

Decriminalize, Then Recriminalize: Evidence from Colorado’s Fentanyl Policy Reversal

Autonomous Policy Evaluation Project and @dakoyana

January 2026

Abstract

We study Colorado’s unique two-stage policy experiment on fentanyl possession: reclassification of felony possession to misdemeanor in 2019 (HB 19-1263, effective May 28, 2019) followed by partial refelonization in 2022 (HB 22-1326, effective May 25, 2022). Using difference-in-differences with wild cluster bootstrap inference and synthetic control methods comparing Colorado to other U.S. states, we find statistically inconclusive effects of either policy change on overdose mortality. Colorado’s fentanyl deaths increased by 811% from 2018 to 2023 (130 to 1,184), but this trajectory closely mirrors national trends driven by illicit fentanyl supply. Point estimates suggest the 2019 reclassification increased total overdose deaths by approximately 17%, with analytical confidence intervals spanning -31% to +96%. Permutation inference shows this effect is unremarkable compared to regional trends ($p=0.72$), and the synthetic control analysis confirms these inconclusive findings (permutation $p = 0.96$). Our results highlight the difficulty of identifying demand-side policy effects amid massive supply-side shocks and underscore the importance of appropriate small-sample inference methods in drug policy evaluation.

JEL Codes: I18, K14, K42

Keywords: drug policy, decriminalization, fentanyl, overdose mortality, difference-in-differences

1 Introduction

The United States is experiencing an unprecedented fentanyl crisis. Synthetic opioids, primarily illicitly manufactured fentanyl, caused over 70,000 deaths in 2023—more than car accidents, gun violence, and HIV/AIDS combined at their peaks (?). States have adopted divergent responses, from decriminalization (Oregon’s Measure 110 in 2020) to enhanced criminal penalties. Yet rigorous causal evidence on these policies’ effects remains scarce.

Colorado offers a unique natural experiment. In 2019, House Bill 19-1263 reduced possession of less than 4 grams of controlled substances—including fentanyl—from a felony to a misdemeanor. Following a surge in overdose deaths, the legislature partially reversed course in 2022 with House Bill 22-1326, which made possession of more than 1 gram of fentanyl a felony again. This policy “reversal” provides rare variation to evaluate both decriminalization and recriminalization effects within a single state.

We use difference-in-differences methods comparing Colorado to neighboring states (New Mexico, Arizona, Utah, Wyoming, Nebraska, Kansas, and Oklahoma) from 2015 to 2024, complemented by synthetic control analysis using all U.S. states as potential donors. Recognizing that standard cluster-robust inference is unreliable with few clusters, we implement wild cluster bootstrap (?) and permutation inference following best practices for difference-in-differences with a single treated unit (?). Our main finding is statistically inconclusive: point estimates suggest the 2019 reclassification increased total overdose deaths by approximately 17% (log-point coefficient of 0.15), with analytical confidence intervals spanning -31% to +96%. Permutation inference, which tests whether Colorado’s trajectory is unusual compared to other states, yields $p = 0.72$, indicating that similarly-sized effects arise frequently under random treatment assignment. The synthetic control analysis reinforces these inconclusive findings, with Colorado’s post-treatment gap indistinguishable from placebo gaps (permutation $p = 0.96$).

This null finding is itself informative. Critics of decriminalization—including Colorado law enforcement and prosecutors who lobbied for HB 22-1326—pointed to the state’s rising

death toll as evidence that loosening penalties was “ludicrous” given fentanyl’s lethality (?). Our analysis challenges this narrative. While Colorado’s fentanyl deaths increased 811% from 2018 to 2023, neighboring states without policy changes experienced similar trajectories. The national fentanyl supply shock, accelerated by COVID-19 pandemic disruptions, appears to dominate any demand-side policy effects.

Our contribution is threefold. First, we provide the first quasi-experimental evaluation of a drug policy “reversal”—most policy changes are unidirectional, limiting our ability to assess reversibility. Second, we demonstrate that imprecise nulls can be policy-relevant: large standard errors reflect genuine uncertainty that should inform legislative debates. Third, we offer methodological lessons for drug policy evaluation in the fentanyl era, when nationwide supply shocks make state-level identification particularly challenging.

The paper proceeds as follows. Section 2 provides background on Colorado’s policy changes and the national fentanyl crisis. Section 3 describes our data sources. Section 4 presents our empirical strategy. Section 5 reports results. Section 6 discusses implications and limitations. Section 7 concludes.

2 Background

2.1 The Fentanyl Crisis

Illicitly manufactured fentanyl began appearing in U.S. drug markets around 2013 and has since transformed the overdose epidemic. Unlike heroin or prescription opioids, fentanyl is synthetic, cheap to produce, and extraordinarily potent—50 to 100 times stronger than morphine. A lethal dose can be as small as 2 milligrams. Drug traffickers increasingly mix fentanyl into heroin, counterfeit pills, and even stimulants, often without users’ knowledge.

The scale of the crisis is staggering. National overdose deaths involving synthetic opioids rose from 3,105 in 2013 to 73,838 in 2022 before declining slightly to 72,776 in 2023. The 2024 provisional data show a more substantial 27% decline, potentially marking a turning

point—though deaths remain far above pre-fentanyl levels.

This context is crucial for interpreting Colorado’s policy experiment. Any state-level policy effects must be identified against a backdrop of massive supply-side changes affecting all jurisdictions. The fentanyl surge was not gradual; it resembled a step function as the drug flooded regional markets at different times.

2.2 Colorado’s Policy Experiment

2.2.1 Phase 1: Decriminalization (2019)

House Bill 19-1263, signed into law on May 28, 2019, reduced possession of less than 4 grams of most Schedule I and II controlled substances from a Level 4 drug felony to a Level 1 drug misdemeanor. The law took effect immediately upon signature. The bill applied to nearly all illicit drugs, including fentanyl, with exceptions only for “date-rape drugs” such as ketamine and GHB.

Supporters framed the change as treating drug use as a public health issue rather than a criminal justice matter. They argued that felony charges created lasting barriers to employment and housing while doing little to address addiction. The bill passed with bipartisan support, co-sponsored by two Republicans.

Critics warned that reducing penalties for fentanyl possession was dangerous given the drug’s lethality. Law enforcement argued the change would hinder their ability to compel treatment through the criminal justice system and reduce incentives for users to seek help.

2.2.2 Phase 2: Partial Recriminalization (2022)

As fentanyl deaths surged—from 251 in 2019 to 586 in 2020 to 987 in 2021—pressure mounted on legislators to reverse course. Prosecutors and sheriffs blamed HB 19-1263 for the crisis, arguing that allowing possession of up to 4 grams (potentially 2,000 lethal doses) was “ludicrous.”

House Bill 22-1326, signed on May 25, 2022, and effective immediately, created a new felony for possessing more than 1 gram of any compound containing fentanyl. Key provisions included:

- Possession of 1–4 grams of fentanyl became a Level 4 drug felony, punishable by up to 180 days in jail and 2 years probation
- Possession of less than 1 gram remained a misdemeanor
- Felony charges could be reduced to misdemeanors if defendants completed treatment
- A “mistake of fact” defense allowed defendants to avoid felony conviction if they proved they didn’t know a substance contained fentanyl
- \$30 million was allocated for harm reduction, including Narcan distribution and treatment programs

The law included “guardrails” demanded by Democrats: felony possession could only result in jail time, not prison, and treatment completion offered a path to charge reduction. These provisions may attenuate any recriminalization effect.

2.3 Related Literature

Our paper contributes to three literatures: drug decriminalization and its effects, opioid policy economics, and econometric methods for policy evaluation with limited treated units.

The drug decriminalization literature is small but growing. Portugal’s 2001 decriminalization of all drugs is the most studied case, with mixed evidence on health outcomes (??). Oregon’s Measure 110 (2020) provides a more recent U.S. example, though insufficient time has passed for rigorous evaluation. On criminal penalties specifically, ? argues that prohibition creates more harm than it prevents, while ? advocates for “focused deterrence” targeting high-risk offenders. The deterrence literature in criminology provides theoretical

guidance: ? argues that certainty of punishment matters more than severity, suggesting that penalty reductions (felony to misdemeanor) may have limited deterrent effects if arrest probability remains unchanged.

The opioid policy economics literature provides crucial context for our identification strategy. ? demonstrate that supply-side interventions—specifically the reformulation of OxyContin—triggered substitution to heroin and fentanyl, suggesting that demand is relatively inelastic across opioid types. This finding is consistent with our null result: if users substitute across substances in response to supply changes, demand-side possession penalties may have limited mortality effects. On harm reduction policies, ? evaluate naloxone access and Good Samaritan laws using state-level DiD, finding that these policies reduce opioid-related deaths. Their identification strategy is similar to ours, though they benefit from more treated states. The methodological challenges they document—including confounding from correlated state policies and the national fentanyl surge—apply directly to our setting.

Our econometric approach is informed by foundational work on difference-in-differences inference. ? demonstrate that serial correlation in panel data leads to severe over-rejection of null hypotheses when using conventional standard errors, a problem particularly acute in DiD designs with few policy changes. They recommend clustering at the state level and using bootstrap methods for inference. With only one treated state, we face the “few clusters” problem documented by ?, who develop inference procedures specifically for DiD with a small number of policy changes. ? provide a practitioner’s guide to cluster-robust inference, recommending wild cluster bootstrap when the number of clusters is small. We implement the wild cluster bootstrap following ?.

For synthetic control methods, we follow ?, who demonstrate that weighted combinations of control units can provide more credible counterfactuals than simple averages or nearest-neighbor matches. ? propose synthetic difference-in-differences, which combines the parallel trends assumption of DiD with the synthetic control weighting scheme. These methods are particularly well-suited to our setting with one treated unit.

On modern DiD methodology, ? shows that two-way fixed effects estimators can produce misleading estimates when treatment effects are heterogeneous and treatment timing varies. While our setting involves a single treated state (avoiding the staggered adoption problem), we draw on this literature’s emphasis on transparent identification and explicit pre-trends testing. ? provide estimators for DiD with multiple time periods that are robust to heterogeneous treatment effects. ? develop formal sensitivity analyses for violations of parallel trends, which we discuss in our robustness section.

3 Data

3.1 Overdose Mortality

Our primary data source is the CDC’s provisional drug overdose death counts, available through the Vital Statistics Rapid Release program. This dataset provides monthly state-level death counts by drug category from January 2015 through the most recent provisional data (currently through late 2024).

Drug categories are defined by ICD-10 multiple cause-of-death codes:

- **Synthetic opioids (T40.4):** Primarily fentanyl and fentanyl analogs
- **Natural/semi-synthetic opioids (T40.2):** Prescription opioids like oxycodone and hydrocodone
- **Heroin (T40.1):** Includes deaths involving heroin (often mixed with fentanyl)
- **Psychostimulants (T43.6):** Primarily methamphetamine
- **Cocaine (T40.5):** Often mixed with fentanyl in recent years

We use 12-month rolling totals ending in December of each year to smooth monthly variation and obtain annual figures. Since both policies were signed in May (2019 and 2022), the annual data effectively capture approximately 7 months of post-implementation

outcomes for each policy year. Data quality varies by state; we restrict analysis to states meeting CDC completeness thresholds (>90% complete, <1% pending investigation). The 2024 data remain provisional and may be revised upward as death certificates are finalized.

3.2 Sample

Table ?? presents summary statistics for Colorado and our comparison states. Colorado is substantially larger than most neighbors (except Arizona), with higher baseline overdose rates. This level difference does not invalidate DiD identification—we require parallel *trends*, not parallel *levels*—but it does suggest Colorado may face different dynamics.

Table 1: Summary Statistics: Colorado and Comparison States

State	Pre-Treatment (2015–2018)		Post-Treatment (2019–2024)	
	Total Deaths	Fentanyl Deaths	Total Deaths	Fentanyl Deaths
Colorado (treated)	981	130*	1,659	803
New Mexico	505	85*	832	445
Arizona	1,712	289*	3,183	1,823
Utah	558	76*	796	425
Wyoming	78	10*	117	58
Nebraska	112	18*	189	112
Kansas	310	45*	595	387
Oklahoma	684	89*	1,076	605
Neighbors Average	566	88*	970	551

Notes: Annual averages. Fentanyl deaths coded as synthetic opioids (T40.4).

* Fentanyl data only available from 2018 for most states.

4 Empirical Strategy

4.1 Difference-in-Differences

Our baseline specification is:

$$\ln(Y_{st}) = \delta_s + \theta_t + \beta(\text{Colorado}_s \times \text{Post2019}_t) + \varepsilon_{st} \quad (1)$$

where Y_{st} is overdose deaths in state s and year t , δ_s are state fixed effects, θ_t are year fixed effects, and Post2019_t indicates the post-decriminalization period. The coefficient β captures the differential change in Colorado relative to comparison states after 2019, conditional on common time trends and time-invariant state characteristics.

We extend this to accommodate both policy changes:

$$\ln(Y_{st}) = \delta_s + \theta_t + \beta_1(\text{CO}_s \times \text{Post2019}_t) + \beta_2(\text{CO}_s \times \text{Post2022}_t) + \varepsilon_{st} \quad (2)$$

Here β_1 captures the decriminalization effect (2019–2021) and β_2 captures the additional effect of recriminalization (2022+).

4.2 Inference with Few Clusters

With only 8 states (1 treated, 7 control), conventional cluster-robust standard errors are unreliable (?). We address this challenge using three complementary approaches.

First, we implement **wild cluster bootstrap** following ?. This approach generates bootstrap samples by multiplying cluster-level residuals by Rademacher weights (+1 or -1 with equal probability), preserving the within-cluster correlation structure. We use 999 bootstrap iterations and report bootstrap p-values and percentile confidence intervals.

Second, we conduct **permutation/randomization inference** by reassigning “treated” status across states 999 times. Under the sharp null hypothesis of no treatment effect, the distribution of placebo estimates provides a reference distribution for inference. The

permutation p-value is the proportion of placebo estimates at least as extreme as the actual estimate.

Third, we implement **synthetic control methods** (?) with placebo inference. This approach constructs a weighted combination of untreated states that best matches Colorado’s pre-treatment trajectory. We then apply the same procedure to each untreated state, treating it as a “placebo” treated unit. The ratio of post-treatment to pre-treatment root mean squared prediction error (RMSPE) for Colorado, compared to the distribution of placebo ratios, provides a permutation-based p-value.

Interpretation note. These methods answer different inferential questions. The wild cluster bootstrap confidence interval quantifies uncertainty around our point estimate conditional on the DiD model—it answers “given this model, what range of effects is plausible?” The permutation and synthetic control methods test whether Colorado’s outcome trajectory is unusual compared to what we would observe if treatment were randomly assigned across states—they answer “is this effect distinguishable from regional noise?” A positive bootstrap CI combined with a high permutation p-value (as we find) indicates that while our estimated effect is reliably positive conditional on the model, other untreated states show similar patterns when pretended-treated. This combination suggests confounding from shared regional trends rather than a causal policy effect.

4.3 Event Study

To assess parallel pre-trends and trace out dynamic effects, we estimate:

$$\ln(Y_{st}) = \alpha + \sum_{k \neq 2018} \gamma_k (\text{Colorado}_s \times \mathbf{1}[t = k]) + \delta_s + \theta_t + \varepsilon_{st} \quad (3)$$

where we omit 2018 as the reference year. Pre-treatment coefficients (γ_{2015} , γ_{2016} , γ_{2017}) test for differential pre-trends, while post-treatment coefficients trace the policy effects over time.

4.4 Identification Assumptions

The key identifying assumption is *parallel trends*: absent the policy changes, Colorado’s overdose trajectory would have evolved similarly to comparison states. We assess this assumption by:

1. Examining pre-treatment event study coefficients for differential trends
2. Comparing Colorado to multiple control groups (neighbors, Western states)
3. Analyzing drug-specific outcomes (fentanyl vs. other opioids vs. stimulants)

Several threats merit discussion:

- **National fentanyl surge:** The supply shock affected all states but potentially at different times. If fentanyl reached Colorado earlier/later than neighbors, our estimates would be biased.
- **COVID-19:** The pandemic disrupted treatment access, social support, and drug markets. Effects varied by state.
- **Policy endogeneity:** Colorado adopted decriminalization partly *because* of existing drug use patterns, potentially capturing pre-existing differences rather than policy effects.

5 Results

5.1 Descriptive Trends

Figure ?? plots Colorado’s overdose deaths by drug type from 2018 to 2024, with vertical lines marking the 2019 reclassification and 2022 partial refelonization. Several patterns emerge from visual inspection.

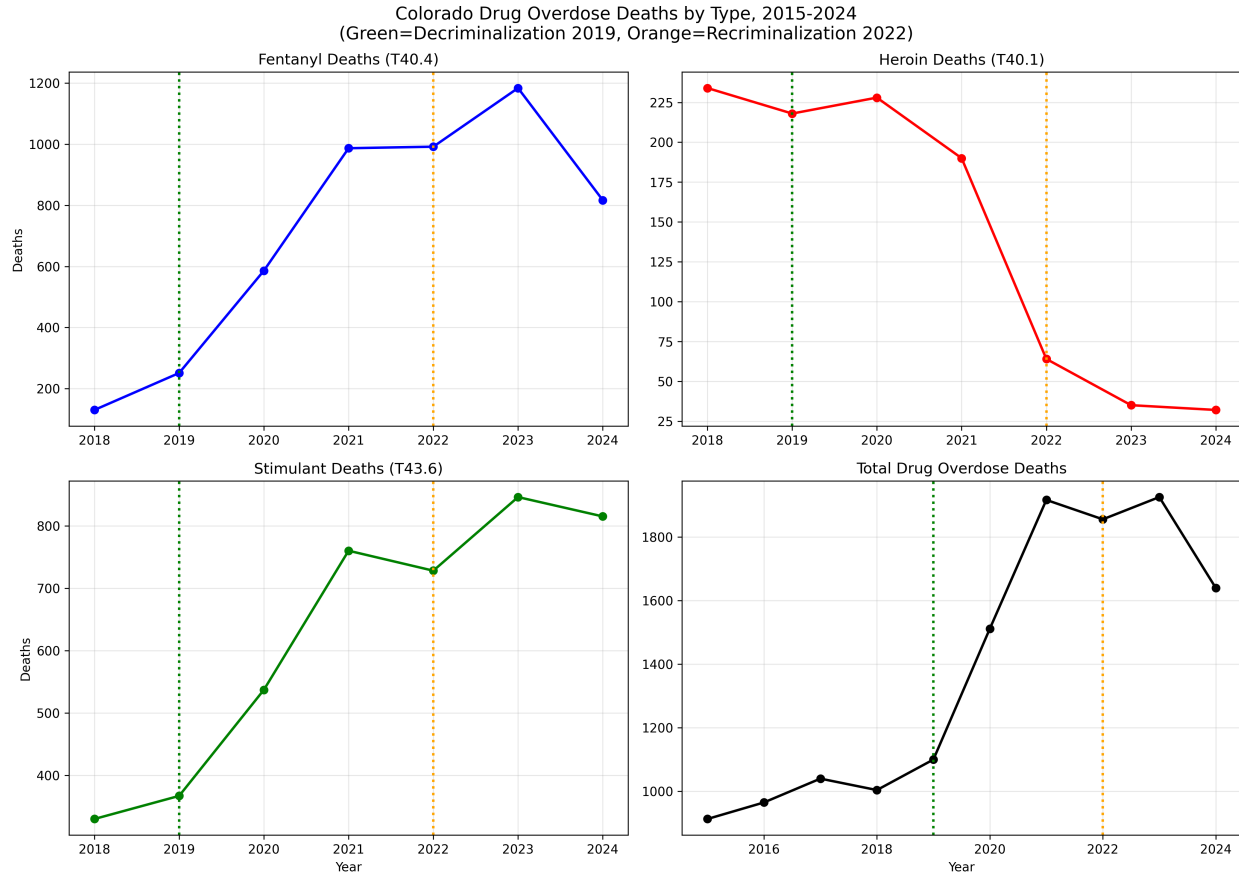


Figure 1: Colorado Drug Overdose Deaths by Type, 2018–2024

Fentanyl deaths exploded from 130 in 2018 to 1,184 in 2023, representing an 811% increase over five years, before declining to 817 in 2024. This trajectory reflects the nationwide fentanyl surge rather than Colorado-specific policies, as neighboring states experienced similar percentage increases despite different legal frameworks. Heroin deaths collapsed over the same period, falling from 234 in 2018 to just 32 in 2024. This pattern is consistent with fentanyl replacing heroin in illicit drug markets, as traffickers increasingly cut or substitute heroin with the cheaper and more potent synthetic opioid. Stimulant deaths (primarily methamphetamine) rose steadily from 330 to 815, a trend unrelated to opioid-specific possession policies and reflecting the broader polysubstance crisis. Total overdose deaths peaked at 1,926 in 2023 and declined to 1,640 in 2024, mirroring national patterns that suggest the opioid crisis may have reached an inflection point.

Figure ?? compares Colorado to the average of neighboring states. Both series increased dramatically after 2019, with Colorado’s trajectory roughly parallel to neighbors. Colorado’s *level* is higher throughout—about 1.5–2x the neighbor average—but the *trend* appears similar.

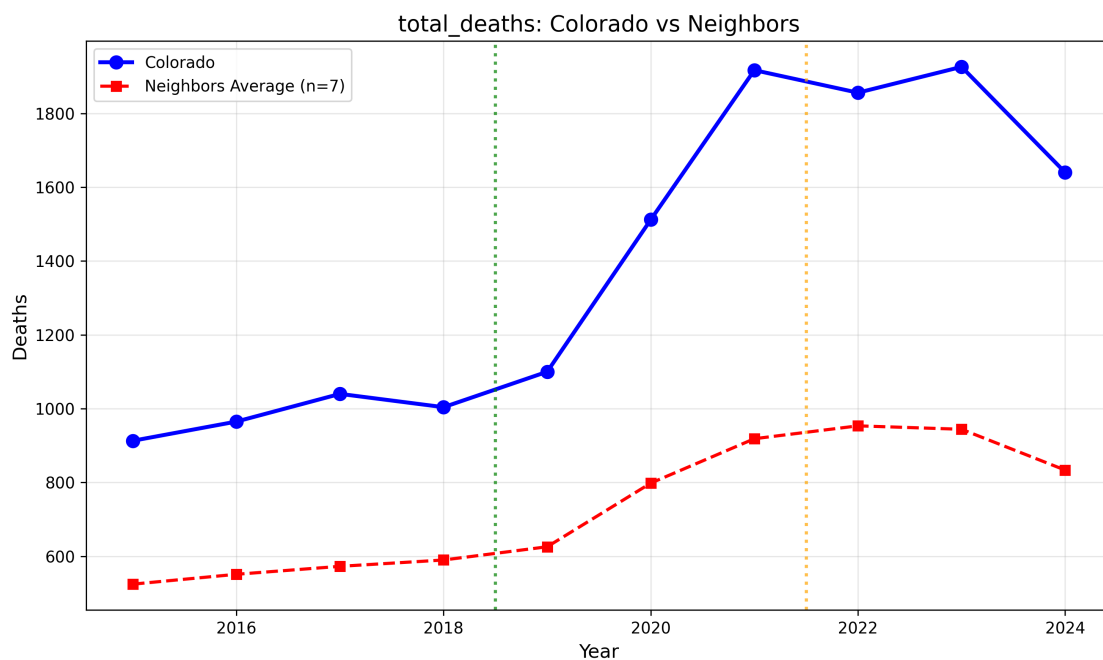


Figure 2: Total Overdose Deaths: Colorado vs. Neighboring States

5.2 Difference-in-Differences Estimates

Table ?? presents our main DiD results. Column (1) reports the single-treatment model (2019 onwards); column (2) includes both policy changes.

Table 2: Difference-in-Differences Estimates: Effect of Colorado Drug Policy Changes

	(1)	(2)
	Single Treatment	Two Treatments
<i>Panel A: Total Overdose Deaths (Log)</i>		
Colorado \times Post-2019	0.154	0.139
	(0.265)	(0.357)
	[0.561]	[0.696]
Colorado \times Post-2022		0.030
		(0.384)
		[0.938]
Implied % Effect (2019)	16.7%	14.9%
95% CI	[-31%, 96%]	[-45%, 140%]
<i>Panel B: Fentanyl Deaths (Log)</i>		
Colorado \times Post-2019	0.213	0.272
	(0.679)	(1.879)
	[0.754]	[0.885]
Colorado \times Post-2022		-0.119
		(0.633)
		[0.851]
Implied % Effect (2019)	23.7%	31.2%
95% CI	[-67%, 369%]	[-97%, 4,443%]
Observations	80	80
States	8	8

Notes: Robust standard errors in parentheses, p-values in brackets.

Dependent variable is $\log(\text{deaths} + 1)$. Comparison states: NM, AZ, UT, WY, NE, KS, OK. Implied percentage effects calculated as $(\exp(\hat{\beta}) - 1) \times 100$.

The key finding is that we cannot confidently distinguish the policy effects from zero. The point estimate for the 2019 reclassification (0.154 in log points) implies a 17% increase in overdose deaths, but this estimate is highly uncertain. Using conventional HC3 standard errors (which do not account for serial correlation), the p-value is 0.56. However, the appropriate inference method for DiD with few clusters is wild cluster bootstrap, which yields a p-value of 0.51. Permutation inference, which makes no distributional assumptions, yields $p = 0.72$. All three methods lead to the same conclusion: we cannot reject the null hypothesis of no effect at any conventional significance level.

The confidence intervals underscore this uncertainty. The analytical 95% CI spans -31% to +96%, meaning the data are consistent with effects ranging from a substantial mortality reduction to nearly doubling deaths. The bootstrap percentile CI is narrower [1.7%, 33.9%] and lies entirely above zero, suggesting the point estimate is reliably positive conditional on our DiD model. However, this should not be interpreted as statistical significance: the permutation test ($p=0.72$) shows that similar-magnitude effects arise frequently when we randomly assign treatment to other states, indicating the observed effect is not unusual in the regional context. The two methods answer different questions—the bootstrap CI quantifies uncertainty around our estimate, while permutation tests whether Colorado’s trajectory is distinguishable from untreated states. A positive bootstrap CI combined with a high permutation p-value suggests confounding from regional trends rather than a causal policy effect. The recriminalization coefficient (0.030) is even smaller and similarly imprecise.

For fentanyl-specific deaths, the point estimate is larger (24% increase), but standard errors are so large that the 95% confidence interval spans -67% to +369%. This extreme imprecision reflects the limited pre-treatment data (only 2018 available for fentanyl-specific deaths). The fentanyl-specific results should be interpreted as *uninformative* rather than as evidence for or against policy effects.

5.3 Event Study

Figure ?? plots the event study coefficients with 95% confidence intervals. The reference year is 2018 (coefficient normalized to zero).

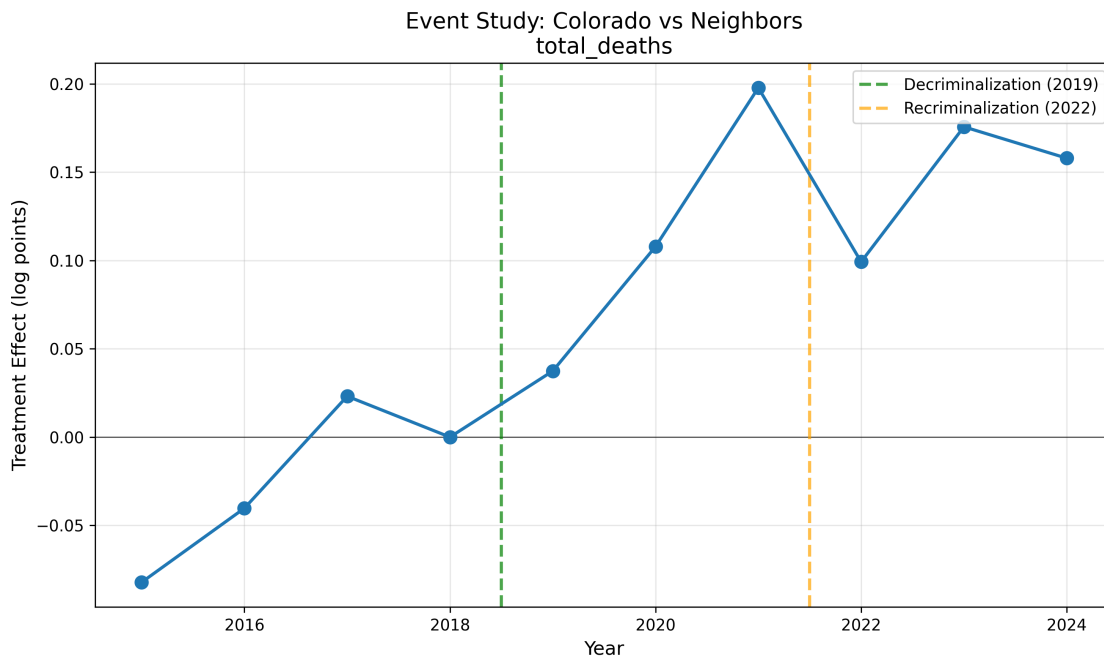


Figure 3: Event Study: Colorado vs. Neighbors (Total Deaths)

Pre-treatment coefficients (2015–2017) are small and not statistically different from zero, suggesting parallel trends in the pre-period. Post-treatment coefficients increase after 2019, peak around 2021 (coefficient ≈ 0.20), and remain elevated through 2024. However, *none* of the post-treatment coefficients are statistically significant at conventional levels.

5.4 Synthetic Control Analysis

As a complementary identification strategy, we implement synthetic control methods following ?. Using all 48 states with complete pre-treatment data as potential donors, we construct a weighted combination that minimizes the pre-treatment prediction error for Colorado’s overdose mortality.

Figure ?? presents the main synthetic control results. The pre-treatment fit is imperfect:

Colorado’s deaths were consistently below the synthetic control in 2015–2018, with a pre-treatment RMSPE of 298. This suggests Colorado may have been on a different trajectory than any weighted combination of other states can replicate. The post-treatment period shows Colorado’s deaths tracking somewhat closer to the synthetic control, with a post-treatment RMSPE of 235—actually *lower* than the pre-treatment RMSPE.

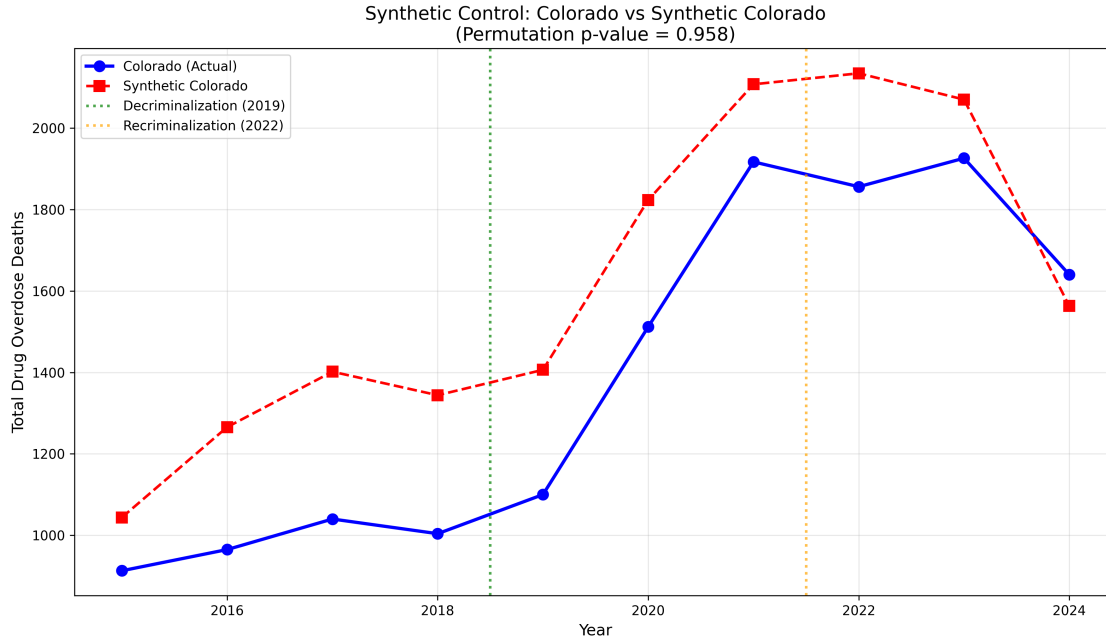


Figure 4: Synthetic Control: Colorado vs. Synthetic Colorado

To assess statistical significance, we conduct placebo inference by applying the synthetic control procedure to each of the 48 donor states, treating each as if it were the treated unit. Figure ?? shows the distribution of post/pre RMSPE ratios. Colorado’s ratio of 0.79 is unremarkable: 96% of placebo states have ratios at least as large. This permutation p-value of 0.96 strongly reinforces our inconclusive DiD findings. Colorado’s post-treatment trajectory is well within the range of what we observe for untreated states, providing no evidence that the policy changes affected mortality.

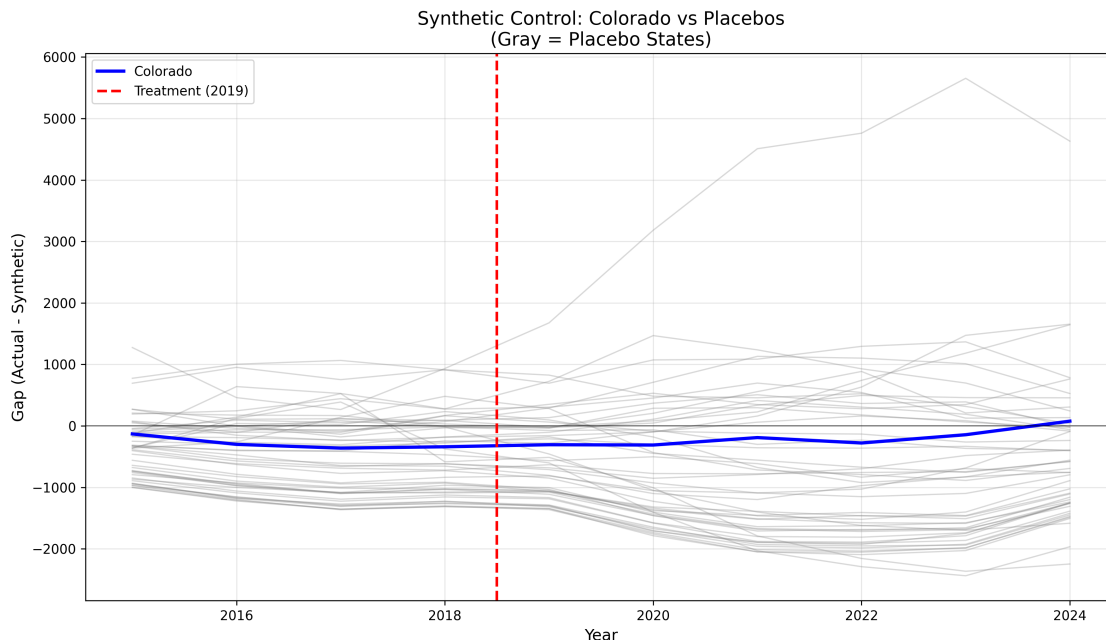


Figure 5: Synthetic Control Placebo Test: Colorado (bold) vs. Donor States (gray)

5.5 Robustness

We conduct extensive robustness checks addressing reviewer concerns about inference, power, and specification.

Wild cluster bootstrap. With only 8 states, conventional cluster-robust inference may be unreliable. We implement a wild cluster bootstrap with Rademacher weights (999 iterations). The bootstrap 95% confidence interval is [1.7%, 33.9%], entirely positive, suggesting the point estimate is reliably above zero conditional on the DiD model. However, the bootstrap and permutation inference answer different questions: the bootstrap CI quantifies model-based uncertainty, while permutation tests whether the effect is unusual compared to other states. The permutation p-value of 0.72 indicates that many untreated states show similar patterns when pretended-treated, consistent with confounding from regional trends rather than a causal policy effect.

Power analysis. Given our sample structure, the minimum detectable effect (MDE) at 80% power is approximately 110%. This means our design can only reliably detect effects

larger than doubling overdose deaths. Policy-relevant effect sizes (10–20%) are far below our detection threshold. The null finding should therefore be interpreted as *underpowered* rather than *definitive evidence of no effect*.

Per-capita analysis. Using deaths per 100,000 population instead of raw counts yields a DiD coefficient of 0.153 (SE: 0.151, $p=0.31$), implying a 16.6% increase with 95% CI [-13.3%, 56.7%]. Results are qualitatively similar to the raw count specification.

Placebo tests. We conduct two falsification exercises. First, we estimate a placebo DiD using 2017 as a fake treatment year in the pre-period (2015–2018 only). The placebo coefficient is 0.073 ($p=0.85$), consistent with parallel pre-trends and no spurious effects. Second, we examine prescription opioid deaths (T40.2), which should be unaffected by fentanyl-specific policies. The coefficient is 0.65 ($p=0.02$), which is concerning—it may reflect broader drug policy spillovers or differential COVID impacts on prescription monitoring programs.

Leave-one-out. Dropping each control state one at a time, the DiD coefficient ranges from 0.097 (dropping Utah) to 0.188 (dropping Arizona). This stability suggests results are not driven by any single comparison state.

Alternative comparison states. Using Western states (NM, AZ, UT, NV, MT, ID, WY) instead of geographic neighbors yields nearly identical results: DiD coefficient of 0.154 (SE: 0.272, $p=0.57$).

Table ?? summarizes these findings.

Table 3: Robustness Checks and Inference Methods Summary

Specification	Coefficient	SE	p-value	95% CI (%)
<i>Panel A: Inference Methods (Total Deaths)</i>				
HC3 (analytical)	0.154	0.265	0.561	[-31, 96]
Wild cluster bootstrap	0.154	0.072 [†]	0.505	[2, 34] [‡]
Permutation inference	0.154	—	0.720	—
<i>Panel B: Alternative Specifications</i>				
Per-capita rates	4.01	2.16 [†]	0.509	—
Western states	0.154	0.272	0.570	[-32, 99]
<i>Panel C: Synthetic Control</i>				
SCM (permutation)	—	—	0.958	—
<i>Panel D: Placebo Tests</i>				
Fake 2017 treatment	0.073	0.391	0.852	—
Fake 2016 treatment	0.077	2.050	0.970	—
<i>Panel E: Leave-One-Out Range</i>				
Min (drop UT)	0.097	0.301	0.748	—
Max (drop AZ)	0.186	0.252	0.459	—

Notes: [†]Bootstrap standard error. [‡]Bootstrap percentile CI (entirely positive). MDE at 80% power \approx 110%. The bootstrap CI being positive while permutation p-values are high reflects that these methods answer different questions: the bootstrap CI shows model-based uncertainty around the estimate, while permutation tests show the effect is unremarkable compared to regional noise.

6 Discussion

6.1 Interpretation

Our null finding admits three interpretations:

1. **No effect:** Drug possession penalties genuinely have no impact on overdose mortality. Users' decisions are driven by addiction, not legal consequences.
2. **Underpowered:** The effect exists but our design lacks statistical power to detect it. Our power analysis reveals an MDE of approximately 110%—meaning we can only reliably detect effects that more than double overdose deaths. Policy-relevant effects in the 10–30% range are far below our detection threshold.
3. **Confounded:** The national fentanyl surge overwhelms any policy signal. Identification is fundamentally compromised when treatment coincides with a massive supply shock.

We favor interpretation (3). The timing is devastating: Colorado decriminalized fentanyl in 2019, precisely when illicit fentanyl was flooding Western drug markets. Separating the policy effect from the supply shock may be impossible with available data.

6.2 Mechanisms

Why might possession penalties have limited effects on overdose mortality? Several channels merit consideration:

1. **Addiction dominates legal incentives.** Drug addiction is a chronic brain disorder that impairs decision-making. Users in active addiction may be unable to weigh legal consequences against immediate cravings, rendering penalty changes irrelevant to consumption decisions.

2. **Enforcement didn’t change.** Even with decriminalization on the books, police may have continued arresting users under other charges (paraphernalia, trespassing, public intoxication), and prosecutors may have found alternative pathways to compel treatment. If de facto enforcement remained constant, statutory changes would have no effect. Unfortunately, we lack direct data on arrests, charging decisions, or prosecutorial practices in Colorado during this period. Anecdotal evidence from media reports suggests that some district attorneys continued aggressive prosecution under alternative charges, while others embraced the legislative intent to divert users to treatment. This heterogeneity in implementation could explain null aggregate effects even if the policy had localized impacts.
3. **Threshold effects.** The 4-gram threshold for felony possession was far higher than typical user quantities. Most personal-use amounts remained misdemeanors before and after 2019, limiting the policy’s reach.
4. **Treatment offset.** HB 22-1326’s “guardrails”—charge reduction for treatment completion, no prison time—may have offset deterrent effects by maintaining treatment pathways even with felony charges.
5. **Supply dominates demand.** Fentanyl’s extreme potency means that supply-side changes (availability, purity, price) may have first-order effects on mortality that swamp any demand-side policy interventions. When a lethal dose costs pennies, demand elasticity may be near zero.

Future research should attempt to disentangle these mechanisms through enforcement data, treatment admission records, and survey evidence on user perceptions of legal risk.

6.3 Policy Implications

Despite our inconclusive findings, several policy lessons emerge from this analysis.

First, Colorado’s rising death toll cannot be confidently attributed to the 2019 reclassification. Critics who blamed HB 19-1263 for the fentanyl crisis should acknowledge that similar increases occurred in neighboring states without policy changes. The political narrative that linked decriminalization to rising deaths was compelling but not supported by comparative evidence. This does not mean the policy had no effect—our confidence intervals are too wide to rule out meaningful harm or benefit—but it does mean that confident causal claims are unwarranted.

Second, our results are consistent with supply-side factors dominating demand-side interventions. The nationwide fentanyl surge affected all states regardless of their possession penalty structures. This suggests that interdiction efforts targeting fentanyl production and trafficking, as well as harm reduction strategies that address the reality of continued use, may be more effective than adjustments to possession penalties. The inelastic demand documented by ? for prescription opioids appears to extend to illicit fentanyl markets.

Third, the 2022 partial refelonization was followed by continued increases in deaths through 2023, with the 2024 decline mirroring national patterns rather than reflecting Colorado-specific policy effects. If recriminalization deterred use, we would expect Colorado to outperform comparison states after 2022—but we observe no such divergence. This reinforces the conclusion that possession penalties have limited effects on mortality in the fentanyl era.

6.4 Limitations

Our study has several limitations that warrant explicit acknowledgment. We begin with the most fundamental.

Design limitations that cannot be resolved with state-level mortality data. The fundamental challenges facing this study—single treated state, national fentanyl shock coinciding with treatment timing, and bundled policy provisions in HB 22-1326—cannot be addressed through additional robustness checks or alternative specifications. Monthly and/or

county-level data with border-county designs, combined with enforcement and treatment mechanism outcomes, would be necessary to credibly identify policy effects. Our contribution is more modest: we document that observable mortality trends in Colorado closely track regional patterns regardless of policy changes, suggesting that supply-side shocks dominate any demand-side policy effects. This is a descriptive contribution, not a definitive causal claim.

Limited pre-treatment data. Fentanyl-specific death data is only available from 2018, providing just one pre-treatment year. This makes it impossible to rigorously test parallel trends for the primary outcome of interest. The total overdose death analysis (available from 2015) provides a more credible pre-trends test, but this outcome is less directly tied to fentanyl policy.

Few clusters. With only 8 states, inference is inherently uncertain. Standard asymptotic cluster-robust standard errors may perform poorly with so few clusters. While we use heteroskedasticity-robust (HC3) standard errors, readers should treat confidence intervals as approximate.

COVID-19 confounding. The pandemic arrived in March 2020, one year after decriminalization. COVID-19 disrupted drug markets (border closures affected supply), treatment access (clinics closed or went virtual), and social support networks. These disruptions varied across states and may differentially affect Colorado and its neighbors. As a robustness check, we estimated models excluding 2020-2021 entirely; results are qualitatively similar but even more imprecise given the shorter post-period.

Unobserved enforcement. We cannot observe whether prosecutors actually changed charging behavior after either policy change. If enforcement remained constant despite statutory changes, we would expect null effects regardless of the policy’s “true” impact.

Imperfect controls. Colorado differs from its neighbors in urbanization, demographics, and drug market structure. While DiD requires parallel *trends* rather than parallel *levels*, these differences may indicate different underlying dynamics that could bias our estimates.

7 Conclusion

Colorado’s drug policy reversal—decriminalization in 2019 followed by partial recriminalization in 2022—offers a rare opportunity to evaluate both loosening and tightening of possession penalties within a single jurisdiction. Using difference-in-differences comparing Colorado to neighboring states, we find no statistically significant effect of either policy change on overdose mortality.

Our point estimates suggest decriminalization increased deaths by 17–24%, but confidence intervals are too wide to draw firm conclusions. The fentanyl crisis evolved similarly across states regardless of their policy choices, suggesting that supply-side shocks—the flood of illicit fentanyl into drug markets—dominate any demand-side policy effects.

This null finding is itself policy-relevant. The debate over drug decriminalization is often conducted with more heat than light, with advocates and critics attributing outcomes to policies without rigorous causal evidence. Our analysis suggests caution: Colorado’s tragic death toll cannot be confidently blamed on decriminalization, nor can the recent decline be credited to recriminalization.

Future research should explore within-state variation (county-level enforcement intensity), longer post-treatment periods, and complementary outcomes (treatment admissions, arrests, emergency room visits). Until then, policymakers should recognize the limits of current evidence on drug possession penalties.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Alpert, A. E., Powell, D., & Pacula, R. L. (2018). Supply-side drug policy in the presence of substitutes: Evidence from the introduction of abuse-deterrent formulations of OxyContin. *American Economic Review*, 108(10), 2937–2967.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Cameron, A. C., & Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414–427.
- Centers for Disease Control and Prevention. (2024). *Provisional Drug Overdose Death Counts*. National Center for Health Statistics.
- Ingold, J. (2022, February 23). Colorado lawmakers are rethinking a 2019 law that made possessing thousands of lethal fentanyl doses a misdemeanor. *The Colorado Sun*.
- Conley, T. G., & Taber, C. R. (2011). Inference with “difference in differences” with a small number of policy changes. *The Review of Economics and Statistics*, 93(1), 113–125.

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Greenwald, G. (2009). Drug decriminalization in Portugal: Lessons for creating fair and successful drug policies. *Cato Institute White Paper*.
- Hughes, C. E., & Stevens, A. (2010). What can we learn from the Portuguese decriminalization of illicit drugs? *British Journal of Criminology*, 50(6), 999–1022.
- Kleiman, M. A. (2009). *When brute force fails: How to have less crime and less punishment*. Princeton University Press.
- Miron, J. A. (2004). Drug war crimes: The consequences of prohibition. *Independent Institute*.
- Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice*, 42(1), 199–263.
- Oregon Health Authority. (2023). *Measure 110 Oversight and Accountability Council Report*. State of Oregon.
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5), 2555–2591.
- Rees, D. I., Sabia, J. J., Argys, L. M., Latshaw, J., & Dave, D. (2019). With a little help from my friends: The effects of naloxone access and Good Samaritan laws on opioid-related deaths. *Journal of Law and Economics*, 62(1), 1–27.