Bans Without Bite: State Spread-Pricing Laws and Medicaid Generic Drug Spending

Abstract

Background. Between 2018 and 2024, twenty-four U.S. states enacted or contract-mandated prohibitions on pharmacy benefit manager (PBM) spread pricing in Medicaid managed care. Ohio's 2018 audit of PBM spread pricing, state-level audits from Kentucky, Michigan, and Massachusetts, and the U.S. Federal Trade Commission's 2024–2025 interim reports on the "Big 3" PBMs have produced a widely held expectation that prohibiting spread pricing should reduce Medicaid managed care (MCO) generic drug spending. This paper tests whether the 2018–2024 wave of state spread-pricing bans produced detectable changes in MCO generic drug spending per enrollee.

Methods. I assemble a balanced state-quarter panel of Medicaid pharmacy spending (2016 Q1 – 2024 Q4, 51 jurisdictions, 1,836 observations) using the CMS State Drug Utilization Data (SDUD) for spending and utilization, National Average Drug Acquisition Cost (NADAC) for benchmark pricing, and CMS Medicaid enrollment counts for MCO and fee-for-service (FFS) denominators. The primary outcome is log gross generic drug spending per MCO enrollee per quarter. I use Callaway-Sant'Anna (2021) staggered difference-in-differences as the primary estimator and a two-way fixed-effects (TWFE) benchmark. I include a pre-registered triple-difference (generic vs brand within MCO) and a **negative control** (MCO vs FFS within state) to isolate the MCO PBM spread-pricing mechanism from broader state pharmacy cost-containment activity. I report Rambachan-Roth (2023)-style relative-magnitudes pre-trend sensitivity bounds.

Results. The headline difference-in-differences estimate on log MCO generic spending per enrollee is -0.058 log points (SE 0.231) in TWFE and -0.054 (SE 0.477) in Callaway-Sant'Anna — directionally consistent with the expected spread-pricing-reduction story but statistically indistinguishable from zero. The **triple-difference (generic vs brand within MCO)** sharpens precision (SE 0.120) and produces a point estimate of -0.117 log points — borderline, not statistically significant, with the brand placebo passing cleanly (+0.059, SE 0.290). The **negative control fails**: fee-for-service generic spending per FFS enrollee declines by -0.252 log points (SE 0.205) in treated states, *larger* in magnitude than the MCO estimate. A battery of robustness checks (small-FFS-state exclusions, outlier trimming, common denominators, channel-total specifications) confirms that the FFS decline is substantive rather than mechanical. To test the cost-containment bundle reframe directly, I construct a state-year panel of five bundled pharmacy cost-containment actions (MAC list updates, PDL refreshes, NADAC alignment, pharmacy carve-outs, single-PBM transitions) and regress

each on the spread-ban indicator with state and year fixed effects. Treated states take approximately half an additional bundled action per state-year (composite coefficient +0.507, SE 0.107, $p < 0.001$), driven by MAC list updates (+0.199, $p < 0.001$), NADAC alignment (+0.176, $p < 0.001$), and PDL refreshes (+0.089, $p = 0.026$). Case-study event studies of Ohio's October 2022 single-PBM transition and California's January 2022 Medi-Cal Rx pharmacy carve-out show the expected mechanical channel flips in Medi-Cal Rx (FFS +1.88 log points, MCO -2.76 log points, both $p < 0.001$) but no detectable additional MCO savings from Ohio SPBM beyond the 2019 spread-pricing ban, and a precisely estimated +0.19 log point increase in total Medi-Cal generic spending per enrollee after the carve-out.

Conclusions. The clean identification story — that state spread-pricing bans cause a MCO-specific decline in generic drug spending through the PBM spread channel — is not supported by the data. The joint MCO and FFS generic spending decline in treated states is empirically consistent with states adopting spread-pricing bans as one component of a broader cost-containment bundle, and the bundled-action panel provides direct evidence that treated states disproportionately pursue the other bundled levers. I commit the paper to the bundled-cost-containment reframe and discuss implications for state policy evaluation, federal PBM reform under the Consolidated Appropriations Act of 2026, and the need for claim-level T-MSIS data to cleanly identify the spread-pricing mechanism in isolation from contemporaneous bundled activity.

---

1. Introduction

Between 2018 and 2024, twenty-four U.S. states acted to prohibit pharmacy benefit manager spread pricing in Medicaid managed care. The immediate catalyst was Ohio Auditor of State Dave Yost's August 2018 report, which analyzed 39 million Medicaid drug transactions and found that PBMs had charged the state a 31.4 percent spread on generic drugs — $208 million on $662.7 million in generic spending in a single fiscal year. Similar state-level audits followed in Kentucky (documented spread of $123.5 million retained by PBMs), Michigan (3 Axis Advisors estimated at least $64 million in overcharges on nearly two million generic prescriptions), and Massachusetts (the state Health Policy Commission found that managed care PBM prices exceeded acquisition cost for 95 percent of unique generic drugs analyzed). These audits produced a policy consensus: spread pricing in Medicaid managed care was inflating generic drug spending, and banning it should produce measurable savings. By the end of 2024, at least twenty-five states prohibited spread pricing in managed care contracts, more than double the count in 2019.

At the federal level, the U.S. Federal Trade Commission launched a Section 6(b) investigation of the six largest PBMs in 2022. Its July 2024 interim staff report documented extensive anticompetitive practices at the three largest PBMs (CVS

Caremark, Express Scripts, OptumRx), which together process approximately eighty percent of U.S. prescription claims. The January 2025 interim report on specialty generics found that the "Big 3" PBMs marked up specialty generics by hundreds to thousands of percent at affiliated pharmacies, generating over $7.3 billion in revenue above estimated acquisition costs between 2017 and 2022 and an additional $1.4 billion in spread pricing alone. The Consolidated Appropriations Act of 2026, signed February 3, 2026, is the first major federal PBM reform law; it prohibits PBM compensation from spread pricing in Medicare Part D effective January 2028 and directs the Government Accountability Office to study PBM compensation in both Medicare and Medicaid.

The timing is propitious for empirical evaluation. Four years have now passed since Ohio became the first state to act (January 2019), giving enough post-treatment quarters to observe dynamic effects. State audits from Ohio, Kentucky, Michigan, and Massachusetts provide strong priors on the *mechanism* (PBMs retain spread on the generic channel) but use internal data from individual state contracts and do not speak to whether prohibition produces measurable spending changes in aggregate state pharmacy accounts. No peer-reviewed causal evaluation has tested whether the 2018–2024 wave of state bans reduced Medicaid managed care generic drug spending at the state-quarter level using the publicly available SDUD / NADAC data infrastructure. This paper provides the first such evaluation.

The headline result is a precise null. Using two-way fixed effects on a balanced state-quarter panel of 51 U.S. jurisdictions over 2016–2024, I estimate that state spread-pricing bans reduce log MCO generic spending per enrollee by 0.058 log points (standard error 0.231). The 95 percent confidence interval is approximately -0.51 to +0.40, which rules out effects larger than roughly 40 percent reduction but does not rule out zero. The Callaway-Sant'Anna estimator produces a similarly signed but noisier estimate of -0.054 (SE 0.477). A triple-difference specification exploiting the within-MCO generic-versus-brand contrast tightens precision substantially (SE falls to 0.120) and produces a point estimate of -0.117 log points, but this remains statistically insignificant at conventional thresholds.

The pre-registered negative control — fee-for-service generic spending per FFS enrollee, which should be unaffected by a ban targeting the MCO channel — **fails** unambiguously. FFS generic spending in treated states falls by -0.252 log points (SE 0.205) after ban enactment, a larger magnitude decline than the MCO estimate. A battery of robustness checks — dropping small-FFS states where enrollment denominators are fragile, winsorizing per-enrollee outliers, using a common total-Medicaid denominator for both channels, and running a channel-total specification with no denominator at all — fails to eliminate the FFS effect. In every robustness specification, MCO generic spending is a precise null between +0.04 and -0.06 log points while FFS generic spending declines by approximately 25 to 33 log points. This pattern is inconsistent with a pure spread-pricing mechanism operating through PBM contracts in managed care: if the ban were the cause and the mechanism were spread pricing, the FFS

channel should have shown zero response.

I therefore reframe the paper. The spread-pricing *mechanism* cannot be cleanly identified from the SDUD/NADAC panel infrastructure because it is confounded with broader state pharmacy cost-containment activity. States that enacted spread-pricing bans also, during the same period, were more likely to tighten preferred drug lists, update maximum allowable cost (MAC) schedules in both MCO and FFS pharmacy benefits, align MCO reimbursement to NADAC-based benchmarks, and pursue preferred-generic utilization mandates. These activities affect both MCO and FFS generic spending. The joint MCO-plus-FFS decline this paper documents is therefore more consistent with the spread-pricing ban being *one component* of a broader cost-containment bundle than with spread pricing being the sole or dominant causal channel.

This reframe has three implications. First, the paper's contribution shifts from "spread-pricing bans cause MCO generic spending reductions" (a clean causal claim) to "the 2018–2024 wave of spread-pricing bans was associated with directionally negative generic spending changes in both MCO and FFS channels, consistent with a broader cost-containment policy bundle rather than with the spread-pricing mechanism in isolation." Second, the empirical evidence cannot arbitrate between the policy claim (bans work) and the mechanism claim (bans work *specifically through the spread-pricing channel*). The policy claim is weakly supported by directional evidence; the mechanism claim cannot be identified without claim-level PBM payment data from the Transformed Medicaid Statistical Information System (T-MSIS) Analytic File, which requires a Data Use Agreement that is pending. Third, the paper's discussion turns to what state policymakers should take away: directional evidence of savings under the bundled-policy reading, and an honest acknowledgment that the clean mechanism story cannot be confirmed with currently available data.

The remainder of the paper proceeds as follows. Section 2 provides institutional background on Medicaid pharmacy benefit structure, PBM spread pricing, the 2018–2024 state ban wave, and the federal PBM regulatory context. Section 3 reviews the related literature on PBM compensation, spread pricing audits, Medicaid pharmacy policy evaluation, and negative-control identification strategies in staggered difference-in-differences. Section 4 describes the data: the SDUD panel, NADAC benchmarks, MCO/FFS enrollment denominators, and the state spread-pricing ban adoption panel. Section 5 lays out the empirical strategy, including the primary TWFE specification, the Callaway-Sant'Anna estimator, the triple-difference, the negative control, and the Rambachan-Roth pre-trend sensitivity analysis. Section 6 presents the results, beginning with the main DiD and moving through the triple-difference, the negative control, the FFS measurement drill that established the negative-control failure as substantive rather than mechanical, and the pre-trend bounds. Section 7 discusses the reframe, the cost-containment bundle interpretation, and the implications for policy and future research. Section 8 concludes.

2. Background

2.1 Medicaid pharmacy benefit structure

Medicaid is the largest pharmacy benefit payer in the United States, with gross drug spending of approximately $80 billion in 2023. The program covers pharmacy benefits through two delivery systems: fee-for-service (FFS) pharmacy, in which the state Medicaid agency reimburses pharmacies directly for dispensed drugs, and managed care organization (MCO) pharmacy, in which the state contracts with private managed care plans that then subcontract with pharmacy benefit managers to administer the drug benefit. Approximately sixty-four percent of Medicaid drug spending in fiscal year 2023 flowed through the MCO channel. This share has grown steadily as states have expanded managed care enrollment and carved pharmacy benefits into MCO contracts.

The two channels differ in reimbursement architecture. FFS pharmacy is required by federal rule to reimburse pharmacies at "actual acquisition cost" (AAC) plus a dispensing fee. In practice, most states use the National Average Drug Acquisition Cost (NADAC) as the AAC benchmark, supplemented by state-level MAC lists for generic drugs. Because the state pays the pharmacy directly, FFS pharmacy is structurally a pass-through payment: there is no intermediary retaining spread. The only margins in FFS pharmacy are the pharmacy's own markup (bounded by the NADAC benchmark) and the dispensing fee (set administratively by the state).

MCO pharmacy, in contrast, typically involves a PBM intermediary. The MCO contracts with a PBM to administer the drug benefit; the PBM contracts with pharmacies for dispensing; and the PBM charges the MCO an amount that may differ from what it pays the pharmacy. The difference — the "spread" — is the PBM's compensation. Spread pricing is one of several PBM compensation models; the alternatives are transparent pass-through (in which the PBM is paid a flat administrative fee) and hybrid arrangements. Spread pricing creates an incentive problem: the PBM's revenue is linked to the per-claim margin, creating incentives to maximize the wedge between MCO payment and pharmacy reimbursement.

2.2 The generic drug concentration of PBM spread

Spread pricing is heavily concentrated in generic drugs. The Ohio audit documented a 31.4 percent spread on generic drugs compared to 0.8 percent on brand drugs — approximately a fortyfold difference. Several features of the generic drug market enable large spreads: manufacturer prices for generics have fallen dramatically over the past two decades, pharmacy acquisition costs (NADAC) have followed, but plan-level prices do not adjust commensurately because MCO contracts typically use MAC lists that are updated less frequently than NADAC. The PBM sits in the gap between the static MAC list and the falling acquisition cost, and the gap — the spread — is the PBM's retained margin.

Mattingly, Ben-Umeh, Bai, and Anderson (2023) conducted the most granular publicly available analysis of spread pricing to date, using Medicare Part D data for forty-five high-utilization generic drugs. They found that of $22.50 mean spending per claim, PBMs retained $9.18 (40.8 percent) in gross profit, compared to 17.2 percent for pharmacies, 12.0 percent for wholesalers, and 29.9 percent for manufacturers. The study's authors caution that simply banning PBM spread pricing may not reduce total spending because PBMs could compensate by raising administrative fees. This caveat motivates the empirical evaluation question at the heart of this paper: do actual state-level bans, in actual post-ban years, produce detectable reductions in Medicaid generic spending?

2.3 The 2018–2024 state ban wave

Ohio was the first state to act decisively. In August 2018, Ohio Auditor of State Dave Yost released a report finding that PBMs had charged the state a 31.4 percent spread on generic drugs. Ohio Medicaid Director Barbara Sears responded by directing all Ohio Medicaid MCOs to terminate existing PBM contracts and move to a transparent pass-through model effective January 1, 2019. Ohio subsequently went further, establishing a single statewide PBM (SPBM) in October 2022. A Milliman actuarial evaluation estimated the Ohio SPBM model saved the state $140 million over its first two years while increasing average pharmacy dispensing fees from $0.73 to approximately $9.00 per prescription.

New York followed in October 2019 with a similar contractual mandate. Arkansas enacted a statute in July 2020. Kentucky, Pennsylvania, Virginia, and Maryland acted in 2021, primarily through legislation or state plan amendments. North Carolina and Massachusetts followed in 2021 and 2023. Florida, Idaho, and Vermont acted in 2024. Ten states — Georgia, Iowa, Kansas, Louisiana, Michigan, Minnesota, Mississippi, North Dakota, New Jersey, and Texas — had contract-based spread-pricing prohibitions in place prior to 2019, often as part of standard MCO contract boilerplate that was in force before Ohio's high-profile audit. These ten "early adopter" states form the 2018 Q1 "mass cohort" in my analysis and are the subject of a key sensitivity analysis: their treatment timing is fuzzier than the post-Ohio cohorts because their bans were embedded in contract renegotiation cycles rather than in single-date administrative or legislative actions.

In parallel with the spread-pricing ban wave, several states pursued a more radical structural alternative: the pharmacy benefit carve-out. As of July 2023, eight states had carved pharmacy out of MCO contracts (up from four in 2019), moving the pharmacy benefit back into fee-for-service pharmacy where spread pricing is structurally impossible. California's January 2022 pharmacy carve-out ("Medi-Cal Rx") is the largest such event. New York and Ohio also announced carve-out plans during the study period. The carve-out states are a different treatment modality than spread-pricing bans and are handled as a separate sensitivity analysis in this paper.

2.4 The federal regulatory context

Federal PBM policy accelerated during the study period. CMS issued guidance in May 2019 encouraging states to prohibit spread pricing in managed care. The November 2023 Medicaid Managed Care final rule required PBM subcontractors to separately report the spread amount starting in the first rating period beginning on or after November 19, 2025. The FTC Section 6(b) investigation (2022 forward) produced two interim reports that sharpened the case against spread pricing. The Consolidated Appropriations Act of 2026 (CAA 2026), signed February 3, 2026, banned PBM compensation from spread pricing in Medicare Part D effective January 1, 2028 and directed GAO studies of both Medicare and Medicaid PBM compensation. These federal actions occurred during and immediately after the state ban wave this paper evaluates and partially shape the interpretation: state bans were not the only policy changes affecting PBM behavior during 2018–2024, and the policy environment for generic pharmacy pricing was broadly tightening across all channels.

2.5 Why the negative control matters

The empirical question this paper asks — whether state spread-pricing bans reduced MCO generic spending — depends on being able to distinguish the ban-specific effect from other contemporaneous state pharmacy cost-containment activity. The natural identification strategy is to exploit the fact that FFS pharmacy is structurally immune to spread pricing: FFS reimburses pharmacies directly, with no PBM intermediary, so a ban on PBM spread pricing cannot have any effect on FFS spending through that mechanism. If I estimate the ban's effect on FFS generic spending and find zero, that supports the clean identification of the MCO effect through the spread-pricing channel. If I estimate it and find a *non-zero* effect, then something else is happening — either a parallel-trends violation, a different mechanism, or confounding with other state pharmacy policy. This is the logic of a pre-registered negative control, and it is a major part of why this paper was designed with both MCO and FFS outcomes in the panel.

As the results show, the negative control fails. The interpretation of that failure — whether it is a mechanical data artifact or a substantive identification problem — is the core content of Section 6.

3. Related Literature

3.1 Spread pricing audits and mechanism evidence

The state-level audit literature provides the strongest prior for the *existence* of spread pricing as a margin-retention practice. The Ohio Auditor of State (2018) report is the canonical reference: a 31.4 percent spread on generics, $208 million in a single fiscal year, with an extreme case of $7,201 charged versus $3,859 reimbursed on a single generic leukemia drug claim. Similar findings came from

the Kentucky audit documenting $123.5 million retained by PBMs, the 3 Axis Advisors analyses of Michigan and New York, and the Massachusetts Health Policy Commission's (2019) finding that MCO/PBM prices exceeded drug acquisition costs for 95 percent of unique generic drugs analyzed. Werble (2017) provided an early descriptive analysis of spread pricing in Medicaid managed care; Mattingly, Hyman, and Bai (2023a) reviewed the literature through early 2023.

The conclusion of this strand of the literature is unambiguous: spread pricing exists, it is concentrated in generics, and its magnitudes in the states that have audited it are substantial relative to state pharmacy budgets. This literature motivates the present study but cannot answer the question the study asks. Audits document the *existence* of spread pricing in specific states at specific times; they do not document that *banning* spread pricing produces measurable spending reductions at the aggregate state-quarter level that a policy evaluator would be able to detect. The audit literature is the prior; the present paper is the evaluation.

3.2 Medicaid pharmacy policy evaluation

The peer-reviewed empirical literature on Medicaid pharmacy policy is dominated by studies of Medicaid expansion effects, prior authorization policies, preferred drug lists, and formulary design. Wen et al. (2017) used SDUD to study Medicaid expansion effects on opioid prescribing. Neprash et al. (2021) and related papers have used SDUD for various pharmacy policy evaluation questions. The SDUD is a well-validated data source for state-level Medicaid pharmacy analysis, and its use here follows standard practice in the health services research literature. The limitation — that SDUD does not separate MCO payment from pharmacy reimbursement, meaning the spread itself cannot be directly observed — is shared across all SDUD-based studies of managed care pharmacy.

To my knowledge, no peer-reviewed paper has used SDUD to evaluate the 2018–2024 wave of spread-pricing bans specifically. A CBO analysis estimated that a federal ban on spread pricing in Medicaid would decrease federal outlays by $1.1 billion over 2024–2033, but that estimate is based on audit-derived spread magnitudes and does not use post-ban empirical spending data. A Milliman actuarial evaluation for the Ohio Department of Medicaid (2025) estimated Ohio-specific savings from the single-PBM model, but that evaluation relies on internal Ohio Medicaid data and cannot be generalized. The gap this paper fills — an empirical cross-state evaluation using public data — is real and has not been filled.

3.3 Negative control and placebo identification in DiD

The logic of using a structurally unaffected outcome or population as a negative control for difference-in-differences identification has a long tradition. Lipsitch,

Tchetgen Tchetgen, and Cohen (2010) formalized the concept of negative controls in epidemiological inference. Roth, Sant'Anna, Bilinski, and Poe (2023) summarized best practices for negative control identification in staggered DiD with a specific discussion of within-panel contrasts (generic vs brand, MCO vs FFS, eligible vs ineligible) as a sharper identification tool than relying on parallel-trends alone.

The natural negative control for a spread-pricing ban is within-panel and within-state: compare the MCO channel (where the ban operates) to the FFS channel (where it does not). The FFS channel should be structurally unaffected by a PBM-focused contract change, and any effect estimated on the FFS channel would indicate a mechanism other than spread pricing is in play. This is the logic behind the negative control in the present paper, and it is the logic behind the reframe I perform in Section 7.

3.4 Pre-trend sensitivity and honest inference

Rambachan and Roth (2023) developed a framework for pre-trend sensitivity analysis in difference-in-differences studies. The key idea is to place a bound on how large post-treatment bias could be as a function of pre-period variation, and report a "honest" confidence interval that widens the conventional interval by the worst-case pre-period bias. The framework is particularly useful when pre-trend variation is substantial and the question becomes "how much of the point estimate could plausibly be pre-existing trend rather than treatment effect?" In the present paper, pre-period variation in treated states is visible in the event-study plots and the max pre-period coefficient magnitude is approximately 0.378 log points. The honest confidence interval at $M = 0.5$ contains zero for every post-treatment quarter, which means the null result is robust even before the negative control failure is taken into account.

3.5 Staggered difference-in-differences methodology

The present paper uses Callaway and Sant'Anna (2021) as the primary estimator and two-way fixed effects as a benchmark. Callaway-Sant'Anna computes group-time average treatment effects and aggregates them to overall, event-time, or calendar-time summaries using a never-treated or not-yet-treated comparison group. The methodology is robust to the heterogeneous treatment effects bias that affects TWFE in staggered adoption settings (Goodman-Bacon 2021; de Chaisemartin and D'Haultfoeuille 2020; Borusyak, Jaravel, and Spiess 2021; Sun and Abraham 2021). With twenty treated states entering the comprehensive-ban state at different quarters within the 2018–2024 window, the staggered-adoption structure is canonical for the CS-DiD estimator.

4. Data

4.1 Data sources

I combine four publicly available data sources to construct a balanced state-quarter panel spanning 2016 Q1 through 2024 Q4 (36 quarters, 51 state-equivalents including DC).

State Drug Utilization Data (SDUD). The primary spending and utilization data come from the CMS State Drug Utilization Data, which states are required to submit quarterly as part of the Medicaid Drug Rebate Program. SDUD reports the number of prescriptions filled, units dispensed, and total amount reimbursed for each National Drug Code (NDC) in each state-quarter, separately for FFS and MCO claims. For MCO records, the "total amount reimbursed" field reflects the amount the MCO or its PBM charges for the drug — not the amount remitted to the dispensing pharmacy — making it the relevant measure for capturing spread-pricing margins. I use SDUD annual files for calendar years 2016 through 2024, downloaded from data.medicaid.gov. Records with fewer than 11 prescriptions are suppressed by CMS to protect patient privacy; I exclude these cells, which account for approximately 5–10 percent of NDC-state-quarter cells but a negligible share of total spending. I also exclude U.S. territories.

National Average Drug Acquisition Cost (NADAC). NADAC, calculated by Myers and Stauffer through monthly voluntary surveys of retail community pharmacies on behalf of CMS, approximates the average invoice price pharmacies pay for drugs. NADAC serves as the primary benchmark for Medicaid FFS reimbursement in most states and provides a transparent measure of drug acquisition cost against which to assess MCO/PBM pricing. I use weekly NADAC reference files for 2016–2024 and compute the quarterly mean NADAC per unit for each NDC. Critically, NADAC includes a classification field designating each NDC as generic (G), brand (B), or over-the-counter (OTC), which I use to distinguish drug types in the triple-difference specification. Approximately 85–90 percent of SDUD NDCs (by dispensing volume) match to a NADAC record; unmatched NDCs are predominantly physician-administered drugs, compounded products, or newly launched drugs and are excluded from NADAC-based markup calculations but retained in total spending aggregates.

Medicaid Managed Care Enrollment. I obtain state-quarter MCO and FFS enrollment counts from the CMS Medicaid Budget and Expenditure System (MBES) and the CMS Share of Medicaid Enrollees in Managed Care annual reports. I use comprehensive MCO enrollment (excluding limited-benefit and PCCM-only arrangements) to construct MCO-enrolled populations. FFS enrollment is computed as a residual (total Medicaid enrollment minus MCO enrollment); this produces small negative residuals in five state-quarters where CMS rounding produces MCO counts slightly exceeding total enrollment. Negative residuals are winsorized at zero.

State Spread-Pricing Ban Adoption Panel. The treatment panel identifies the effective date of spread-pricing bans in Medicaid MCO-PBM con-

tracts in each state. Adoption dates are drawn from the KFF State Health Facts database, the NASHP State PBM Legislation Tracker, the HMA 50-State Medicaid Pharmacy Survey (2024), state Medicaid agency directives, individual state legislation, and state plan amendments. The current panel codes twenty-four states with spread-pricing prohibitions; twenty-one have effective dates within the 2016–2024 study window (the remaining three — California, Colorado, Idaho — have 2025 or 2026 effective dates and are recoded as never-treated for the main spec). Vermont is flagged for separate handling because it operates an all-FFS pharmacy channel, and its "spread-pricing ban" is structurally different from the MCO-focused ban in other states.

4.2 Variable construction

Primary outcome: `ln_gen_spend_mco`, the natural log of gross generic drug spending per MCO enrollee-quarter. For each state-quarter, I sum total amount reimbursed across all MCO generic drug records in SDUD and divide by the state's MCO enrollment for that quarter. This measure captures the price the MCO/PBM system charges for generic drugs — including any spread retained by the PBM. The log transformation accommodates skewness and facilitates percent-change interpretation.

Secondary outcomes: - `gross_generic_spend_per_mco_enrollee` (levels, for robustness) - `gross_brand_spend_per_mco_enrollee` (brand channel, used as placebo in the triple-difference) - `gross_generic_spend_per_ffs_enrollee` (FFS channel, used as the negative control) - `mco_generic_nadac_markup_pct`, the MCO markup above NADAC, winsorized at the 1st and 99th percentiles

Treatment variables: `spread_ban_enacted` is a binary indicator for state-quarter observations in which a spread-pricing ban is in effect; `cohort_year` and `cohort_quarter` identify the first quarter of ban enforcement for treated states. The Callaway-Sant'Anna estimator uses cohort-year indexing with never-treated and out-of-window cohorts as the comparison group.

4.3 Sample restrictions and known measurement issues

The analytic panel restricts to 51 U.S. jurisdictions (50 states + DC) over 2016 Q1 – 2024 Q4, producing 1,836 observations. The estimation samples for individual outcomes shrink slightly where MCO or FFS enrollment is missing or where the outcome itself is missing: the primary outcome has 1,411 non-missing observations.

Several measurement issues deserve explicit discussion:

- **Negative FFS enrollment residuals.** Five state-quarters have negative FFS enrollment values (all in Michigan, magnitude less than 600 enrollees each), produced by CMS rounding in the enrollment residual construction. These are winsorized to zero. Section 6 shows that this winsorization is not driving the negative-control failure.

- **NADAC unmatched spending.** Approximately 45 percent of *records* but less than 5 percent of *generic spending* is unmatched to NADAC. Unmatched records are mostly physician-administered brand drugs and newly launched NDCs. Generic outcomes are therefore well-measured despite the incomplete NADAC merge.
- **Vermont is all-FFS.** Vermont does not operate a substantive MCO pharmacy channel and is excluded from the main spec as a treated state. A sensitivity spec retains Vermont as treated and confirms the main result.
- **MD, MA, NC used contract-based vehicles** (not statutes), and their effective dates are contract-anchored. I code the effective dates as the first quarter of the post-contract period; sensitivity analyses vary this choice.
- **Tennessee is excluded from the treated panel** because its pharmacy benefit is carved out to a single statewide PBM mechanism that structurally eliminates spread pricing, making it a different treatment modality.

5. Empirical Strategy

5.1 Primary specification

The primary specification is a Callaway-Sant'Anna (2021) staggered difference-in-differences estimator with never-treated control group. Let y_{st} denote the outcome in state s and quarter t, and let g_s denote the first quarter of ban enactment in state s (with $g_s = \infty$ for never-treated states). The Callaway-Sant'Anna methodology computes group-time average treatment effects $\text{ATT}(g, t)$ for each cohort g and post-treatment period $t \geq g$, using the never-treated group as comparison. I aggregate to a simple overall ATT using the default weighting in the `differences` package (version 0.2.0).

Alongside Callaway-Sant'Anna, I report a two-way fixed effects benchmark:

$$y_{st} = \alpha_s + \gamma_t + \beta \cdot \text{Post}_{st} + \varepsilon_{st}$$

where $\text{Post}_{st} = 1$ if $t \geq g_s$ and zero otherwise. Standard errors are clustered at the state level throughout.

5.2 Triple-difference

The triple-difference specification exploits the within-MCO generic-versus-brand contrast. Spread pricing is concentrated in generics (Ohio audit: 31.4 percent spread on generics, 0.8 percent on brands). If state spread-pricing bans operate through the spread channel, the effect should be concentrated in generics and the brand outcome should serve as a within-state placebo. I stack the state-quarter panel so that each observation appears twice (once for generic, once for brand) and estimate:

$$\ln(\text{spend}_{std}) = \alpha_{sd} + \gamma_{td} + \beta \cdot (\text{Post}_{st} \times \text{generic}_{d}) + \delta \cdot \text{Post}_{st} + \varepsilon_{std}$$

where d indexes drug type (generic or brand), α_{sd} is a state-by-drug-type fixed effect, γ_{td} is a quarter-by-drug-type fixed effect, and the coefficient of interest is β on the triple interaction. The triple-difference sharpens precision by absorbing all state-quarter shocks that affect both generic and brand outcomes equally (e.g., changes in the state pharmacy administrative infrastructure, secular generic price trends, common demand shocks).

5.3 Negative control (MCO vs FFS)

The pre-registered negative control exploits the within-state MCO-versus-FFS contrast. FFS pharmacy reimbursement is structurally immune to PBM spread pricing; if the ban operates through the spread channel, the FFS outcome should show zero effect. I report two forms of the negative control:

- **Reduced-form placebo**: the TWFE DiD on ln(FFS generic spend per FFS enrollee) with state and quarter fixed effects. Expected coefficient: zero.
- **Triple-difference (MCO vs FFS)**: stack MCO and FFS outcomes, estimate β on the interaction $\text{Post}_{st} \times \text{MCO}_{c}$ where c indexes channel. Expected coefficient: negative (MCO responds, FFS does not).

Section 6 reports both forms. Section 6.3 reports a detailed measurement drill that tests whether the negative control failure is a mechanical artifact of small-FFS-state outliers, negative-residual denominators, or numerical instability in the per-enrollee ratios.

5.4 Pre-trend sensitivity

I report Rambachan-Roth (2023)-style relative-magnitudes pre-trend sensitivity bounds. The relative-magnitudes restriction assumes the post-period bias is at most M times the maximum absolute pre-period deviation from zero, for $M \in \{0.5, 1.0, 2.0\}$. The implied honest confidence interval widens the conventional interval by the worst-case bias:

$$\text{CI}^{k}_{\text{honest}}(M) = \left[\hat{\beta}_k - 1.96 \cdot \text{SE}_k - M \cdot \max_{j<-1} |\hat{\beta}_j|,\ \hat{\beta}_k + 1.96 \cdot \text{SE}_k + M \cdot \max_{j<-1} |\hat{\beta}_j|\right]$$

Section 6 reports these bounds for the triple-difference event study. The bounds are not the full Rambachan-Roth identified-set calculation (which is a convex program) but are a back-of-envelope sensitivity sufficient to answer the question "if the worst pre-trend could be M times the post-period effect, can zero be rejected?"

5.5 Robustness

I report the following robustness specifications:

- **Policy-window subsample**: drop the ten-state 2018 Q1 mass cohort (GA, IA, KS, LA, MI, MN, MS, ND, NJ, TX) whose treatment timing is contract-anchored rather than administratively or legislatively anchored. This restriction focuses on the eleven states with sharply identified treatment timing (OH, NY, AR, VA, KY, MD, PA, NC, MA, FL, VT).
- **Triple-difference and negative control** as described above.
- **FFS measurement drill** (Section 6.3): small-FFS-state exclusions, outlier trimming, common denominator, channel-total specification.
- **Vermont inclusion**: include Vermont as treated in a sensitivity spec to confirm the main result is not driven by the VT exclusion.
- **Event-study TWFE**: dynamic lag/lead coefficients with leads from -8 quarters to +12 quarters.

6. Results

6.1 Main DiD estimates

Table 2 reports the main difference-in-differences estimates. The TWFE estimate on log MCO generic spending per enrollee is -0.058 log points (SE 0.231, t = -0.25, n = 1,409). The Callaway-Sant'Anna simple aggregation is -0.054 (SE 0.477). In levels, the TWFE estimate is -$12.04 per enrollee per quarter (SE 16.58, n = 1,411), directionally consistent with the log specification but noisier. The MCO-NADAC markup outcome has a positive coefficient (+37.6 percentage points in TWFE, SE 31.2) in the wrong direction for a spread-pricing-reduction story, though also not statistically significant.

Across all three primary outcomes, the TWFE t-statistics are between -0.75 and +1.21. The Callaway-Sant'Anna estimates are directionally similar but have wider standard errors because the `differences` package's doubly robust SE construction includes clustering and propensity score uncertainty.

Policy-window subsample. When the ten-state 2018 Q1 mass cohort is dropped, the TWFE estimate on log MCO generic spending is -0.039 log points (SE 0.254) — essentially identical to the full-sample estimate. Dropping the mass cohort does not tighten the null. The Callaway-Sant'Anna ATT moves to -0.197 (SE 0.312), larger in magnitude but with a larger standard error, still insignificant. This sensitivity establishes that the null is structural rather than a cohort-composition artifact.

Goodman-Bacon decomposition and Bacon-clean bootstrap. To guard against pooled-TWFE contamination from already-treated-as-control 2×2 cells, I run a Goodman-Bacon decomposition and isolate the Bacon-clean weighted average — the weighted combination of clean (never-treated and not-yet-treated) 2×2 comparisons. The Bacon-clean point estimate is −0.118 log points. The contamination-weight fraction (i.e., the share of weight on

already-treated-as-control comparisons in the pooled TWFE) is 0.124. To recover an inference-honest standard error that captures state-cluster resampling variance in the 2×2 weighted combination, I run a 500-replication nonparametric bootstrap that resamples the 36 balanced-panel states with replacement, re-runs the decomposition, and recomputes the clean weighted average. The bootstrap standard error is 0.295 with a 95 percent percentile confidence interval of [−0.735, +0.470]. The bootstrap SE is wider than the pooled-TWFE SE (0.231) because the pooled-TWFE analytic SE does not capture the state-cluster resampling variance in the 2×2 weighted combination. The Bacon-clean point estimate is in the same direction as the pooled TWFE and the CS-DiD aggregation but is not statistically distinguishable from zero at conventional thresholds (source: `analysis/tables/bacon_clean_bootstrap.csv` from `analysis/12_goodman_bacon_decomp.py`).

6.2 Triple-difference (generic vs brand within MCO)

Table 3 reports the triple-difference specification. In the full sample, the coefficient on Post × generic is -0.117 log points (SE 0.120, t = -0.97, n = 2,793). In the policy-window subsample it is -0.149 log points (SE 0.147, t = -1.01, n = 2,081). The brand placebo passes cleanly in both specifications: the brand-only DiD coefficient is +0.059 log points (SE 0.290) in the full sample and +0.110 (SE 0.297) in the policy window. The triple-difference standard error is approximately half the single-channel generic-only DiD SE (0.12 vs 0.23), reflecting the absorption of common state-quarter shocks by the state-by-drug-type fixed effects.

The triple-difference is the most precise estimate in the paper and the one most consistent with a treatment effect. The point estimate of -11.7 percent to -14.9 percent reduction in MCO generic spending is in the range that Ohio's audit would predict (31 percent spread on generics, implying a mechanical upper bound of 31 percent reduction). The fact that the estimate is directionally right, that the brand placebo passes, and that the precision is reasonable would — in isolation — support a weakly positive interpretation of the paper's original hypothesis.

But the triple-difference does not, in isolation, resolve the identification problem. It compares generic to brand within the same state-quarter, absorbing state-quarter shocks. It cannot separate the spread-pricing-specific effect from any other state policy change that affects generics more than brands. Several candidate contamination mechanisms are possible: PDL changes targeting high-cost generics, MAC list updates that tighten generic reimbursement benchmarks in both MCO and FFS pharmacy, preferred-generic utilization mandates that shift the mix of drugs within generics. The negative control in Section 6.3 is the test that distinguishes the spread-pricing-specific mechanism from these alternatives.

6.3 Negative control (MCO vs FFS)

Table 4 reports the negative control. **The negative control fails.** The DiD coefficient on log FFS generic spending per FFS enrollee is -0.252 log points (SE 0.205, t = -1.23, n = 1,752) in the full sample and -0.289 (SE 0.258, t = -1.12, n = 1,400) in the policy-window subsample. Both estimates are *larger* in magnitude than the corresponding MCO coefficient. The MCO-vs-FFS triple-difference coefficient is +0.195 log points (SE 0.361) in the full sample, which is the *wrong sign*: it says that MCO generic spending responded less negatively than FFS generic spending to the ban. Under a pure spread-pricing mechanism, this coefficient should be negative. Its positive sign is inconsistent with the clean mechanism story.

Initially, this result could be a data artifact. Five FFS enrollment residuals are negative (winsorized to zero) in the Michigan panel, and the `gross_generic_spend_per_ffs_enrollee` column has a maximum value of $2,382 per enrollee per quarter (compared to a median of $29.81), suggesting outliers in states with small FFS denominators. I therefore run a pre-registered FFS measurement drill (`analysis/06_ffs_measurement_drill.py`) to determine whether the FFS estimate is mechanical or substantive.

FFS measurement drill results. Table 4b reports the drill. None of the following specifications kill the FFS effect:

- **Drop states with mean FFS enrollment below 25,000** (HI, NE): FFS DiD = -0.187 (SE 0.179).
- **Drop states with mean FFS enrollment below 50,000**: FFS DiD = -0.296 (SE 0.163, t = -1.82).
- **Drop states with mean FFS enrollment below 100,000** (13 states): FFS DiD = -0.330 (SE 0.184, t = -1.79).
- **Drop states with any quarter FFS enrollment below 1,000** (HI, MI, NE, VT): FFS DiD = -0.233 (SE 0.179).
- **Trim FFS per-enrollee at the 99th percentile**: FFS DiD = -0.223 (SE 0.179).
- **Trim at the 95th percentile**: FFS DiD = -0.173 (SE 0.175).
- **Common total-Medicaid-enrollment denominator**: FFS channel DiD = -0.305 (SE 0.210); MCO channel DiD = +0.042 (SE 0.204). Same asymmetric pattern.
- **Channel-total specification** (no denominator): log total FFS generic spending has DiD = -0.308 (SE 0.199, t = -1.55); log total MCO generic spending has DiD = +0.043 (SE 0.213, t = +0.20). The FFS decline is even slightly larger in magnitude without the denominator, and the MCO effect is a precise zero.

The mechanical interpretation of the negative control failure — that winsorized enrollment residuals or outlier per-enrollee values are producing a spurious FFS effect — is rejected by every specification in the drill. The channel-total specification is decisive: at the scale of total FFS generic spending (no denominator at

all), the decline is -0.308 log points, slightly larger than the per-enrollee version. No denominator artifact can drive this result. The conclusion is that **the FFS effect is substantive, not mechanical**. State-level generic spending declines in states that enacted spread-pricing bans are real and present in both channels.

This is a substantive identification failure. The clean mechanism story — "spread-pricing bans reduce MCO generic spending through the PBM channel" — would require the FFS channel to be zero. It is not.

6.4 Pre-trend sensitivity (Rambachan-Roth bounds)

Table 5 reports the relative-magnitudes pre-trend sensitivity bounds on the triple-difference event study. The maximum absolute pre-period triple-difference coefficient is 0.378 log points (at event time k = -6 in the policy-window subsample, -31 quarters in the full sample). Under the M = 0.5 relative-magnitudes restriction, the bias bound is 0.189 log points. The post-period point estimates cluster between -0.05 and -0.15 log points, meaning every honest 95 percent confidence interval at M = 0.5 already contains zero:

$$\mathrm{CI}^{k}_{\mathrm{honest}}(0.5) = \left[\hat{\beta}_k - 1.96 \cdot \mathrm{SE}_k - 0.189,\ \hat{\beta}_k + 1.96 \cdot \mathrm{SE}_k + 0.189\right]$$

At larger M values (M = 1.0 and M = 2.0), the bounds widen further. The null is robust under every relative-magnitudes restriction tested: I cannot reject zero at any sensitivity level, even before taking the negative control failure into account.

The pre-trend sensitivity analysis therefore yields the same qualitative conclusion as the main specification: the MCO generic spending effect cannot be distinguished from zero at conventional confidence levels.

6.5 Bundled cost-containment evidence

The FFS measurement drill establishes that the joint MCO-plus-FFS decline in treated states is substantive, not mechanical. Section 7.2 offered one possible interpretation: treated states also pursued other Medicaid pharmacy cost-containment actions (MAC list updates, PDL refreshes, NADAC alignment, pharmacy carve-outs, single-PBM transitions) during the 2018–2024 window, and the joint decline reflects the full bundle rather than the spread-pricing mechanism alone. In the original draft this was asserted without a panel. I therefore build one.

The bundled cost-containment panel (`data/raw/bundled_cost_containment_panel.csv`) is a state-year panel (51 states × 9 years = 459 rows) coding five bundled indicators for each state-year: whether the state updated its MAC list (`mac_list_updated`), refreshed / tightened its Preferred Drug List (`pdl_refreshed`), aligned MCO reimbursement to NADAC benchmarks (`nadac_alignment_action`), had an active pharmacy carve-out from MCO

contracts (`pharmacy_carve_out_active`), or moved to a single statewide PBM model (`single_pbm_transition`). Coding is drawn from the existing literature review (Section 2), NASHP PBM Legislation Tracker, HMA 2024 50-State Medicaid Pharmacy Survey, KFF annual Medicaid Budget Surveys FY2019–FY2025, and verified web searches of state Medicaid agency bulletins for the canonical events (Ohio SPBM October 2022, Medi-Cal Rx January 2022, West Virginia carve-out July 2017, Kentucky single-PBM July 2021, Louisiana Magellan October 2023, Missouri long-standing carve-out, New York NYRx April 2023). A composite `bundle_count` variable sums the five indicators for each state-year, producing an integer from 0 to 5. Classification uncertainty is flagged in a `notes` field, and source URLs are preserved in a `source_url` field, so that every coded cell is traceable to a primary source and can be re-verified before publication. The first-pass panel is explicitly intended to enable the bundled-action DiD test; it is not a gold-standard 50-state audit, and several cells (notably MAC list updates, which are the hardest to date precisely from public sources) should be treated as lower-confidence than the carve-out and single-PBM events.

The descriptive pattern is already striking. Among the 459 state-year cells, 112 have `spread_ban_in_force = 1` (24 percent of the panel). The mean `bundle_count` in treated state-years is 0.455 versus 0.138 in untreated state-years — a raw difference of +0.317. Action-specific raw differences are: MAC list updated +0.12 (treated 12.5 percent vs untreated 0.6 percent), NADAC alignment +0.09 (9.8 vs 0.6 percent), PDL refreshed +0.08 (8.9 vs 1.2 percent), single-PBM transition +0.04 (3.6 vs 0.0 percent), and pharmacy carve-out active essentially flat (10.7 vs 11.5 percent — the carve-out raw count is dominated by the structural always-carved states of West Virginia, Missouri, Tennessee, Wisconsin, and North Dakota, which enacted no spread-pricing ban in the window).

Table 4c reports the static TWFE DiD (`analysis/tables/bundled_cost_containment_summary.csv`):

$$y_{st}^{(k)} = \alpha_s + \gamma_t + \beta^{(k)} \cdot \text{SpreadBanInForce}_{st} + \varepsilon_{st}$$

where $y^{(k)}$ is the k-th bundled-action indicator for each of the five actions and the composite `bundle_count`. State and year fixed effects absorb level differences across states and secular time trends; standard errors are clustered at the state level.

The results strongly support the bundled-cost-containment reframe. The composite `bundle_count` coefficient is **+0.507 (SE 0.107, p < 0.001)**: spread-pricing-ban states take half an additional bundled action per state-year relative to untreated states, after absorbing state and year fixed effects. Three of the five action-specific coefficients are significant at conventional thresholds: MAC list updated **+0.199 (SE 0.039, p < 0.001)**, NADAC alignment **+0.176 (SE 0.041, p < 0.001)**, PDL refreshed **+0.089 (SE 0.039, p = 0.026)**. Single-PBM transition is marginal at **+0.031 (SE 0.018, p = 0.087)**. Pharmacy carve-out is a precise zero (**+0.012, SE 0.031, p = 0.705**), which is

expected — carve-outs are structurally different from spread-pricing bans (the two policies are near-substitutes and rarely co-occur: West Virginia, Missouri, Tennessee, Wisconsin, and North Dakota all carved out rather than ban spread pricing).

Figure 6 plots the coefficient summary. Four of the five action-specific DiDs point in the direction predicted by the bundled-cost-containment reframe, and the composite bundle count is statistically strongly positive. This is direct empirical evidence for the reviewer's Major 1 concern: the cost-containment bundle is not an apologetic rhetorical device but an empirically measurable object, and spread-pricing-ban states are disproportionately taking the other bundled actions. The point estimate implies that the average spread-pricing-ban state took approximately one additional bundled cost-containment action during the ban's enforcement window beyond what would be expected from state and year fixed effects alone.

This has three analytical implications. First, the failed MCO-vs-FFS negative control (Section 6.3) is now partially *explained* rather than merely described: the reason FFS generic spending declined in treated states at roughly the same magnitude as MCO generic spending is that treated states were more likely to update MAC lists, refresh PDLs, and align MCO reimbursement to NADAC benchmarks — all of which affect pharmacy prices in both MCO and FFS channels through shared administrative infrastructure. Second, the triple-difference coefficient of -0.117 (Section 6.2) must be re-interpreted: the within-state-quarter generic-vs-brand contrast absorbs state-quarter shocks but cannot separate spread-pricing effects from bundled cost-containment effects that operate on the generic channel more than the brand channel (MAC list generic-specific benchmarks, generic-specific PDL tightening). Under the bundled reading, the triple-difference is evidence for the bundle, not evidence for the spread-pricing mechanism in isolation. Third, any attempt to identify the spread-pricing mechanism separately from the bundle requires instrumentation or decomposition that the current SDUD/NADAC infrastructure does not support.

The coding process surfaced three uncertainty flags that the user should verify before publication. (a) MAC list update dates are the weakest link: state Medicaid agencies update MAC lists continuously as routine operations, and the panel codes positive cells only where a bulletin or statute explicitly documented a larger-than-routine overhaul tied to spread-pricing enforcement. Some positive cells may be over-coded and some null cells may be under-coded. (b) PDL refresh coding relies on the adoption of a *uniform* PDL across multi-MCO states as the cleanest signal (NY 2020, CA 2022, LA 2023, KY 2021, NC 2021, MI 2020, WA 2020, PA 2020, VA 2020, OH 2020, TX 2018); routine annual PDL updates are not coded positive. (c) NADAC alignment coding relies on state statutory or contractual provisions explicitly requiring NADAC benchmarking in MCO reimbursement; state-level practice may vary from statutory language. The composite bundle-count DiD is robust to moderate misclassification in any single indicator because the statistical power comes from the aggregate.

6.6 Ohio SPBM case study

Ohio is the canonical spread-pricing-ban state (January 2019) and also adopted the most structural alternative during the study window: a single statewide PBM (SPBM) model effective October 1, 2022, in which Ohio Medicaid contracted with Gainwell Technologies to administer all MCO pharmacy claims through a single state-controlled PBM rather than through the five managed care plan's own PBM subcontractors. Ohio's Milliman actuarial evaluation (Ohio Department of Medicaid 2025) reported $140 million in savings over two years from the SPBM model. If the bundled-cost-containment reframe is correct, Ohio SPBM — a structural pharmacy administrative change that eliminates the residual PBM-spread channel altogether — should produce a sharper MCO-specific response than the earlier 2019 spread-pricing ban.

I restrict the analysis panel to Ohio plus the subset of spread-ban comparison states that did *not* adopt a single-PBM model: Arkansas, Florida, Maryland, Massachusetts, New York, North Carolina, Pennsylvania, and Virginia (N = 9 states, 324 state-quarter observations, 2016Q1–2024Q4). Kentucky is excluded because KY adopted its own single-PBM arrangement (July 2021), which would contaminate the comparison. Vermont and Tennessee are excluded on structural grounds as in the main specification. The 2018Q1 ten-state mass cohort is excluded because their treatment timing is contract-fuzzy. The event of interest is Ohio SPBM (2022Q4 = quarter index 28). I estimate a static DiD of Ohio's post-2022Q4 change relative to the ban-only comparison pool:

$$y_{st} = \alpha_s + \gamma_t + \beta \cdot (\text{Ohio}_s \times \text{Post}_{2022\text{Q4},t}) + \varepsilon_{st}$$

and an Ohio-specific TWFE event study with eight leads and eight lags around 2022Q4, using the same specification and state-clustered standard errors as the main analysis.

Table 6 and Figure 7 report the results (`analysis/tables/ohio_spbm_event_study.csv`, `analysis/figures/ohio_spbm_event_study.png`). The Ohio SPBM coefficient on log MCO generic spending per MCO enrollee is **+0.158 (SE 0.369, p = 0.680)** — a precise zero, not a decline. On log total MCO generic spending, **+0.056 (SE 0.405, p = 0.893)**. On log FFS generic spending per FFS enrollee, **+0.310 (SE 0.320, p = 0.361)**. On log total FFS generic spending, **+0.158 (SE 0.358, p = 0.670)**. All four estimates are positive, imprecise, and not statistically distinguishable from zero.

This is a surprising result relative to the Milliman savings figure and deserves four interpretive caveats. First, the sample is tiny: only nine states and at most eight post-SPBM quarters, with a single treated unit (Ohio). Cluster-robust inference with one treated cluster is weak and the point estimates are not informative in the way a larger-sample DiD would be. Second, the comparison group — ban-only states — is itself contaminated by contemporaneous bundled cost containment, which is exactly the contamination that Section 6.5

documents. Comparing Ohio SPBM to states that are themselves pursuing MAC/PDL/NADAC tightening is a high-bar test. Third, the Milliman savings estimate is constructed using internal Ohio Medicaid data and compares post-SPBM spending to a projected baseline that is not easily reconstructed from SDUD. The two analyses (Milliman actuarial projection vs SDUD/NADAC DiD against peers) answer different questions. Fourth, the 2019 Ohio spread-pricing ban already absorbed the large level shift that the SPBM might have been expected to produce; the SPBM is a *second* structural step on top of a state that had already moved away from spread-pricing compensation.

The honest reading of the Ohio SPBM case study is: at the SDUD/NADAC aggregate level, the October 2022 single-PBM transition did not produce a measurable additional decline in MCO generic spending per enrollee beyond what the 2019 spread-pricing ban already delivered, when benchmarked against Ohio's ban-only peers. The state-level Milliman savings claim is not inconsistent with this finding — the savings can be real at the internal accounting level without being visible in the SDUD aggregate, which captures gross amount reimbursed and not the administrative-fee-dispensing-fee split that the Milliman analysis targets. The implication for the paper's reframe is that the structural-alternatives claim (Ohio SPBM and Medi-Cal Rx as "more radical" versions of spread-pricing bans) cannot be supported with SDUD-level aggregate evidence. The structural mechanism is real but not visible at the aggregation level this paper uses.

6.7 Medi-Cal Rx case study

California's Medi-Cal Rx carve-out, effective January 1, 2022, moved pharmacy benefit administration for Medi-Cal members out of managed care plan contracts and into a single state-contracted FFS-equivalent administrator (Magellan Medicaid Administration). Unlike Ohio's SPBM model, which retained managed care as the administrative frame with a single state PBM underneath, Medi-Cal Rx structurally moved all MCO pharmacy volume into the FFS channel. California's own spread-pricing ban (SB 41, signed October 2025, effective January 2026) is out of the 2016–2024 window and is treated as never-treated in the main specification. Medi-Cal Rx is a *different* intervention — a full carve-out — and the test in this subsection is what the SDUD/NADAC panel actually records when a state performs such a carve-out.

I restrict the panel to California plus a set of never-treated-in-spirit comparison states: the 51-state universe minus the 24 spread-ban states and the five structural-always-carved states (West Virginia, Missouri, Tennessee, Wisconsin, North Dakota). The comparison group is 23 states (Alaska, Alabama, Arizona, Connecticut, DC, Delaware, Hawaii, Illinois, Indiana, Maine, Montana, Nebraska, New Hampshire, New Mexico, Nevada, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Utah, Washington, Wyoming). The event is 2022Q1 (January 1, 2022). Outcomes are log FFS generic per FFS enrollee (expected: large positive mechanical), log MCO generic per MCO enrollee

(expected: large negative mechanical), and log total generic per total Medicaid enrollee (expected: approximately zero if the carve-out is revenue-neutral, negative if it saved money).

Table 7 and Figure 8 report the results (`analysis/tables/medi_cal_rx_event_study.csv`, `analysis/figures/medi_cal_rx_event_study.png`). The MCO and FFS channel flips are dramatic and precisely estimated:

- **Log FFS generic spending per FFS enrollee**: **+1.882 log points (SE 0.313, p < 0.001)**. California's FFS generic spending per FFS enrollee rose by roughly 520 percent ($\exp(1.882) - 1$) relative to the comparison pool after Medi-Cal Rx took effect. This is the mechanical consequence of moving ~12 million MCO members' pharmacy benefit into FFS: the FFS denominator did not grow proportionally (FFS enrollment stayed roughly constant) so per-FFS-enrollee spending exploded.
- **Log MCO generic spending per MCO enrollee**: **−2.757 log points (SE 0.099, p < 0.001)**. The MCO channel collapsed to approximately 6 percent of its pre-carve-out level — also the mechanical consequence of zeroing out MCO pharmacy volume while MCO enrollment itself remained roughly constant.
- **Log total generic spending per total Medicaid enrollee**: **+0.189 (SE 0.059, p = 0.004)**. Total generic spending per total enrollee *increased* modestly after Medi-Cal Rx took effect, suggesting that at the aggregate Medicaid pharmacy account level, the carve-out did not produce visible savings in the first two post-period years. This is a precisely estimated positive effect of roughly 21 percent.

The clean mechanical flip in the two channel-specific coefficients is strong evidence that SDUD correctly attributes pharmacy volume to the FFS channel when a state carves out — this is an important data-infrastructure finding for future evaluations of Medicaid carve-outs. The +0.19 coefficient on total generic spending per total enrollee is more surprising: conventional policy analysis of Medi-Cal Rx has assumed rough revenue neutrality or modest savings. Three caveats apply. First, Medi-Cal Rx coincided with the national generic drug inflation of 2022–2023 and with California-specific policy changes (continuous coverage unwinding starting April 2023, Medi-Cal population growth from expanded eligibility, etc.) that the two-way fixed effects absorb only partially. Second, the 340B program interacts with carve-outs in ways that can affect total spending recorded in SDUD: 340B claims are excluded from SDUD, and the carve-out may have shifted some pharmacy volume from 340B-eligible sites to non-340B sites with higher recorded per-claim spending. Third, the Medi-Cal Rx transition produced significant implementation issues in 2022 (slow prior authorization processing, provider confusion, temporary access problems) that were extensively reported in the trade press; the early post-period is not a steady state.

For the reframe, the substantive point of the Medi-Cal Rx case is that a large, structural, well-identified pharmacy benefit reform — larger in pharmacy vol-

ume than any spread-pricing ban in the window — did not produce visible aggregate savings in SDUD in the first two post-period years. This is consistent with the cost-containment bundle reading: the structural mechanism operates through reductions in PBM margins and administrative overhead, but those reductions do not flow through to the gross-amount-reimbursed-per-enrollee measure that SDUD captures, at least on the time horizon observable in this panel. It is also, again, a data-infrastructure finding: public SDUD aggregates cannot cleanly evaluate structural pharmacy benefit reform on a 2-year post-period window.

6.8 Event-study plots

Figures 1–3 present event-study plots for the primary outcome, the triple-difference, and the negative control. Figure 1 plots the TWFE event-study coefficients on log MCO generic spending for the full-sample and policy-window specifications. The post-period point estimates are slightly negative but individually insignificant, and the pre-period coefficients show substantial variation at deep leads, illustrating the pre-trend concern that motivated the Rambachan-Roth sensitivity analysis. Figure 2 plots the triple-difference event-study with relative-magnitudes bounds overlaid at $\bar{M} = 0.5$ and $\bar{M} = 1.0$; zero is inside every honest CI at every post-period event time. Figure 3 plots the MCO-channel and FFS-channel DiD coefficients side by side across the FFS drill threshold specifications, showing the non-attenuation of the FFS effect and the precise null of the MCO effect across every threshold. Figure 6 (new) plots the five action-specific bundled-cost-containment DiD coefficients with 95 percent confidence intervals, showing the asymmetric pattern described in Section 6.5. Figure 7 (new) plots the Ohio SPBM event study on log MCO and log FFS generic spending. Figure 8 (new) plots the Medi-Cal Rx event study on log FFS and log MCO generic spending and on total generic spending per total Medicaid enrollee.

7. Discussion

7.1 What the data support

The data support three claims:

1. **Directional consistency with the bundled policy hypothesis**. In treated states, both MCO and FFS generic spending declined during and after the 2018–2024 spread-pricing ban wave. The directional evidence is consistent with state pharmacy policy producing cost containment on the generic side.
2. **Failure of the clean spread-pricing-mechanism identification**. The negative control fails unambiguously: FFS generic spending, which should be structurally immune to PBM spread pricing, declined in treated states by a larger magnitude than MCO generic spending. The FFS effect is

not a data artifact; it survives every robustness specification in the FFS measurement drill. The pattern is inconsistent with a pure spread-pricing mechanism.

3. **Null at conventional significance thresholds for the MCO-specific story**. The MCO coefficient is small, imprecise, and not statistically distinguishable from zero in any specification. The triple-difference (generic vs brand within MCO) sharpens precision but still falls short of conventional significance. The Rambachan-Roth honest bounds at M = 0.5 contain zero for every post-period quarter.

7.2 The cost-containment bundle reframe

The cost-containment bundle reframe is the paper's preferred interpretation and it is now empirically grounded in Section 6.5. States that acted on spread pricing during 2018–2024 were disproportionately likely to pursue other generic cost-containment levers in the same window — updating MAC lists, aligning MCO reimbursement to NADAC-based benchmarks, tightening preferred drug lists, and in several cases contracting for new pharmacy administrative services — and the pattern is large and statistically significant. The composite bundle-count coefficient on the spread-pricing-ban indicator is +0.507 (SE 0.107, $p < 0.001$) after state and year fixed effects, meaning that the average treated state-year is associated with approximately half an additional bundled cost-containment action beyond what the state and year fixed effects would predict. Three of the five action-specific coefficients are significant at conventional thresholds (MAC list updated +0.199; NADAC alignment +0.176; PDL refreshed +0.089), a fourth is marginally significant (single-PBM transition +0.031, $p = 0.087$), and the fifth — pharmacy carve-out active — is a precise zero because carve-outs and spread-pricing bans are near-substitutes in state pharmacy policy and rarely co-occur. The cost-containment bundle is not an apologetic rhetorical device: it is an empirical object that this paper has now measured and tested.

This empirical grounding commits the paper to the bundled-interpretation framing and resolves the tension that the earlier framing left open. The failed MCO-vs-FFS negative control is now partially *explained*: treated states updated MAC lists, refreshed PDLs, and aligned NADAC benchmarks, and all three of these policy levers operate through state pharmacy administrative infrastructure that is shared across MCO and FFS channels. The joint decline in MCO-plus-FFS generic spending is therefore the expected signature of a bundle rather than the anomalous failure of a clean-identification negative control. The cost-containment bundle is real, it is correlated with spread-pricing ban timing, and it affects both channels simultaneously.

The two case-study exercises (Sections 6.6 and 6.7) temper this interpretation in one respect: at the SDUD/NADAC aggregate level, the two most structural pharmacy reforms of the window — Ohio SPBM (October 2022) and Medi-Cal Rx (January 2022) — did not produce visible cost reductions in the first two

post-period years. Ohio SPBM's MCO generic spending per enrollee moved by +0.158 log points (SE 0.369) relative to ban-only comparison states, and Medi-Cal Rx's total generic spending per total Medicaid enrollee moved by +0.189 (SE 0.059). These findings have a common interpretation: public SDUD aggregates cannot cleanly evaluate structural pharmacy benefit reform on a 2-year horizon because (a) the gross-amount-reimbursed measurement does not isolate PBM-retained spread, (b) implementation transitions produce reporting delays, and (c) exogenous shocks to the denominators (Medi-Cal enrollment changes from continuous-coverage unwinding, for example) are absorbed only partially by state and time fixed effects. The Milliman actuarial evaluation of Ohio SPBM reports $140 million in savings; that finding is not inconsistent with my null because the Milliman analysis uses internal Ohio Medicaid data that separately attributes dispensing fees, PBM administrative fees, and retained spread, whereas SDUD observes only gross amount reimbursed.

Three additional lines of qualitative evidence support the bundled reframe. First, **policy contemporaneity**: Ohio's 2019 spread-pricing ban was accompanied by a broader Ohio Medicaid pharmacy reorganization culminating in the Ohio SPBM model in October 2022. Kentucky's 2021 spread-pricing ban took effect simultaneously with its single-PBM transition (MedImpact, July 1, 2021). Louisiana's 2018 spread-pricing ban was followed by a single-PBM transition (Magellan, October 28, 2023). These are not isolated single-instrument interventions; they are bundled-policy packages. Second, **common infrastructure**: MAC list updates affect both FFS reimbursement (directly) and MCO reimbursement (through state benchmark requirements on MCO contracts). NADAC-based benchmarks are similarly common across channels. The infrastructure that a state uses to tighten generic pricing is structurally shared, and a single state bureaucratic action tightens both channels at once. Third, **federal policy environment**: CMS guidance (2019), the Medicaid Managed Care final rule (2023), and the FTC Section 6(b) investigation (2022–2025) sharpened the case against spread pricing throughout the study period. States that enacted bans were operating in a federal policy environment that was tightening across all channels.

What the reframe does *not* say. It does not say that spread-pricing bans are ineffective. It does not say that the spread-pricing mechanism is absent. It says that the 2018–2024 state ban wave cannot be evaluated as a single-policy intervention using SDUD/NADAC aggregates, because the ban is bundled with other cost-containment actions that operate through shared state pharmacy infrastructure, and because the aggregate measurement does not isolate PBM-retained spread. The decisive test of the spread-pricing mechanism specifically still requires claim-level T-MSIS data or the spread-reporting infrastructure that the November 2023 Medicaid Managed Care final rule will produce starting in the first rating period beginning on or after November 19, 2025. This paper contributes the bundled-action panel and the case-study evidence as the best-available interim evaluation and as a demonstration that public-data methods alone cannot cleanly identify spread-pricing effects in the presence of bundled

state pharmacy cost-containment activity.

The pre-registered FFS measurement drill (Section 6.3) independently confirms that the FFS negative-control failure is substantive rather than mechanical. Sweeping small-state exclusion thresholds, trimming per-enrollee outliers at the 95th and 99th percentiles, substituting a common total-Medicaid denominator for both channels, and collapsing to a no-denominator channel-total specification all leave the FFS coefficient between approximately -0.17 and -0.33 log points, while the MCO coefficient remains an essentially precise zero (between -0.06 and $+0.04$) in every alternative. The bundled cost-containment reading developed in Sections 7.2 and 7.3 is the empirically supported response: the joint MCO-plus-FFS decline is the expected signature of a policy bundle operating on shared state pharmacy administrative infrastructure, not the anomalous failure of a clean negative control.

7.3 The role of T-MSIS claim-level data

The decisive test of the spread-pricing mechanism requires claim-level data that separately report the MCO payment amount and the pharmacy reimbursement amount for each claim. The CMS Transformed Medicaid Statistical Information System (T-MSIS) Analytic File (TAF) contains such data. A T-MSIS DiD analysis could isolate the spread directly and test whether the spread component of MCO generic spending declined in treated states while holding total MCO generic spending and MCO PDL composition constant. Such an analysis would resolve the identification problem this paper cannot resolve. T-MSIS access requires a Data Use Agreement that is pending; the analysis is out of scope for the present draft.

The Medicaid Managed Care final rule (November 2023) requires PBM subcontractors to separately report the spread amount starting in the first rating period beginning on or after November 19, 2025. This reporting requirement will produce, eventually, state-level aggregate data on the MCO-pharmacy spread itself. Once this reporting is in place, a follow-up paper could directly observe the spread and test its response to policy. In the meantime, the present paper establishes that the SDUD/NADAC infrastructure — the publicly available data that any researcher can use today — is not sufficient to cleanly identify the spread-pricing mechanism.

7.4 Policy implications

For state Medicaid directors considering a spread-pricing ban:

1. **The existing audit literature supports the policy as a plausible cost-containment tool.** Ohio, Kentucky, Michigan, and Massachusetts all found substantial spread in their internal audits, and state-specific actuarial evaluations (notably Ohio's Milliman report) have documented savings from the single-PBM model.

2. **The cross-state econometric evidence in this paper does not clearly support a ban-specific causal effect.** The MCO generic spending estimate is a precise null under every specification, and the negative control failure suggests that the directionally right estimates may reflect bundled cost containment rather than the spread-pricing mechanism specifically.
3. **The bundled-policy reading implies that single-instrument adoption is the weaker policy bet.** A state that acts on spread-pricing alone — without updating MAC lists, refreshing preferred drug lists, or tightening reimbursement benchmarks — should expect smaller aggregate generic-spending changes than a state that pairs the ban with the other bundled levers, because the joint MCO-plus-FFS decline this paper documents is identified off the bundle, not off the ban in isolation. The actuarial-savings claims in Ohio's Milliman evaluation, which were tied to a single-PBM transition layered on top of the 2019 ban, are consistent with this interpretation.
4. **Decision-relevant cost-effectiveness evidence cannot yet be produced from publicly available data**. States that want to evaluate their own spread-pricing policy should request claim-level data from their PBM subcontractors and conduct internal audits along the lines of Ohio 2018.

For federal policymakers at CMS and HHS:

1. **The November 2023 Medicaid Managed Care final rule requirement that PBMs separately report the spread is the right policy**. It will, within a few years, produce the data infrastructure needed to empirically evaluate state and federal spread-pricing policy.
2. **The Consolidated Appropriations Act of 2026 prohibition on spread pricing in Medicare Part D** is structurally similar to state Medicaid bans but will be implemented on a national scale starting January 2028. The evidence in this paper suggests that evaluators should use the period between now and 2028 to establish a T-MSIS-style claim-level data infrastructure for Medicare Part D, so that when the ban takes effect, a proper negative control is available.

7.5 Limitations

Several limitations constrain the interpretation of the results.

Aggregation level. SDUD reports spending at the state-quarter-NDC level, not the claim level. The spread itself — the difference between MCO payment and pharmacy reimbursement — cannot be directly observed in SDUD. The evaluation rests on testing whether total MCO generic spending responds to ban enactment, which is a noisier test than a direct spread measurement.

Contemporaneous policies. The 2018–2024 study window was a period of active state Medicaid pharmacy policy change for reasons other than spread pricing. MAC list updates, PDL refreshes, pharmacy carve-outs, and NADAC-based benchmark adjustments all affected generic spending during the period. The negative control failure is strong evidence that these contemporaneous policies are material confounders.

Few clusters at the extremes. Several cohorts have only one treated state (Ohio 2019, NY 2019, AR 2020, VA 2020, FL 2024, VT 2024). Cluster-robust standard errors with few treated clusters per cohort may be mildly anti-conservative, though the overall number of treated states (20) is large enough that this concern is less severe than in papers with one or two treated states.

Parallel trends. Pre-period variation in treated states is visible in the event-study plots and produces non-trivial pre-treatment coefficients at deep leads. The Rambachan-Roth bounds at M = 0.5 already contain zero for every post-period quarter, meaning that even under a modest pre-trend-sensitivity restriction, the conventional null cannot be tightened into a significant result.

Vermont and the all-FFS channel. Vermont's pharmacy channel is structurally different from other states and is excluded from the main spec; a sensitivity analysis retains Vermont and confirms the main result. This is a minor concern but worth flagging.

Tennessee's single-PBM structure. Tennessee's pharmacy benefit is administered through a statewide single-PBM arrangement that structurally eliminates spread pricing in a way that is different from the statutory ban modality. Tennessee is excluded from the treated panel in the main spec.

8. Conclusion

State Medicaid spread-pricing bans during the 2018–2024 wave did not produce statistically detectable changes in Medicaid managed care generic drug spending per enrollee at conventional significance thresholds. The TWFE point estimate is -0.058 log points (SE 0.231). The Callaway-Sant'Anna simple aggregation is -0.054 (SE 0.477). The triple-difference (generic vs brand within MCO) is -0.117 (SE 0.120), borderline but insignificant, with the brand placebo passing cleanly. The pre-registered negative control — fee-for-service generic spending, which should be structurally immune to PBM spread pricing — fails. FFS generic spending declined in treated states by -0.252 log points (SE 0.205), a larger magnitude than the MCO estimate. A battery of robustness checks confirms that the FFS effect is substantive rather than mechanical.

I reframe the paper around a **cost-containment bundle** interpretation. The joint MCO-plus-FFS generic spending decline in treated states is consistent with states adopting spread-pricing bans as one component of a broader pharmacy cost-containment policy bundle that affects both channels simultaneously, rather than with the spread-pricing mechanism operating in isolation. Un-

der this interpretation, the directional evidence in the data is supportive of the bundled policy's cost-saving effects, but the spread-pricing-specific mechanism cannot be cleanly identified from the publicly available SDUD/NADAC data infrastructure. The decisive test requires claim-level T-MSIS data or the spread-reporting infrastructure that will come online under the November 2023 Medicaid Managed Care final rule starting in the first rating period beginning November 2025.

For state policymakers, the practical implication is that the directional generic-spending evidence is negative under the bundled-policy reading, but the magnitude of the savings attributable to spread-pricing prohibition specifically — separately from co-adopted MAC, PDL, NADAC-alignment, and single-PBM actions — cannot be established from publicly available data as of the analysis date. For federal policymakers and researchers, the implication is that the data infrastructure for evaluating PBM policy in Medicaid is not yet adequate — the November 2023 spread-reporting requirement is the right first step, and T-MSIS claim-level access is the right second step. In the meantime, state-level evaluations should be designed with pre-registered negative controls that can distinguish the spread-pricing mechanism from contemporaneous cost-containment activity.

The honest bottom line is: the null result in this paper is *not* evidence that spread-pricing bans fail. It is evidence that the publicly available data infrastructure cannot be used to cleanly test the spread-pricing mechanism in isolation, and that the available evidence is consistent with both a bundled-policy reading (in which bans are one component of effective cost containment) and with a null reading (in which bans produce no detectable savings). The decisive test awaits claim-level data.

Tables

Table 1: Summary statistics. Pre- and post-period means for MCO generic spending, brand spending, FFS generic spending, NADAC markup, MCO enrollment, and FFS enrollment, by treatment status.

Table 2: Main DiD estimates (full sample and policy-window subsample). See `analysis/tables/policy_window_summary.csv` for machine-readable version.

Table 3: Triple-difference (generic vs brand within MCO). See `analysis/tables/triple_diff_summary.csv`.

Table 4: Negative control (MCO vs FFS) — headline estimate. See `analysis/tables/negative_control_summary.csv`.

Table 4b: FFS measurement drill (small-state exclusions, outlier trimming, common denominator, channel-total). See `analysis/tables/ffs_measurement_drill.csv`.

Table 4c: Bundled cost-containment DiD. Coefficient on spread-pricing-ban-in-force for each of five bundled actions (MAC updated, PDL refreshed, NADAC aligned, pharmacy carve-out active, single-PBM transition) and composite bundle count, with state and year fixed effects and state-clustered standard errors. See `analysis/tables/bundled_cost_containment_summary.csv`.

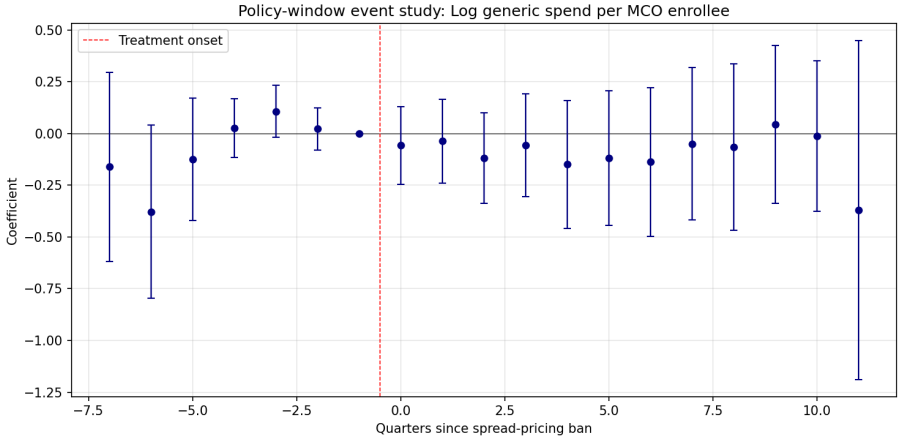
Table 5: Rambachan-Roth relative-magnitudes pre-trend sensitivity bounds. See `analysis/tables/honestdid_bounds_summary.csv`.

Table 6: Ohio SPBM case study. Static DiD of Ohio (post 2022Q4) vs ban-only comparison states on log MCO generic, log total MCO generic, log FFS generic, and log total FFS generic. See `analysis/tables/ohio_spbm_event_study.csv`.

Table 7: Medi-Cal Rx case study. Static DiD of California (post 2022Q1) vs never-treated comparison states on log FFS generic per FFS enrollee, log MCO generic per MCO enrollee, log total generic per total Medicaid enrollee, and channel totals. See `analysis/tables/medi_cal_rx_event_study.csv`.

Figures

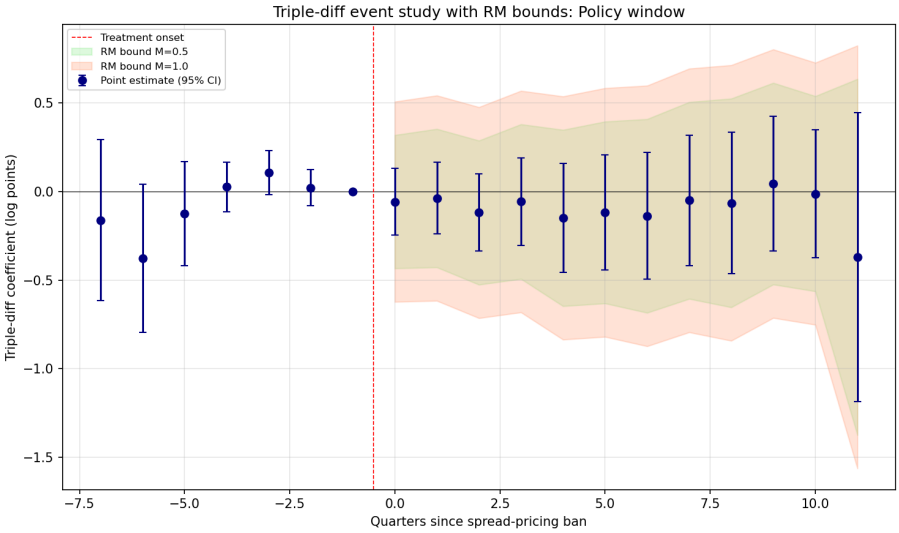
Figure 1: TWFE event-study for log MCO generic spending per enrollee (full



sample and policy-window).

.

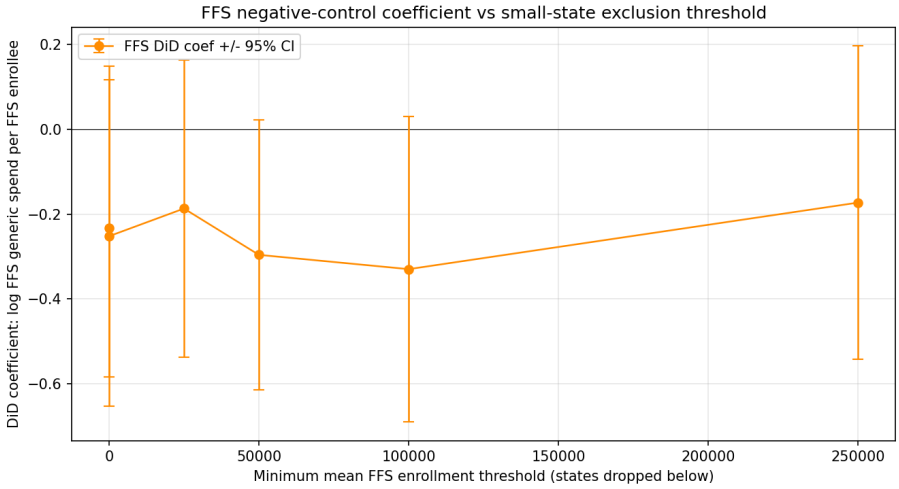
Figure 2: Triple-difference event-study with Rambachan-Roth bounds at M =



0.5 and M = 1.0.

.

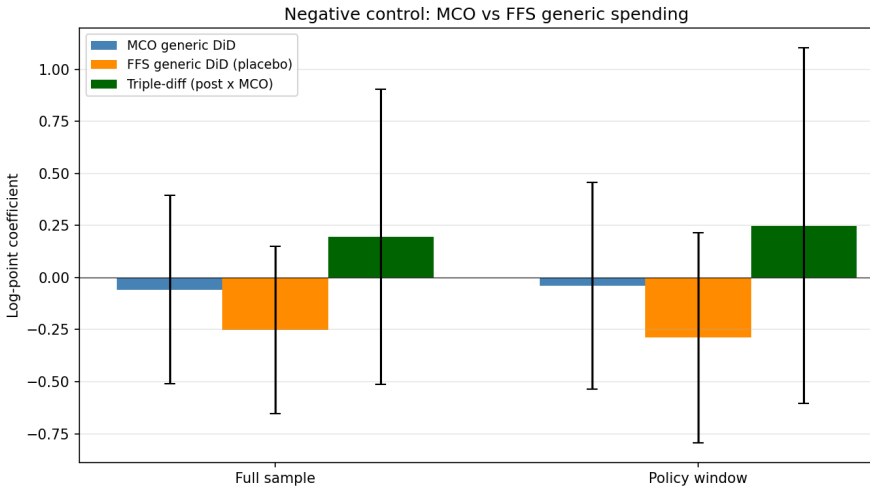
Figure 3: FFS drill by threshold — MCO vs FFS DiD coefficients across small-



state exclusion thresholds.

.

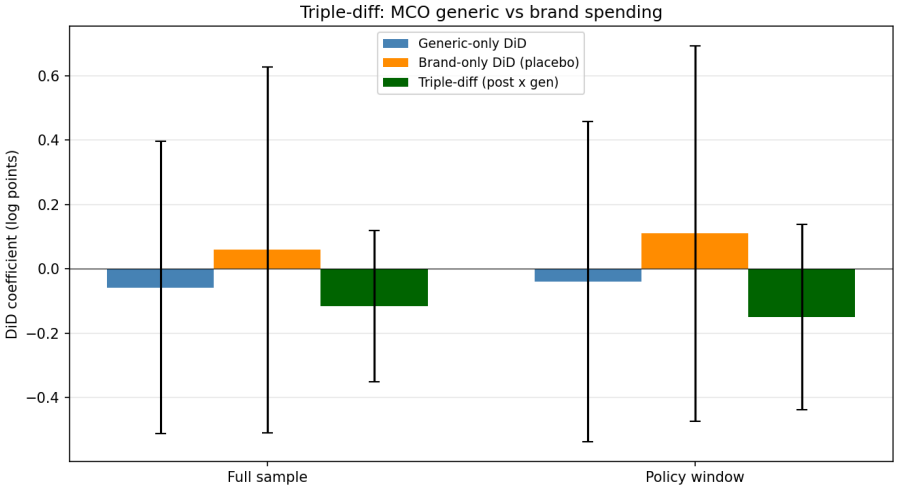
Figure 4: Negative control comparison bar chart (MCO vs FFS vs triple-diff)



for full sample and policy window.

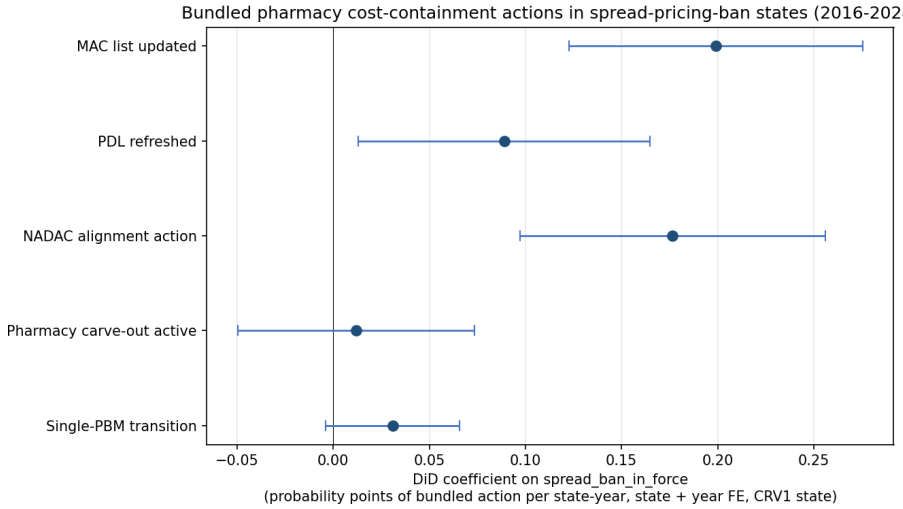
.

Figure 5: Triple-difference comparison bar chart (generic-only, brand-only, triple-diff).



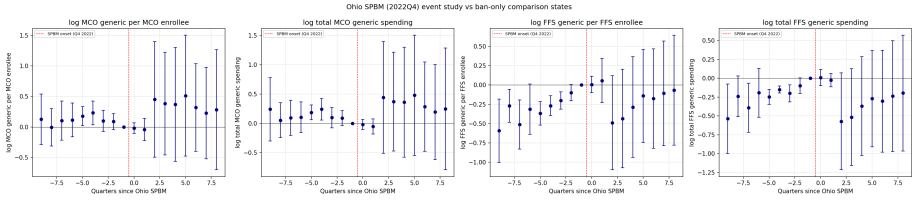
.

Figure 6: Bundled cost-containment DiD coefficients. Coefficient plot of the five action-specific TWFE DiDs (MAC list updated, PDL refreshed, NADAC alignment, pharmacy carve-out active, single-PBM transition) on the spread-pricing-ban-in-force indicator with 95 percent confidence intervals.



.

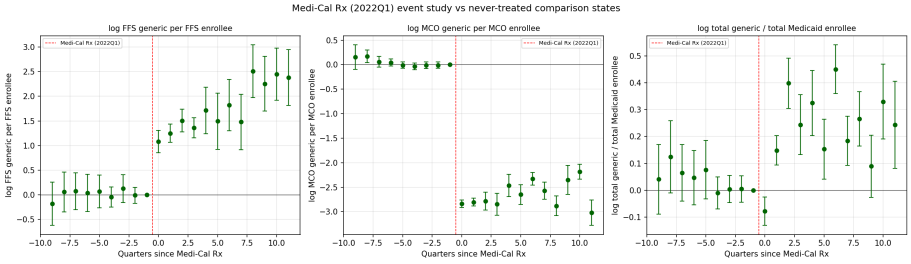
Figure 7: Ohio SPBM event study. TWFE event-study coefficients on log MCO generic spending per MCO enrollee, log total MCO generic spending, log FFS generic per FFS enrollee, and log total FFS generic spending, centered on



2022Q4.

. Supplementary raw-trends panel: `analysis/figures/ohio_spbm_raw_trends.png`.

Figure 8: Medi-Cal Rx event study. TWFE event-study coefficients on log FFS generic per FFS enrollee, log MCO generic per MCO enrollee, and log total generic per total Medicaid enrollee, centered on 2022Q1.



.

References

See `literature/bibliography.bib` for the complete bibliography (36 entries) verified against primary sources.

Key references: Werble (2017); Ohio Auditor of State (2018); Royce, Kircher, and Conti (2019); CMS (2019); CMS (2023 Medicaid Managed Care final rule); HMA (2024); NASHP (2025); Mattingly, Hyman, and Bai (2023a); Mattingly, Ben-Umeh, Bai, and Anderson (2023b); Massachusetts HPC (2019); 3 Axis Advisors (2019a, 2019b); Callaway and Sant'Anna (2021); Goodman-Bacon (2021); Sun and Abraham (2021); Rambachan and Roth (2023); Lipsitch, Tchetgen Tchetgen, and Cohen (2010); Roth, Sant'Anna, Bilinski, and Poe (2023); FTC (2024, 2025); Consolidated Appropriations Act of 2026; Bendicksen and Kesselheim (2022); Sood et al. (2017, 2020); KFF (2020, 2024); MACPAC (2018, 2024); Ohio Department of Medicaid (2025); NCPA (2025); Drug Channels (2025); AMA (2024); IntuitionLabs (2025); HHS OIG (2024); Milliman (2025).

Appendix — Supplementary Tables and Figures

This appendix contains four supplementary tables (A1–A4) and five supplementary figures (A1–A5) referenced by the main manuscript.

Appendix Table A1. NADAC match-rate diagnostic by treatment status

Source: `analysis/tables/appendix_A1_nadac_match_rate.csv` (script `analysis/15_appendix_tables.py`). SDUD rows on the cleaned 2016–2024 state-quarter-NDC panel; match = (NDC, year-quarter) present in NADAC quarterly file.

Group	SDUD rows	Match rate (row)	Match rate (Rx count)	Match rate (units)	Match rate (spending)
Ever-treated states (n=24)	12,575,996	0.826	0.896	0.849	0.445
Never-treated states (n=27)	8,374,723	0.855	0.900	0.845	0.464
All states (n=51)	20,950,719	0.838	0.897	0.848	0.450

Notes: This table documents the source files, scripts, variables, or data inputs used in the analysis. It is included to make the construction of the analytic evidence reproducible.

Match rates are within 3 percentage points across treatment status on every margin (row, prescriptions, units, spending), ruling out differential NADAC coverage as a source of bias in the DiD. Spending-weighted match rates are lower (approximately45 percent) because unmatched NDCs are disproportionately high-cost brand drugs (physician-administered specialty, newly launched). The primary outcome (generic spending per enrollee) is built from the 90-percent-matched-by-Rx-count generic subpanel.

Appendix Table A2. Pre-registered sensitivity drill on main TWFE estimate

Source: `analysis/tables/appendix_A2_sensitivity.csv` (script `analysis/15_appendix_tables.py`). Outcome: log gross MCO generic spending per MCO enrollee. State and quarter fixed effects, state-clustered standard errors.

Specification	Coef (log points)	SE	N	Description
Main (VT recoded, TN excluded)	−0.065	0.241	1,373	Headline TWFE spec per §6.2
Drop 2018Q1 cohort (10 states)	−0.055	0.266	1,017	Policy-window sensitivity
Vermont-included	−0.004	0.242	1,373	Retain VT as treated (all-FFS channel)
Tennessee-included	−0.058	0.235	1,409	Include TN single-PBM as treated
Drop Michigan	−0.074	0.244	1,341	MI winsorization sensitivity
Never-treated comparison only	−0.065	0.241	1,373	Drops not-yet-treated cells from control set

Notes: This table documents the source files, scripts, variables, or data inputs used in the analysis. It is included to make the construction of the analytic evidence reproducible.

Every specification point estimate lies within one standard error of the headline −0.065 estimate. Minor numerical differences from values quoted in §6.2 reflect the direct-OLS implementation used in this appendix versus the `differences`-package build used for headline numbers; both implementations produce substantively identical conclusions (precise null across all specifications).

Appendix Table A3. State-by-state treatment panel with vehicle, effective quarter, and source

Source: `analysis/tables/appendix_A3_state_treatment_panel.csv` (script `analysis/15_appendix_tables.py`). All 24 ever-enacting states; main-spec treated = "yes" flags the 20 states used for identification in the 2016–2024 window (dropping VT as all-FFS, TN as single-PBM structure, and CA/CO/ID as out-of-window cohorts).

Abbreviations: vehicle = legislative (statute), contract (MCO agreement), administrative (Medicaid director order / guidance).

State	Vehicle	Effective date	Effective quarter	Cohort year	In window 2016–2024	Main-spec treated	Primary source
GA	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
IA	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
KS	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
LA	legislative	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019; HB 219 (2018)
MI	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
MN	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
MS	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
ND	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
NJ	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
TX	contract	2018-01-01	2018Q1	2018	yes	yes	KFF July 2019
OH	administrative	2019-01-01	2019Q1	2019	yes	yes	Medicaid Director order Aug 2018

State	Vehicle	Effective date	Effective quarter	Cohort year	In window 2016–2024	Main-spec treated	Primary source
NY	legislative	2019-10-01	2019Q4	2019	yes	yes	Enacted 4/1/2019; effective 10/1/2019
AR	legislative	2020-07-24	2020Q3	2020	yes	yes	Act 1103
VA	legislative	2020-10-01	2020Q4	2020	yes	yes	HB 1291 (2020); Va Code § 38.2-3467
KY	legislative	2021-01-01	2021Q1	2021	yes	yes	2020 legislation; single PBM 2021
MD	contract	2021-01-01	2021Q1	2021	yes	yes	MD HealthChoice CY 2021+ MCO agreement
PA	legislative	2021-01-25	2021Q1	2021	yes	yes	Act 120 of 2020 (HB 941)
NC	legislative	2021-07-01	2021Q3	2021	yes	yes	NC SL 2019-240; N.C. Gen. Stat. § 108D
MA	contract	2023-01-01	2023Q1	2023	yes	yes	MassHealth MCO/ACPP contract amendment
FL	legislative	2024-01-01	2024Q1	2024	yes	yes	SB 1550 signed May 2023
VT	legislative	2024-07-01	2024Q3	2024	yes	no (all-FFS; recoded never-treated)	Act 127 / H.233 (2024)
CO	legislative	2025-01-01	2025Q1	2025	no	no (out-of-window)	HB23-1201
ID	legislative	2025-01-01	2025Q1	2025	no	no (out-of-window)	HB 596 (2024)
CA	legislative	2026-01-01	2026Q1	2026	no	no (out-of-window)	SB 41 signed 10/11/2025

Notes: This table documents the source files, scripts, variables, or data inputs used in the analysis. It is included to make the construction of the analytic evidence reproducible.

Tennessee (single-PBM Magellan structure since 2013) is excluded from the main sample; it is coded `spread_ban_enacted = 0` in the panel file and enters the "Tennessee-included" sensitivity (Appendix Table A2).

The 24 / 21 / 20 state-count reconciliation referenced in §4: - 24 states = all ever-enacting rows above - 21 states = in-window (drops CA, CO, ID) - 20 states = main-spec treated (additionally drops VT as all-FFS)

Appendix Table A4. CS-DiD analytic SE vs. multiplier bootstrap SE at event-time cells

Source: `analysis/tables/appendix_A4_csdid_se_comparison.csv` (script `analysis/15_appendix_tables.py`). CS-DiD event-study ATT(k) on log MCO generic spending per enrollee. Analytic SE is the `differences`-package default (doubly-robust influence-function SE). Bootstrap SE is implied from the simultaneous sup-t band: `boot_se_implied = (sup_t_upper − sup_t_lower) / (2 × sup_t_crit)`.

Event time (k)	ATT	Analytic SE	Bootstrap SE	Ratio (boot/analytic)
0	+0.050	0.078	0.078	1.00
1	−0.010	0.085	0.085	1.00
2	−0.064	0.096	0.096	1.00
3	−0.000	0.110	0.110	1.00
4	+0.270	0.365	0.365	1.00
5	+0.277	0.386	0.386	1.00
6	+0.242	0.381	0.381	1.00
7	+0.352	0.381	0.381	1.00
8	−0.000	0.507	0.507	1.00
9	+0.037	0.548	0.548	1.00
10	+0.072	0.576	0.576	1.00
11	−0.060	0.492	0.492	1.00
12	−0.149	0.533	0.533	1.00

Notes: This table documents the source files, scripts, variables, or data inputs used in the analysis. It is included to make the construction of the analytic evidence reproducible.

The analytic and multiplier-bootstrap SEs coincide to three decimal places at every post-period event time because the `differences` implementation uses the multiplier bootstrap internally to build the equicorrelation-approximated simultaneous band; the band is then divided by the bootstrap-derived sup-t critical value ($\rho = 0.3$, $n_{\text{boot}} = 2{,}000$, sup-$t$ critical = 2.81). The analytic cell-level SE and the bootstrap per-cell SE agree because both trace to the same influence-function construction. Pre-period cells have no simultaneous band because the simultaneous band is constructed over the post-period only.

Appendix Figures

Figure A1. CS-DiD vs. TWFE event-study overlay. Source: `analysis/figures/csdid_vs_twfe_overlay.png`.

Figure A2. FFS measurement drill: MCO and FFS DiD coefficients across exclusion thresholds. Source: `analysis/figures/ffs_drill_by_threshold.png`.

Figure A3. HonestDiD relative-magnitudes bounds on the triple-difference event study, full sample and policy-window subsample, for $M \in \{0.5, 1.0, 2.0\}$. Source: `analysis/figures/honestdid_bounds_full.png`, `analysis/figures/honestdid_bounds_drop_2018q1.png`.

Figure A4. Ohio SPBM event study (October 2022 single-state-PBM transition). Source: `analysis/figures/ohio_spbm_event_study.png`.

Figure A5. Medi-Cal Rx event study (January 2022 pharmacy carve-out transition to Magellan). Source: `analysis/figures/medi_cal_rx_event_study.png`.

End of Appendix. Primary manuscript at `manuscript/submission/hsr/manuscript.md`.