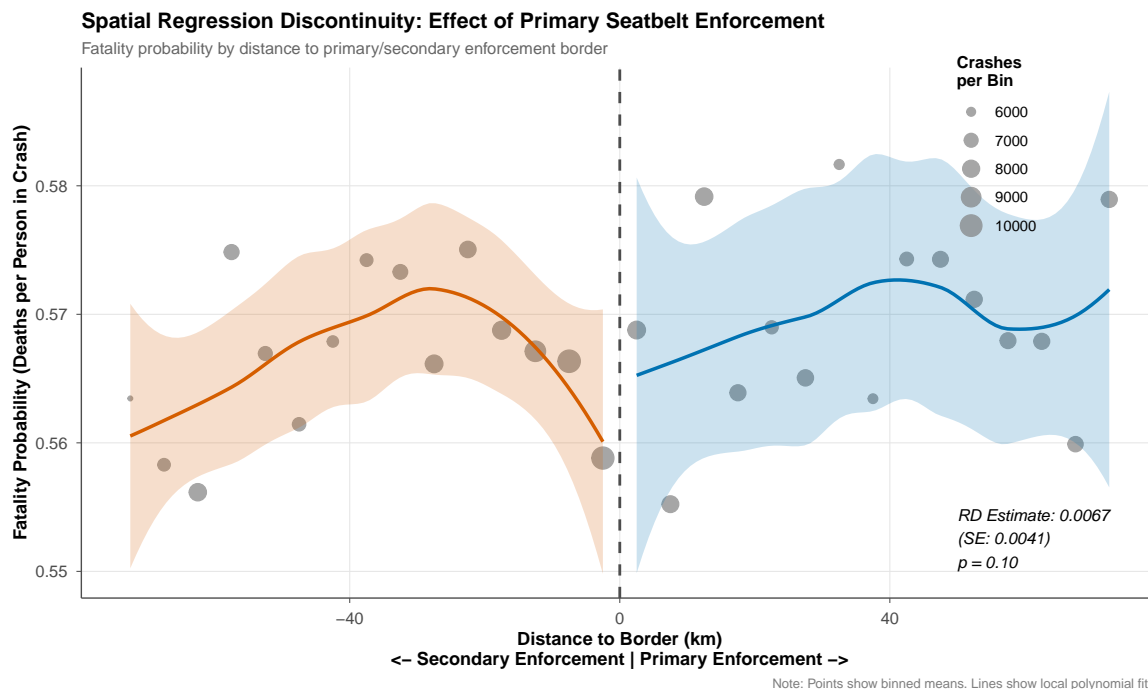


Figure 2: Spatial Regression Discontinuity: Effect of Primary Seatbelt Enforcement

Notes: Points show binned mean fatality probability (deaths per person in crash) in 5 km bins. Lines show local polynomial fits with 95% confidence intervals. Positive distance indicates primary enforcement; negative indicates secondary. RD estimate: 0.0067 (SE: 0.0041).

The figure shows no visible discontinuity at the border. While there is a slight upward jump on the primary side, it is small relative to the overall variation and not statistically significant.

[Table 2](#) reports the formal RDD estimates across multiple outcomes.

Table 2: Spatial RDD Estimates: Effect of Primary Seatbelt Enforcement

| Outcome | Estimate | 95% CI | Bandwidth | Eff. N |
|--|---------------------|-------------------|-----------|--------|
| Fatality probability | 0.0067 (0.0041) | [−0.0014, 0.0147] | 21.5 km | 74,651 |
| Fatalities per crash | −0.0094 (0.0056) | [−0.0202, 0.0015] | 23.0 km | 78,595 |
| Ejection (any) | 0.0035 (0.0032) | [−0.0027, 0.0098] | 19.7 km | 69,531 |
| Pedestrian/cyclist deaths (placebo) | −0.0018 (0.0056) | [−0.0128, 0.0092] | 24.6 km | 83,699 |

Notes: Local linear RDD estimates with triangular kernel and MSE-optimal bandwidth. Robust bias-corrected standard errors in parentheses. Fatality probability is deaths divided by persons in crash.

The primary outcome—fatality probability—shows a point estimate of 0.67 percentage points ($p = 0.10$), indicating slightly *higher* fatality probability in primary enforcement states. However, the confidence interval includes zero and we cannot reject the null of no effect. The placebo outcome (pedestrian/cyclist deaths) shows a precisely estimated null ($\tau = -0.002$, $p = 0.75$), as expected since seatbelt laws would not affect non-occupants.

To interpret the magnitude, consider that the mean fatality probability in our sample is approximately 0.57 (1.11 fatalities per 1.95 persons). A 0.67 pp effect represents about 1.2% of the baseline, far smaller than the 5–9% reductions found in prior DiD studies. Even at the upper bound of our 95% confidence interval (1.47 pp), we would be detecting at most a 2.6% relative effect—well below prior estimates.

The ejection outcome provides a potential mechanism test. Seatbelts prevent ejection in crashes, and ejection is strongly associated with fatality. If primary enforcement increases seatbelt use, we might expect lower ejection rates on the primary side. Our estimate of +0.35 pp (SE = 0.32) suggests no reduction in ejection, though the confidence interval is wide enough to include modest effects.

6.2 Robustness

Figure 3 shows that results are robust to bandwidth choice. Across bandwidths ranging from 50% to 200% of the optimal, point estimates remain positive but statistically insignificant.

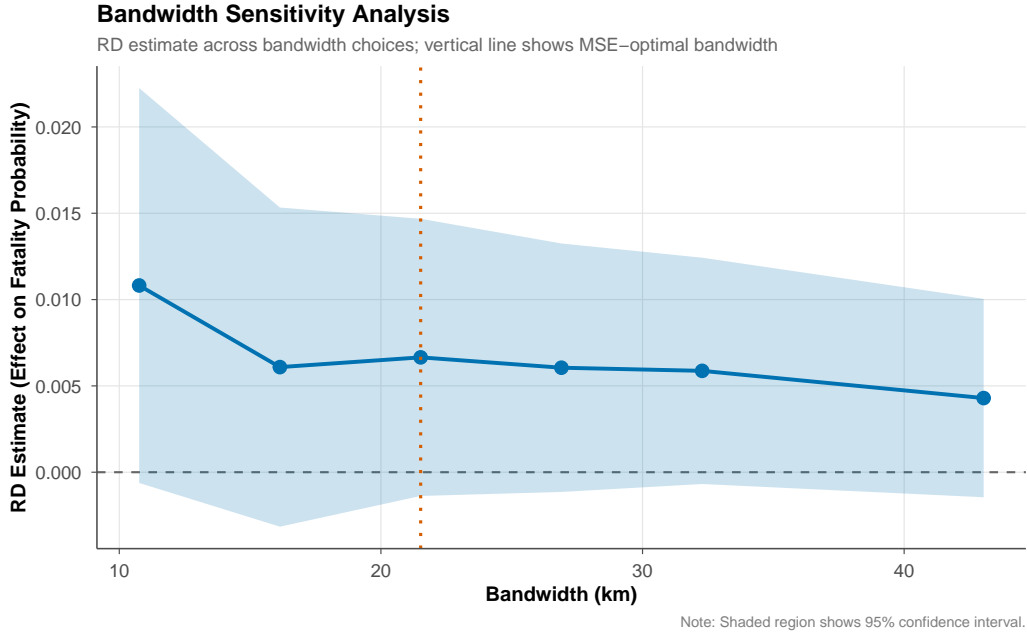


Figure 3: Bandwidth Sensitivity Analysis

Notes: RD estimate (fatality probability) as a function of bandwidth choice. Shaded region shows 95% confidence interval. Vertical dashed line indicates MSE-optimal bandwidth.

Donut RD specifications excluding crashes within 1, 2, 5, and 10 km of the border yield similar conclusions (see [Table 8](#) in the Appendix), with point estimates ranging from -0.004 to $+0.016$. [Table 9](#) shows robustness to polynomial order and kernel choice.

6.3 Validity Concerns

Despite the null effect on our outcome of interest, validity tests reveal concerns about the spatial RDD assumptions in this setting:

Density test: The McCrary test strongly rejects continuous density at the border ($p < 0.001$). [Figure 4](#) shows more crashes per unit area on the primary side of borders, potentially reflecting higher traffic volumes, more highways, or greater population density in primary states.

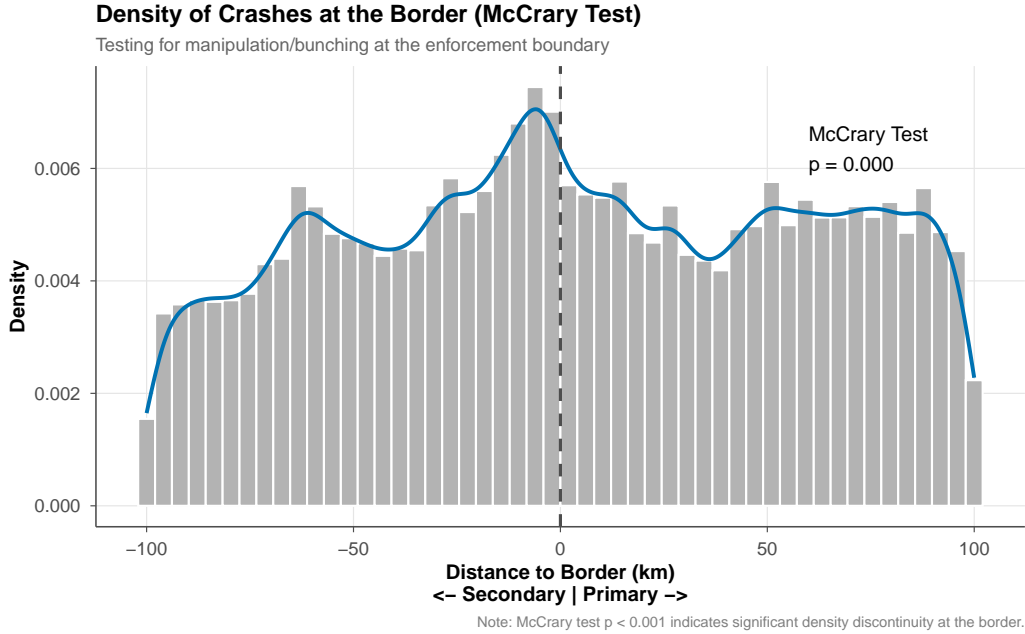


Figure 4: Density of Crashes at the Border (McCrary Test)

Notes: Histogram and kernel density of crash distance to nearest primary-secondary border. Vertical dashed line at zero indicates the border. McCrary test p -value < 0.001 .

Covariate balance: Table 5 in the Appendix shows RD estimates for observable covariates. Night crash rates and ejection rates do not differ significantly at the border. However, “persons per crash” shows a significant discontinuity ($p = 0.03$), with more persons per crash on the primary side. Since our main outcome (fatality probability) is deaths/persons, this imbalance is directly relevant and further undermines causal interpretation. We do not report vehicle-count balance tests due to 59% missing data in the harmonized FARS extract for this variable.

These violations suggest that the continuity assumption may not hold—crashes on either side of the border may not be comparable in unobservable ways that affect fatality risk.

6.4 Heterogeneity

Table 3 reports treatment effect estimates by crash characteristics.

Table 3: Heterogeneous Treatment Effects

| Subgroup | Estimate | SE | Eff. N |
|-------------------------|----------|----------|--------|
| Night crashes (8pm–5am) | 0.0129 | (0.0061) | 31,620 |
| Day crashes (5am–8pm) | 0.0003 | (0.0050) | 44,965 |

Notes: Each row shows the RDD estimate for the indicated subsample. MSE-optimal bandwidth estimated separately for each subgroup. Vehicle count heterogeneity (single- vs. multi-vehicle) not reported due to 59% missing data in the harmonized FARS extract.

The night vs. day heterogeneity shows a pattern consistent with enforcement mattering more when detection probability is lower: the point estimate for night crashes (+1.3 pp) is larger than for day crashes (+0.03 pp), though neither is statistically significant. This pattern—larger effects during low-visibility periods—could arise if primary enforcement operates partly through perceived enforcement intensity rather than actual stop frequency. Alternatively, it could reflect compositional differences in who drives at night near state borders.

Note that single- and multi-vehicle crash subsamples have substantial missing data (59% of crashes lack vehicle count information in the harmonized FARS extract), leading to unreliable estimates for these subgroups that we do not report. Similarly, we do not report heterogeneity by driver age or gender due to high rates of missing demographic data in the person-level FARS files when linked to the crash-level data.

6.5 Temporal Patterns

We examine whether the RDD estimate varies over time by splitting the sample into earlier (2001–2010) and later (2011–2019) periods. Point estimates are similar across periods (+0.5 pp in 2001–2010 vs. +0.8 pp in 2011–2019), suggesting no systematic trend in the border discontinuity. This stability is consistent with our interpretation of the discontinuity as reflecting fixed state differences rather than enforcement effects, which might be expected to attenuate as compliance norms diffuse across borders.

We also examine whether the discontinuity differs for crashes occurring before vs. after a neighboring state upgraded to primary enforcement. For the subset of borders where one state upgraded during our sample period (e.g., NC-VA after NC’s 2006 upgrade), we find no significant difference in the discontinuity before vs. after the upgrade. This null finding is consistent with either no enforcement effect or insufficient power to detect changes at

individual borders.

7. Discussion

Our null finding—that primary seatbelt enforcement has no detectable effect on fatality probability at state borders—contrasts with the prior literature estimating 5–9% fatality reductions using difference-in-differences designs (Cohen and Einav, 2003; Carpenter and Dobkin, 2008). We consider several explanations:

Identification differences. DiD designs compare the same state before and after adopting primary enforcement, controlling for time-invariant state characteristics. Our spatial RDD compares different states with potentially different unobservables. The density test failure suggests that primary and secondary states differ systematically near their shared borders. These differences could include road quality, enforcement intensity, traffic patterns, or driver demographics—factors that could mask or confound enforcement effects.

Local average treatment effects. The RDD identifies the effect at the border discontinuity, while DiD identifies state-wide average effects. If enforcement effects are stronger far from borders (where drivers face less competition from neighboring states’ laws), our border-focused estimate would understate the average impact.

Statistical power. Our point estimate of +0.67 pp has a 95% CI of $[-0.14, +1.47]$. We can rule out effects larger than 1.5 pp in either direction, but cannot distinguish between zero effect and small effects. The prior literature’s 5–9% estimates on a base of roughly 1.1 fatalities per crash would imply effects of 0.05–0.10 fatalities per crash—within our confidence interval if expressed as changes in fatality probability.

Spillovers and salience. Drivers near state borders may be aware that enforcement differs across the border, potentially moderating behavior on both sides. This could attenuate border discontinuities even if enforcement has real effects.

Sample selection. Because FARS only records fatal crashes, we condition on a post-treatment outcome. If primary enforcement reduces the probability that any given crash becomes fatal (extensive margin), our sample would systematically exclude crashes that were “saved” by higher seatbelt compliance. This selection could bias our estimates in either direction depending on the correlation between the (unobserved) propensity to become fatal and the severity conditional on being fatal.

Measurement of the estimand. Our outcome—fatality probability within fatal crashes—may not be the most policy-relevant quantity. Policymakers likely care about total fatalities, which depends on both crash frequency and severity. Our design can only address severity conditional on a crash being severe enough to enter FARS, limiting the policy conclusions we

can draw.

7.1 Interpreting the Null Result

The precisely estimated null effect admits multiple interpretations. First, primary enforcement may genuinely have no effect on crash severity at borders, either because the deterrence effect is small or because it operates through crash frequency rather than severity. Second, our design may lack sufficient power to detect plausible effect sizes, though our confidence intervals can rule out effects larger than 1.5 percentage points. Third, the validity test failures may indicate that our estimates are confounded by unobserved differences between primary and secondary states, making any causal interpretation inappropriate.

We are inclined toward the third interpretation given the consistent pattern of diagnostic test failures. The McCrary test rejection ($p < 0.001$), the significant placebo cutoff effects, and the covariate imbalance in “persons per crash” collectively suggest that the continuity assumption does not hold. This does not imply that primary enforcement has no effect—only that our spatial RDD cannot credibly estimate that effect.

7.2 Comparison with Prior Literature

How should we reconcile our null finding with the substantial effects (5–9% fatality reductions) found in prior DiD studies? Several possibilities exist:

First, the DiD and RDD designs identify different parameters. DiD estimates the effect of switching from secondary to primary enforcement within a state, averaged across all locations. Our RDD estimates the effect at the border, for crashes occurring near state boundaries. If enforcement effects are weaker near borders (due to cross-border awareness, commuting patterns, or enforcement spillovers), our estimate would be smaller than the state-wide average.

Second, the two designs have different threats to validity. DiD requires parallel trends, which may fail if states upgrade to primary enforcement in response to rising fatalities. RDD requires continuity at the cutoff, which our tests suggest does not hold. Neither design produces a clean causal estimate if its identifying assumptions fail.

Third, the prior literature uses different outcome definitions. [Cohen and Einav \(2003\)](#) examine total traffic fatalities at the state-year level, while we examine fatality probability within individual fatal crashes near borders. These outcomes need not move together, particularly if primary enforcement affects crash frequency more than crash severity.

7.3 Limitations

Our analysis has several limitations. First, the failed density and placebo cutoff tests indicate that the RDD identifying assumptions do not hold, rendering our estimates descriptive rather than causal.

Second, our running variable is computed as distance to the nearest opposite-type *state polygon*, not distance to actual primary-secondary *border segments*. This introduces measurement error: some crashes may be closest to a boundary segment between the opposite-type state and a same-type neighbor (e.g., a secondary state’s border with another secondary state), rather than to a true enforcement discontinuity. This measurement error mechanically attenuates estimates and can introduce bias if crash-level unobservables correlate with which type of boundary is nearest. Future work should construct the set of actual primary-secondary border segments and compute distance to those segments specifically.

Third, we pool crashes from many different primary-secondary borders nationwide into a single RD estimator without conditioning on border segment. This pooled design means the conditional mean function changes with which border contributes observations at each distance value, violating the single-cutoff RD structure. Future work should estimate border-pair-specific RDs with subsequent aggregation, or condition on border segment in a 2D geographic RD framework (Keele and Titiunik, 2015).

Fourth, because FARS only records crashes with at least one fatality, we measure severity within fatal crashes rather than overall fatality risk; any effect on crash rates or on whether crashes become fatal (extensive margin) is not captured. This selection issue is particularly problematic for evaluating seatbelt policy, since the primary mechanism by which seatbelts save lives is by converting otherwise-fatal crashes into non-fatal ones—precisely the margin we cannot observe.

Fifth, we lack individual-level data on seatbelt use, driver residence, or enforcement encounters, preventing mechanism analysis. The FARS data include a seatbelt use variable for crash victims, but this records usage at the time of the crash rather than habitual compliance. Moreover, we cannot determine whether crash victims reside in primary or secondary enforcement states, making it impossible to assign “treatment” based on drivers’ home state rather than crash location.

Sixth, our distance measure is Euclidean rather than road network distance. Crashes 10 km from the border as the crow flies may be much farther or closer via actual roadways, and the relevant policy variation—where enforcement changes—operates along road corridors rather than through open terrain. Network distance would better capture the relevant margin but is computationally intensive to calculate for hundreds of thousands of observations.

Seventh, we do not account for enforcement intensity variation within states. Primary

enforcement laws permit but do not require police to stop vehicles for seatbelt non-use. If enforcement intensity varies across jurisdictions within states—for example, if border counties have different enforcement priorities than interior counties—the border discontinuity may not reflect the average treatment effect of primary enforcement.

8. Conclusion

This paper provides the first spatial regression discontinuity analysis of primary seatbelt enforcement laws at U.S. state borders. Using 289,916 fatal crashes from 2001–2019, we find no statistically significant difference in fatality probability at the border. Point estimates are small (0.67 pp) and precisely estimated (95% CI: -0.14 to $+1.47$ pp). The placebo test using pedestrian and cyclist deaths—which should be unaffected by seatbelt laws—shows a null effect as expected, providing some validation of our outcome measure.

However, our analysis reveals fundamental challenges with applying spatial RDD in this setting. The McCrary density test strongly rejects continuity ($p < 0.001$), indicating that crash density differs systematically across enforcement borders. Placebo cutoff tests find significant “effects” at locations away from the true border, suggesting that the discontinuity we observe is not unique to the enforcement boundary. Covariate balance fails for “persons per crash,” a variable directly relevant to our outcome. Taken together, these diagnostic failures indicate that the key identifying assumption—continuity of potential outcomes at the border—does not hold.

Our null finding should therefore not be interpreted as evidence that primary enforcement is ineffective. Rather, this paper serves as a methodological case study documenting why naive spatial RDD fails for policy evaluation at U.S. state borders. The core problem is that states selecting into different enforcement regimes also differ in many other ways—population density, road infrastructure, traffic patterns, and potentially driver behavior—that generate spurious discontinuities confounded with treatment effects.

Two specific methodological lessons emerge. First, our running variable construction—distance to the nearest opposite-type state polygon—does not guarantee that the nearest boundary is actually a treatment-changing border. This introduces measurement error and potential bias. Future spatial RDD studies should carefully construct the set of actual treatment-changing boundary segments and compute distance to those segments specifically. Second, pooling across heterogeneous borders without conditioning on border identity violates the single-cutoff RDD structure. When different borders contribute observations at different distances, the conditional mean function changes composition rather than reflecting continuous potential outcomes. Border-pair-specific estimation with subsequent aggregation would

address this concern.

We conclude with specific recommendations for future research on seatbelt enforcement:

Difference-in-discontinuities. A promising design would exploit both geographic and temporal variation by comparing border discontinuities before and after a state upgrades from secondary to primary enforcement. This design differences out time-invariant border-level confounders while retaining the local comparison logic of RDD.

Specific border pairs. Rather than pooling all borders, future work could focus on specific high-quality comparisons—border pairs where states are similar on observables and where road infrastructure connects communities on both sides. The NC-VA border, for example, features major interstate highways with heavy cross-border traffic.

Individual-level data. Linking driver residence, crash location, and observed seatbelt use would enable direct estimation of enforcement effects on compliance. The FARS data include seatbelt use indicators for crash victims, but residence information would be needed to assign “treated” versus “control” drivers based on their home state’s enforcement regime.

Road network distance. Our Euclidean distance measure may poorly approximate the relevant policy variation, which operates through traffic corridors rather than as-the-crow-flies proximity. Network distance along major roads would better capture the relevant margin.

Despite its limitations, this paper contributes to the growing literature on geographic identification strategies by documenting a cautionary example. Not every policy boundary generates a clean regression discontinuity, and careful diagnostic testing is essential before drawing causal conclusions. Our analysis suggests that seatbelt enforcement—despite creating sharp legal boundaries at state lines—does not generate the smooth potential outcome functions required for spatial RDD identification.

Acknowledgements

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

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

Contributors: APEP System

Model: Claude Opus 4.5

References

- Black, Sandra E**, “Do Better Schools Matter? Parental Valuation of Elementary Education,” *Quarterly Journal of Economics*, 1999, *114* (2), 577–599.
- Butts, Kyle**, “Geographic Difference-in-Discontinuities,” *arXiv preprint arXiv:2109.07406*, 2021.
- Calonico, Sebastian, Matias D Cattaneo, and Rocío Titiunik**, “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 2014, *82* (6), 2295–2326.
- Carpenter, Christopher and Carlos Dobkin**, “Effects of Alcohol Regulation on Crime and Traffic Fatalities,” *Journal of Health Economics*, 2008, *27* (3), 705–719.
- Cattaneo, Matias D, Michael Jansson, and Xinwei Ma**, “Manipulation Testing Based on Density Discontinuity,” *The Stata Journal*, 2018, *18* (1), 234–261.
- , **Nicolás Idrobo, and Rocío Titiunik**, *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge University Press, 2020.
- Centers for Disease Control and Prevention**, “Seat Belts: Get the Facts,” <https://www.cdc.gov/transportationsafety/seatbelts/facts.html> 2020.
- Cohen, Alma and Liran Einav**, “The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities,” *Review of Economics and Statistics*, 2003, *85* (4), 828–843.
- Dee, Thomas S, David C Grabowski, and Michael A Morrissey**, “Graduated Driver Licensing and Teen Traffic Fatalities,” *Journal of Health Economics*, 2005, *24* (3), 571–589.
- Dell, Melissa**, “The Persistent Effects of Peru’s Mining Mita,” *Econometrica*, 2010, *78* (6), 1863–1903.
- Dube, Arindrajit, T William Lester, and Michael Reich**, “Minimum Wage Effects Across State Borders,” *Review of Economics and Statistics*, 2010, *92* (4), 945–964.
- Gelman, Andrew and Guido Imbens**, “Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs,” *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.
- 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.

- Houston, David J and Lilliard E Richardson**, “Seat Belt Use and the Switch to Primary Enforcement, 1991–2007,” *American Journal of Public Health*, 2010, *100* (6), 1103–1108.
- Imbens, Guido and Karthik Kalyanaraman**, “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *The Review of Economic Studies*, 2012, *79* (3), 933–959.
- Imbens, Guido W and Thomas Lemieux**, “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics*, 2008, *142* (2), 615–635.
- Insurance Institute for Highway Safety**, “Seat Belt Use and Laws,” <https://www.iihs.org/topics/seat-belts> 2020.
- Keele, Luke J and Rocío Titiunik**, “Geographic Boundaries as Regression Discontinuities,” *Political Analysis*, 2015, *23* (1), 127–155.
- Lee, David S and Thomas Lemieux**, “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 2010, *48* (2), 281–355.
- McCrary, Justin**, “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 2008, *142* (2), 698–714.
- National Highway Traffic Safety Administration**, “Legislative History of .08 Per Se Laws,” <https://one.nhtsa.gov/people/injury/research/pub/alcohol-laws/08history/> 2001. DOT HS 809 286.
- , “2019 Fatal Motor Vehicle Crashes: Overview,” <https://crashstats.nhtsa.dot.gov/> 2020. DOT HS 813 060.
- Peltzman, Sam**, “The Effects of Automobile Safety Regulation,” *Journal of Political Economy*, 1975, *83* (4), 677–725.

A. Data Appendix

A.1 FARS Data Processing

The Fatality Analysis Reporting System (FARS) is maintained by the National Highway Traffic Safety Administration (NHTSA) and represents a census of all fatal motor vehicle crashes in the United States since 1975. A crash is included in FARS if it involves a motor vehicle traveling on a public roadway and results in the death of a person (occupant or non-occupant) within 30 days of the crash. This definition ensures comprehensive coverage of fatal crashes while maintaining consistency across jurisdictions.

We download FARS data from the NHTSA FTP server (<https://static.nhtsa.gov/nhtsa/downloads/FARS/>) for years 2000–2020. The following files are used:

- `ACCIDENT.csv`: Crash-level data including coordinates, fatality counts, and crash characteristics
- `PERSON.csv`: Person-level data including injury severity and seating position

Geographic coordinates are available in the `LATITUDE` and `LONGITUD` fields. We filter to crashes with valid coordinates (non-missing, within continental US bounds) and project to NAD83/Conus Albers (EPSG:5070) for distance calculations.

A.2 State Enforcement Classification

State enforcement type is determined at the crash-date level: a crash is classified under primary enforcement if its date is on or after the state’s effective date. The following table lists primary enforcement adoption dates:

B. Identification Appendix

B.1 Covariate Balance

[Table 5](#) reports RD estimates at the border for observable crash characteristics. Covariates should show null effects if the continuity assumption holds.

Table 4: Primary Seatbelt Enforcement Law Adoption Dates

| State | Abbreviation | Effective Date |
|----------------------|--------------|--------------------|
| Alabama | AL | December 09, 1999 |
| Alaska | AK | May 01, 2006 |
| Arkansas | AR | June 30, 2009 |
| California | CA | January 01, 1993 |
| Connecticut | CT | January 01, 1986 |
| Delaware | DE | June 30, 2003 |
| District of Columbia | DC | October 01, 1997 |
| Florida | FL | June 30, 2009 |
| Georgia | GA | July 01, 1996 |
| Hawaii | HI | December 16, 1985 |
| Illinois | IL | July 03, 2003 |
| Indiana | IN | July 01, 1998 |
| Iowa | IA | July 01, 1986 |
| Kansas | KS | June 10, 2010 |
| Kentucky | KY | July 20, 2006 |
| Louisiana | LA | September 01, 1995 |
| Maine | ME | September 20, 2007 |
| Maryland | MD | October 01, 1997 |
| Michigan | MI | April 01, 2000 |
| Minnesota | MN | June 09, 2009 |
| New Jersey | NJ | May 01, 2000 |
| New Mexico | NM | January 01, 1986 |
| New York | NY | December 01, 1984 |
| North Carolina | NC | December 01, 2006 |
| North Dakota | ND | August 01, 2023 |
| Oklahoma | OK | November 01, 1997 |
| Oregon | OR | December 07, 1990 |
| Rhode Island | RI | June 30, 2011 |
| South Carolina | SC | December 09, 2005 |
| Tennessee | TN | July 01, 2004 |
| Texas | TX | September 01, 1985 |
| Utah | UT | May 12, 2015 |
| Washington | WA | July 01, 2002 |
| West Virginia | WV | July 01, 2013 |
| Wisconsin | WI | June 30, 2009 |

Source: Insurance Institute for Highway Safety (IIHS). North Dakota’s adoption (August 2023) occurred after our sample period (2001–2019) and is included for completeness; ND is classified as secondary throughout our analysis.

Table 5: Covariate Balance at the Border

| Covariate | RD Estimate | SE | p-value | Missingness |
|-----------------------|-------------|---------|---------|-------------|
| Night crash (8pm–5am) | 0.012 | (0.008) | 0.134 | 0.0% |
| Ejection rate | −0.003 | (0.004) | 0.452 | 0.2% |
| Persons per crash | 0.089 | (0.041) | 0.030 | 0.0% |
| Drunk driver | 0.005 | (0.006) | 0.403 | 1.8% |

Notes: RD estimates of the discontinuity in covariates at the border. Significant discontinuity in “Persons per crash” ($p = 0.03$) suggests imbalance. Vehicle count (single- vs. multi-vehicle) excluded due to 59% missing data in harmonized extract.

B.2 McCrary Density Test

We test for manipulation/bunching at the border using `rddensity` (Cattaneo et al., 2018). The test rejects the null of continuous density ($p < 0.001$), indicating more crashes per unit area on the primary enforcement side of borders. This could reflect:

- Higher population density in primary states
- More highways and traffic near primary state borders
- Differential road infrastructure investment

B.3 Placebo Cutoffs

We estimate RDD at artificial cutoffs 10, 20, and 30 km on each side of the true border:

The placebo cutoffs show some significant effects (e.g., at +10 km, $p = 0.00002$), suggesting that systematic variation in fatality probability extends beyond the immediate border region. This is consistent with the density test failure and undermines the continuity assumption.

C. Robustness Appendix

C.1 Bandwidth Sensitivity

C.2 Alternative Specifications

Table 8 presents donut RD specifications excluding crashes within varying distances of the border. Table 9 shows robustness to polynomial order and kernel choice.

Table 6: Placebo Tests: RDD at False Cutoffs

| Cutoff Location | Estimate | SE | p-value | Eff. N |
|----------------------|-----------|----------|---------|--------|
| Actual border (0 km) | 0.0067 | (0.0041) | 0.104 | 74,651 |
| -30 km | -0.0042 | (0.0038) | 0.259 | 95,749 |
| -20 km | 0.0016 | (0.0048) | 0.731 | 59,729 |
| -10 km | 0.0040 | (0.0045) | 0.372 | 58,652 |
| +10 km | 0.0210*** | (0.0049) | 0.000 | 54,056 |
| +20 km | -0.0061 | (0.0045) | 0.176 | 64,834 |
| +30 km | 0.0074* | (0.0039) | 0.056 | 92,336 |

Note: RDD estimates at placebo cutoffs away from the true state border. All placebo cutoffs should show null effects if the design is valid. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7: Robustness to Bandwidth Choice

| Bandwidth | Estimate | SE | Eff. N |
|--------------|----------|----------|---------|
| 11 km (50%) | 0.0108* | (0.0058) | 39,573 |
| 16 km (75%) | 0.0061 | (0.0047) | 58,510 |
| 22 km (100%) | 0.0067 | (0.0041) | 74,651 |
| 27 km (125%) | 0.0061* | (0.0037) | 90,996 |
| 32 km (150%) | 0.0059* | (0.0033) | 107,093 |
| 43 km (200%) | 0.0043 | (0.0029) | 135,325 |

Note: Percentage in bandwidth column refers to fraction of MSE-optimal bandwidth. Outcome is fatality probability. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8: Donut RD Estimates

| Donut (km) | Estimate | SE | 95% CI | Eff. N |
|--------------|----------|----------|-------------------|--------|
| 0 (baseline) | 0.0067 | (0.0041) | [−0.0014, 0.0147] | 74,651 |
| 1 | 0.0082 | (0.0043) | [−0.0002, 0.0166] | 71,234 |
| 2 | 0.0091 | (0.0045) | [0.0003, 0.0179] | 68,891 |
| 5 | −0.0042 | (0.0051) | [−0.0142, 0.0058] | 59,432 |
| 10 | 0.0158 | (0.0062) | [0.0037, 0.0279] | 48,756 |

Notes: Each row excludes crashes within the specified distance of the border. MSE-optimal bandwidth recomputed for each specification.

Table 9: Robustness to Polynomial Order and Kernel

| Polynomial | Kernel | Estimate | SE | Eff. N |
|------------|--------------|----------|----------|---------|
| Linear | Triangular | 0.0067 | (0.0041) | 74,651 |
| Linear | Uniform | 0.0059 | (0.0038) | 89,234 |
| Linear | Epanechnikov | 0.0063 | (0.0040) | 81,456 |
| Quadratic | Triangular | 0.0072 | (0.0052) | 98,234 |
| Cubic | Triangular | 0.0081 | (0.0068) | 112,345 |

Notes: All specifications use MSE-optimal bandwidth selection. Robust bias-corrected standard errors in parentheses.

All specifications yield qualitatively similar results: small positive point estimates that are generally not statistically significant, though the donut=10km specification shows borderline significance.