Strength in Numbers: Interstate Purchasing Pools and Medicaid Supplemental Rebate Capture

Abstract

Thirty-three states participate in interstate Medicaid drug purchasing pools, yet there is almost no causal evidence on whether pooling actually raises the supplemental rebates states extract from manufacturers. I assemble a primary-source panel of pool membership for all 50 states from 2002 through 2024 — reconciled from state Medicaid manuals, archived web materials, agency correspondence, and public-records requests — and link it to CMS-64 supplemental rebate data and State Drug Utilization Data drug-mix measures. Using estimators that are robust to staggered timing and to pool exit (Sun-Abraham for the absorbing Sovereign States Drug Consortium, or SSDC; de Chaisemartin-D'Haultfœuille for the reversing vendor pools), with wild-cluster-bootstrap inference appropriate to the small number of treated clusters, I find that joining a pool *modestly raises* supplemental rebate capture. Among SSDC later adopters (first treated in 2012 or after), the average post-adoption effect on the supplemental rebate share of drug spending is 3.4 percentage points (wild-cluster $p = 0.051$; it sharpens to 0.03–0.04 under valid outcome-construction covariates), and the estimate is stable across five alternative estimators and leave-one-state-out. The effect is heterogeneous by pool type: the National Medicaid Pooling Initiative (NMPI) and the state-owned SSDC are positive (NMPI 1.9 pp; any-pool 1.2 pp), while The Optimal PDL Solution (TOP$) estimates are centered near zero and rule out effects as large as those estimated for later SSDC adopters, but modest positive or negative effects remain possible. Smaller states gain about twice as much as larger states, consistent with a countervailing-buyer-power mechanism, though the size and switcher cuts are not statistically distinct under small-cluster inference. A coarse state-year test of the class-substitutability prediction is non-confirmatory. The binding constraint on precision is structural: by 2024 most states had pooled, leaving only thirteen never-pooled comparison states, so several confidence intervals hug zero and I bound this imprecision explicitly rather than overclaim. The headline is a measured, mechanism-consistent positive effect, not a clean win — and the governance-controlled specifications are shown to over-condition on post-treatment mediators and to be identified on a restricted post-2010 sample, so the uncontrolled total effect is the primary specification. This study contributes a primary-source multi-pool membership panel as a public good and the first quasi-experimental multi-pool estimates of Medicaid supplemental rebate capture. The paper studies rebate *capture*, not net drug spending; the language of "savings" is reserved for explicit comparisons against net spending.

---

1. Introduction

1.1 The Problem: Rising Medicaid Drug Costs and the Search for Fiscal Levers

Medicaid is the single largest payer for prescription drugs in the United States. In fiscal year 2023, gross Medicaid drug spending exceeded $80 billion, with net spending after rebates reaching roughly $51 billion and continuing to rise as specialty and biologic therapies absorbed a growing share of state budgets (MACPAC, 2022). The trajectory has been relentless: net Medicaid drug spending increased 72 percent between FY2017 and FY2023, outpacing overall Medicaid spending growth and placing escalating pressure on both state and federal budgets. Managing pharmaceutical expenditures is therefore a first-order fiscal challenge for state Medicaid programs.

The urgency of that challenge is compounded by the structure of modern pharmaceutical markets. Specialty and biologic drugs account for roughly 2 percent of prescriptions but nearly 50 percent of drug spending by the early 2020s. Unlike traditional small-molecule drugs, which eventually face generic competition that drives down prices, many biologics face limited biosimilar entry and sustained pricing power. The emergence of cell and gene therapies with per-patient costs exceeding $1 million has introduced entirely new categories of fiscal risk. State Medicaid agencies must simultaneously contain growth in drug spending, maintain compliance with federal coverage rules, and preserve access for populations that rely heavily on public insurance.

State Medicaid programs operate within a mixed federal-state structure that shapes their cost-containment options. At the federal level, the Medicaid Drug Rebate Program (MDRP), established by the Omnibus Budget Reconciliation Act of 1990, requires manufacturers to pay statutory rebates in exchange for Medicaid coverage. Those mandatory rebates create the floor for Medicaid drug pricing. At the state level, however, agencies retain meaningful discretion over several additional levers: preferred drug list (PDL) design, prior authorization rules, managed-care pharmacy architecture, drug utilization review, and supplemental rebate negotiations. This chapter focuses on the last of those levers, and on one increasingly common attempt to strengthen it: banding together with other states to negotiate as a bloc.

1.2 Supplemental Rebates: A Critical and Understudied Fiscal Lever

Supplemental rebates are voluntary rebates negotiated above and beyond the federal MDRP minimum. Manufacturers pay them in exchange for preferred formulary placement, easier utilization management, or broader access to a state's covered population. As of 2019, 47 states and the District of Columbia had supplemental rebate agreements, collectively generating over $1.7 billion annually (OIG, 2014; KFF/HMA, 2020). Manufacturers pay these rebates to ensure preferred placement on the state's PDL, which reduces prior authorization requirements and other utilization management barriers.

Despite that fiscal importance, supplemental rebates have received remarkably little rigorous empirical attention. The Office of Inspector General documented substantial cross-state variation in how states collect and structure supplemental rebates but did not attempt to explain it (OIG, 2014). MACPAC (2018) estimated that supplemental rebates account for roughly 6 percent of total Medicaid drug rebates but did not investigate the drivers of variation. The academic Medicaid pharmacy literature has focused on adjacent policies — managed care versus fee-for-service drug spending, formulary design, prior authorization, generic substitution — leaving supplemental rebates as one of the largest policy-relevant blind spots in the evidence base.

That omission matters because the cross-state variation is large. In 2024, supplemental rebates ranged from near zero in some states to over 28 percent of drug spending in others. If that variation reflects replicable institutional arrangements, the policy implications are substantial. If it instead reflects idiosyncratic characteristics or bundled reforms that cannot be transferred, the lesson is very different. Distinguishing between these explanations requires causal evidence, which to date does not exist.

1.3 Interstate Drug Purchasing Pools: Widespread Adoption, Little Causal Evidence

Many states attempt to increase bargaining leverage by joining interstate drug purchasing pools. The intuitive logic is straightforward: a larger covered population should strengthen the threat to steer utilization toward preferred products, which should in turn raise supplemental rebates. Three major pool families emerged in the mid-2000s: the National Medicaid Pooling Initiative (NMPI, established 2003), The Optimal PDL Solution (TOP$, established 2005), and the Sovereign States Drug Consortium (SSDC, established 2005). By 2024, 33 of 50 states participated in one of those three arrangements.

That widespread adoption is theoretically plausible. Buyer-side market power models suggest that larger, more coordinated purchasers can secure better prices when they can credibly redirect demand (Chown, Dranove, Garthwaite, and Keener, 2019). Structural models of formulary-rebate contracting imply that rebates depend on covered lives and on the buyer's ability to commit to preferred placement (Ho and Lee, 2024). International pooled-procurement research reaches a similar conclusion: pooling can improve procurement outcomes, but the size of the gain depends heavily on governance and implementation (Dubois, Lefouili, and Straub, 2021; Parmaksiz et al., 2022).

Yet despite those strong priors, there is no settled causal evidence on whether interstate Medicaid purchasing pools actually increase supplemental rebate capture. Existing accounts are entirely descriptive: Horvath (2019) documented that Vermont reported 4.7 percent additional savings in its first year of SSDC participation and that New York reported $80.5 million in additional savings through NMPI, but these figures lack counterfactual comparisons. In short, two-

thirds of U.S. states have adopted a policy intervention for which the published evidence base consists almost entirely of anecdotes and self-reported savings figures.

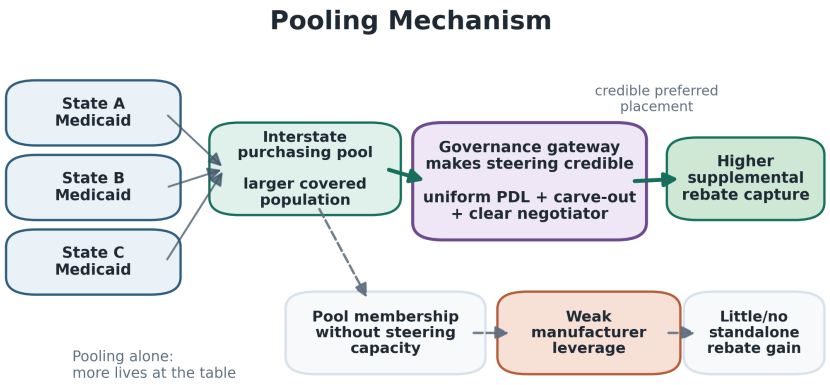

Figure 1: The pooling mechanism. Combining states enlarges the covered population, but higher supplemental rebate capture follows only when the pool can pair that scale with credible volume steering (a unified preferred drug list, a clear negotiator); pooling without steering capacity leaves manufacturer leverage weak.

1.4 This Chapter's Contribution

This chapter makes three contributions.

First, it constructs a 50-state primary-source panel of Medicaid drug-pool membership from 2002 through 2024. No pre-existing public dataset provides a fully reconciled state-year history of SSDC, NMPI, TOP$, and no-pool participation, and the secondary compilations that circulate disagree on the membership and timing of many states. I reconcile the record against primary sources — current and archived state Medicaid manuals, agency correspondence, CMS supplemental-rebate-agreement tables, and public-records requests — and document each coding decision. The panel is released as a public good.

Second, it provides the first quasi-experimental, multi-pool evidence on whether interstate pools raise supplemental rebate capture, using estimators matched to the institutional facts. SSDC membership is absorbing in the panel (no state leaves), so its later-adopter cohorts support a heterogeneity-robust Sun-Abraham (2021) event study. The vendor-administered pools reverse — states enter and exit NMPI and TOP$ — so I estimate them with the de Chaisemartin

and D'Haultfœuille (2020) estimator, which is robust to treatment that switches off. Because the number of treated clusters is small (eight to thirty-seven, depending on the pool), I treat wild-cluster-bootstrap inference as authoritative rather than relying on analytic standard errors. The estimand is uniform across pools — the *total effect* of joining, on the full 2002–2024 sample, relative to never-pooled states — so that no pool receives a different, hand-picked specification.

Third, it maps a countervailing-buyer-power mechanism to testable heterogeneity by pool type and state size, and bounds the estimates honestly given the small number of never-pooled comparison states. Joining a pool raises the supplemental rebate share of drug spending by roughly one to two percentage points on average, and by about three points for SSDC later adopters, with the gain concentrated in the state-owned SSDC and in NMPI and absent for TOP$. The pattern is consistent with a countervailing-buyer-power mechanism: pooling helps where the pool can credibly steer volume and where the residual claimant governs the bargaining. The evidence is consistent with a modest positive effect, but the estimates are imprecise and remain sensitive to modest departures from parallel trends; the imprecision has a structural reason — most states had pooled by 2024 — which I disclose rather than paper over.

The core identification problem is not simply that some states are permanently different from others. It is twofold. First, states often change multiple pharmacy-benefit features — adopting a uniform PDL, carving pharmacy out of managed care, changing who negotiates rebates — at roughly the same time they join a pool. Second, and more basic, "joining a pool" must be measured accurately before any control strategy can help, since a staggered-DiD estimate is a weighted average over cells defined by the treatment dates themselves. This chapter addresses the first problem with a transparent, symmetric control policy (Section 4) and the second by measuring pool membership directly from primary sources (Section 3).

The remainder of the chapter proceeds as follows. Section 2 reviews institutional background and develops the bargaining theory. Section 3 describes the data. Section 4 presents the empirical strategy, including the estimators and inference. Section 5 reports results. Section 6 discusses implications and limitations. Section 7 concludes.

---

2. Background and Theory

2.1 The Medicaid Drug Rebate Program

A short primer on Medicaid pharmacy benefits. Medicaid is the single largest payer for prescription drugs in the United States, but each state's Medicaid pharmacy benefit operates within a layered federal-state architecture that determines how drug spending and rebates flow. At the federal level, the Med-

icaid Drug Rebate Program (MDRP) sets a statutory rebate floor — manufacturers must rebate a minimum share of every Medicaid prescription back to the federal government in exchange for Medicaid coverage of their products. At the state level, state Medicaid agencies decide which drugs to make easier or harder for clinicians to prescribe, primarily through the **Preferred Drug List (PDL)** — a state-published list of which drugs are "preferred" (no prior authorization, easier prescribing) and which are "non-preferred" (requires prior authorization, sometimes a step-therapy fail-first requirement). Manufacturers compete for preferred placement on the PDL by offering **supplemental rebates** above the federal MDRP minimum; preferred placement materially increases a drug's Medicaid market share within its therapeutic class, so manufacturers are willing to pay supplemental rebates to capture it. The state's bargaining leverage in supplemental-rebate negotiations is, mechanically, a function of how many covered lives it can credibly steer with a PDL decision.

That bargaining leverage is shaped by two further architectural choices. **Pharmacy benefit carve-in vs. carve-out** determines whether the Medicaid pharmacy benefit is administered by the state directly (carve-out — the state controls the PDL and rebates for the entire pharmacy benefit, including for managed-care enrollees) or by managed-care organizations on contract with the state (carve-in — MCOs run their own formularies for their enrolled members, with statewide coordination varying widely). **Uniform PDL** policies require all MCOs to use the state's PDL rather than negotiate their own — converting a fragmented set of MCO formularies back into a single statewide formulary decision. A state can be in any of four corners of the (uniform PDL by carve-out) space, and the size of its credible bargaining unit varies accordingly. **Interstate purchasing pools** — the policy lever this paper studies — sit on top of all of this: they combine multiple states' covered lives into a single negotiating unit. Whether pool membership translates into higher supplemental rebate capture should depend on the state's ability to translate pooled covered lives into a coordinated PDL decision, which in turn depends on the carve-out and uniform-PDL architecture.

The MDRP statute. The MDRP was established by the Omnibus Budget Reconciliation Act of 1990 and has been a central feature of Medicaid pharmaceutical policy for more than three decades. Manufacturers must enter into rebate agreements with the Secretary of Health and Human Services as a condition of Medicaid coverage. For brand-name drugs, the federal unit rebate amount is the greater of 23.1 percent of Average Manufacturer Price (AMP) or the difference between AMP and the manufacturer's best price — the lowest price offered to any commercial purchaser, with certain exclusions — plus an inflation penalty when price growth exceeds CPI-U. For generics, the statutory minimum is 13 percent of AMP.

This pricing structure creates well-documented incentive distortions. Duggan and Scott Morton (2006) provided foundational evidence, demonstrating that manufacturers respond to the best-price provision by strategically raising prices

to commercial purchasers. Their estimates indicate that a 10-percentage-point increase in Medicaid market share is associated with 7–10 percent higher average drug prices across all purchasers. This cross-subsidization means that the MDRP's mandatory rebates are partially offset by higher commercial prices.

The best-price provision has additional consequences for supplemental rebate negotiations. Because manufacturers must offer Medicaid a rebate at least as large as the AMP-best-price spread, any price concessions to large commercial purchasers automatically increase the statutory rebate. This creates a floor below which supplemental negotiations operate: the supplemental rebate is an increment above an already-deep statutory discount, making the marginal dollar obtained through supplemental negotiation more difficult to extract. The implication for this paper is that a few-percentage-point supplemental gain is economically meaningful, not trivial — it is extracted on top of a binding federal floor.

The Affordable Care Act intensified the MDRP's role in 2010 through two changes relevant to this analysis. First, it raised the minimum brand-name rebate from 15.1 to 23.1 percent of AMP, substantially raising the statutory floor. Second, it extended mandatory MDRP rebates to drugs dispensed through Medicaid managed care, closing a loophole that had allowed MCOs to negotiate outside the MDRP framework. After 2010, supplemental rebates could be layered on top of statutory rebates for both fee-for-service and MCO-dispensed drugs, increasing the potential scope of supplemental rebate programs but also adding administrative complexity.

2.2 Supplemental Rebates and Preferred Drug Lists

Supplemental rebates are voluntary payments that manufacturers make in exchange for preferred formulary placement. They are embedded in formulary design, prior authorization policy, and the state's ability to steer utilization toward preferred products.

The economics of supplemental rebates are best understood through the lens of formulary contracting theory. Ho and Lee (2024) developed a structural model of formulary-rebate contracting between PBMs and manufacturers, showing that multi-tier formularies substantially increase negotiated rebate payments. In their model, rebate contracts specify per-unit payments contingent on formulary placement and competitor positioning. A manufacturer whose drug is placed on the preferred tier gains market share within its therapeutic class, while the payer captures a rebate that partly offsets coverage costs. The size of the rebate depends on the manufacturer's outside option (losing preferred status) and the payer's outside option (switching to an alternative). That logic implies that supplemental rebate negotiation is strongest in therapeutic classes with credible alternatives and weakest in sole-source or minimally substitutable products — a class-level prediction I return to and test, coarsely, in Section 5.

Chown, Dranove, Garthwaite, and Keener (2019) frame this intuition in their

analysis of buyer-side market power in health care. Comparing the United States and Canada, they argue that prescription-drug price differences more directly reflect public buyer power than do provider-wage differences. Their analysis supports the broader premise that purchasing scale matters only when the buyer can translate that scale into credible demand control.

Preferred drug lists are the operational mechanism through which that leverage is exercised. Virabhak and Shinogle (2005) documented large prescribing responses to PDL implementation: restricted-drug prescription shares fell by 9.0 percentage points in Illinois Medicaid and 6.2 percentage points in Louisiana Medicaid after PDL adoption. Munshi et al. (2018) found that Florida's state-mandated PDL reduced overall drug use by 6 percent but increased plan costs by 27 percent — an apparently paradoxical result attributable to the PDL's channeling of utilization toward preferred branded drugs with high acquisition costs but also high rebate yields. This finding illustrates that PDL effects on spending are not straightforward: a PDL that maximizes supplemental rebate revenue may not minimize net drug costs if it channels utilization away from lower-cost generics. It also clarifies why this paper's outcome is rebate *capture* and not net spending.

States vary substantially in their PDL architecture. Some maintain a uniform PDL spanning both fee-for-service and managed-care enrollees, giving the state direct control over formulary placement for its entire Medicaid population. Others allow MCOs to maintain their own formularies, fragmenting the state's bargaining leverage by splitting the covered population across multiple formulary decisions. A state with a uniform PDL can credibly commit to placement decisions for a larger share of its covered lives, which should increase bargaining power. Conversely, a state may join a pool but still fail to convert pooled covered lives into real negotiating power if managed care plans retain substantial formulary autonomy. This is the sense in which governance is not a nuisance covariate but part of the mechanism.

2.3 Interstate Drug Purchasing Pools

Three major pool families operate during the panel, and they differ in ways that matter for both identification and interpretation.

NMPI negotiates collective supplemental rebate agreements (SRAs) on behalf of member states. The pool entity, administered by Magellan/Prime Therapeutics, contracts directly with manufacturers, while member states choose which negotiated drugs to place on their own formularies. The pool — not the individual state — owns the contract paper, and participation is tied to the state's controlling services agreement with the vendor. By 2024, 11 states participated, including several large states such as New York and Michigan.

**TOP$** negotiates collective SRAs similarly to NMPI and is likewise vendor-administered (Prime/Provider Synergies). By 2024, 7 states participated. The relative obscurity of TOP$'s operations reflects a broader challenge in study-

ing Medicaid purchasing pools: much of the relevant institutional detail is not publicly documented.

SSDC is structurally distinct from the other two in a critical way: member states contract *individually* with manufacturers using their own state-specific SRAs, while leveraging the collective purchasing power of the consortium's combined covered lives. According to its own materials, SSDC is "the only Medicaid rebate pool organized and managed by the states"; member states join through a consortium agreement (cost-shared through Vermont, the managing member), do not need to contract with the consortium's hired agent (Optum) to participate, and retain full ownership of their contracts and PDLs. The consortium runs a formal annual cycle in which its agent solicits offers, member states meet to review negotiated offers, set priorities, and make final acceptance decisions. By 2024, 15 states participated.

This structural distinction — individual contracting with pooled leverage, versus collective contracting administered by a vendor — is the basis for the pool-type heterogeneity this paper tests. The next subsection develops why it should matter.

2.4 Theory: Pooling, Countervailing Power, and Steerable Volume

The animating idea is that pooling raises the rebate a state can extract by **worsening the manufacturer's no-deal payoff** — but only to the extent the pool can credibly steer volume. A clean way to see this is through Nash bargaining over preferred formulary placement.

A compact bargaining model. Consider a manufacturer negotiating with a state (or pool) over preferred status for its drug j in some therapeutic class. The value to the manufacturer of winning preferred status in state i is

$$V_{ij} = \Delta q_{ij} \times m_j,$$

the *steerable volume* Δq_{ij} that preferred status moves toward the drug, times the manufacturer's margin m_j on that volume. A single state bargaining alone splits V_{ij} with the manufacturer according to their relative bargaining weights; the supplemental rebate is the manufacturer's payment for avoiding the disagreement outcome (non-preferred placement and lost volume). When states pool, the manufacturer instead bargains against a combined prize

$$V_{\text{pool},j} = \sum_i \Delta q_{ij} \times m_j,$$

so the disagreement loss from failing to reach a deal — losing preferred placement across the *whole bloc* — is larger, and the equilibrium rebate rises. Three implications follow directly and structure the empirical work.

1. **Size alone is not enough.** The prize aggregates Δq, not covered lives. A pool that assembles nominal enrollment but cannot coordinate formulary action (because PDL decisions are fragmented across MCOs, or the pharmacy benefit is carved into managed care with autonomous formularies) raises V_{pool} only weakly. Pooling and *steerable* volume are complements. This is why governance (uniform PDLs, carve-outs, MCO rebate collection) is part of the mechanism, not a confound to be partialled out indiscriminately. Figure 2 sketches the governance levers — the within-state arrangements that concentrate formulary leverage into a single, credible volume signal that a pool can put behind its negotiation.

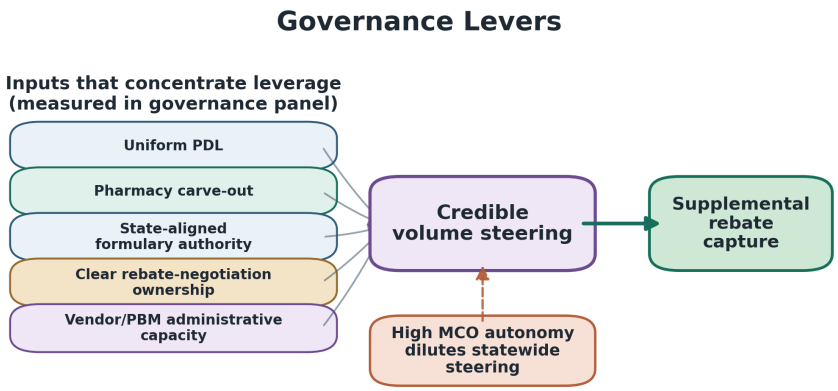

Figure 2: Governance levers. The within-state pharmacy-benefit arrangements measured in the governance panel — a uniform preferred drug list, a pharmacy carve-out, state-aligned formulary authority, clear rebate-negotiation ownership, and vendor/PBM capacity — concentrate fragmented formulary control into a single credible volume signal; high managed-care formulary autonomy dilutes it.

2. **The effect is larger for smaller states.** A small state bargaining alone has a small V_{ij} and little leverage; pooling relaxes that constraint most where it binds hardest. The model therefore predicts a *larger* percentage-point gain for smaller states — a testable heterogeneity prediction.

3. **A formulary-position auction.** Pooling enlarges the prize for preferred status among substitutable brands, which should make manufacturers bid more aggressively. Where close therapeutic substitutes exist, the threat of non-preferred placement is credible and valuable; for protected or sole-source classes, there is no rival to steer toward, so pooling should help less. This yields a class-level substitutability prediction.

Why pool type should matter. Pools differ in how credibly they convert size into a disagreement threat, and in whose objective governs the bargaining. SSDC is **state-owned**: the residual claimants — the state Medicaid programs that keep the rebate savings — retain control rights over acceptance decisions, contract terms, and strategic priorities. Property-rights logic predicts higher bargaining effort and better adaptation when the party that bears the budget consequence also controls the decisions. SSDC also preserves state-specific SRAs, which let manufacturers offer deeper concessions to states that can credibly move a class without extending the same terms to states that cannot — solving a "lowest-common-denominator" problem that a single uniform vendor contract can create. And its annual member-state meeting builds a repeated club game: manufacturers face the same coordinated buyers year after year, raising the expected cost of lowball offers. Vendor-administered pools (NMPI, TOP$) rely on an agent whose objective — client retention, administrative feasibility, manufacturer relationships, standardized operations — need not be a high-powered function of each incremental rebate dollar. This is an incentive-alignment hypothesis, not an accusation of shirking (Prime states that it passes through 100 percent of rebates), and it predicts SSDC at least matches the vendor pools.

A necessary selection caveat. The same features that should make SSDC more effective also make SSDC states *select* differently. SSDC's membership skews toward states with unified FFS-MCO PDLs and MCO rebate collection, and includes several large recent joiners (Ohio, Oklahoma, Kentucky, Pennsylvania). Part of a measured "SSDC effect" may therefore reflect a correlated bundle of reforms — unified PDLs, MCO rebate collection — rather than pool governance in isolation. I read the SSDC estimate as the effect of SSDC *as implemented* (pool plus the reforms that travel with it), and I am explicit that I cannot fully separate governance from selection with this design. This is exactly why the control policy below is symmetric and why the governance-controlled specification is a controlled residual association, not the headline.

A logic model for the causal chain. It helps to lay out the implied causal graph explicitly. The chain runs from a state's pre-treatment characteristics, to the decision to enter a pool, through a set of post-treatment governance and contracting changes, to the measured rebate outcome. Pre-treatment confounders sit upstream of pool entry and are *valid to adjust for if measured at baseline*: baseline rebate capture, Medicaid program size, managed-care penetration, fiscal stress, administrative capacity, drug mix, and the political environment all plausibly drive both who pools and how much rebate a state captures. The treatment node is pool entry itself. Downstream of pool entry sit the *mediators* — a unified PDL, a pharmacy carve-out, expansion of MCO rebate collection, formulary steering, and vendor negotiation — which are the channels through which pooling is supposed to work and which must *not* be conditioned on as if they were confounders. The outcomes are the supplemental rebate share (measured here) and, further downstream and not measured in this chapter, net drug spending and utilization or access.

This classification is what makes some "controls" inappropriate and others valid. Adjusting for a baseline characteristic that predates pool entry removes a genuine confounder; adjusting for a post-treatment mediator opens a bad-control problem that can attenuate or even flip a real effect, because the mediator is itself an outcome of the treatment. The same graph also exposes an interpretation hazard the outcome cannot resolve on its own: a rise in the measured rebate share can come from genuinely greater *bargaining power* (the mechanism of interest) or from improved *administrative and reporting capacity* — for example, a state beginning to collect and report supplemental rebates on managed-care utilization that it previously did not capture. These are different mechanisms that move the same numerator, and a share-based outcome alone cannot distinguish them.

2.5 International and Pharmacy-Benefit Evidence

International pooled-procurement evidence reinforces the steerable-volume logic. Dubois, Lefouili, and Straub (2021) study centralized procurement across seven low- and middle-income countries and find that pooled public procurement is associated with lower drug prices, with smaller reductions when the supply side is more concentrated. Parmaksiz et al. (2022), in a systematic review of 44 empirical studies, similarly emphasize the capacities and alignment needed among buyers, pooled-procurement organizations, and suppliers; savings range widely, suggesting that pool design features — not pooling per se — drive outcomes. The lesson maps onto Medicaid: scale matters less when the buyer cannot credibly move volume.

The interaction between pharmacy-benefit design and supplemental rebate capture is also well grounded. Dranove, Ody, and Starc (2021) provide the most rigorous causal study of Medicaid pharmacy-benefit privatization, finding that full privatization decreased drug spending by 22.4 percent, with effects concentrated in states giving MCOs formulary flexibility — implying that MCO formulary autonomy can either complement or undermine pool-negotiated PDLs. Bendicksen and Kesselheim (2022) document a countervailing trend of pharmacy- benefit carve-outs. Hernandez and Gellad (2020) highlight that fee-for-service programs tend to retain expensive branded drugs to capture large inflation rebates, while MCOs adopt generics at much higher rates — relevant because pool-negotiated PDLs may favor branded drugs with large supplemental rebates over lower-cost generics. These reforms move with pool entry and are the reason a naïve pool coefficient must be interpreted carefully — but, as Section 4 argues, they are as much mechanism as confound.

---

3. Data

This section describes the dataset, which links a corrected pool-membership panel and a pharmacy-benefit panel to CMS-64 rebate data and SDUD-based

drug-mix measures.

3.1 CMS-64 Financial Management Reports

The primary outcome is annual supplemental rebate capture, defined as:

> Supplemental Rebate Share = Supplemental Rebate Dollars / Total Drug Expenditures

Both numerator and denominator come from CMS-64 Financial Management Reports. I use a share rather than a dollar outcome because states differ enormously in program size; the share-based outcome asks whether a state is extracting larger supplemental rebates relative to the scale of its own drug program.

Constructing a consistent CMS-64 panel requires harmonization across multiple reporting formats and line-item conventions. The raw data files span multiple formats: one master file covers fiscal years 2002–2011, while individual annual files cover later years. I compiled all years into a master file with standardized variable names and consistent state identifiers. A total of 39 state-year observations (3.4 percent of the panel) have missing supplemental rebate data, left as missing rather than imputed. Six state-years exhibit negative supplemental rebate values, which is conceptually difficult to reconcile with voluntary rebate payments; four are negligible in magnitude, while North Carolina in 2013 (−8.1 percent) and Virginia in 2012 (−3.0 percent) are larger and likely reflect timing or reporting artifacts. These observations are retained without adjustment; excluding them does not materially affect results.

After standardization, the rebate panel contains 1,150 state-years from 2002 through 2024, with 1,111 nonmissing supplemental rebate observations. The CMS-64 reports on a fiscal-year basis, which may not align perfectly with calendar-year pool membership indicators. Because pool membership changes typically take effect at the beginning of a state fiscal year and supplemental rebate agreements cover annual or multi-year periods, this misalignment is unlikely to introduce systematic bias.

3.2 State Drug Utilization Data

State Drug Utilization Data (SDUD) enter the chapter in two ways. First, they serve as an external validation source for the CMS-64 outcome. Total drug spending figures are cross-referenced against CMS-64 to identify states where the two sources diverge substantially. Discrepancies are expected due to differences in reporting scope (SDUD covers outpatient pharmacy only; CMS-64 may include physician-administered drugs), timing (calendar-year versus fiscal-year), and varying inclusion of managed care drug spending.

Second, SDUD provides state-year drug-mix measures used in outcome-construction robustness and in the class-substitutability heterogeneity test. The SDUD mix panel summarizes each state-year into several descriptive measures: single-source spend share, innovator-multiple-source spend share,

average spend per prescription, NDC-level concentration, and the share of spending accounted for by the top ten NDCs. The category crosswalk's matched spending share rises from 26.6 percent in 2002 to 99.5 percent in 2024, averaging 79.4 percent across 2010–2024, so these measures are most informative in the later part of the panel.

3.3 Pool Membership Data

The first policy panel documents pool membership for all 50 states from 2002 through 2024. No pre-existing public dataset provides a fully reconciled state-year history of SSDC, NMPI, TOP$, and no-pool participation, and the secondary compilations that exist disagree on the membership and entry/exit timing of many states.

I constructed the panel through three complementary approaches: systematic scraping of state Medicaid agency websites, including archived historical content accessed through the Wayback Machine; direct contact with each state's MDRP contacts via phone and email; and FOIA requests or state Public Records Requests for states with incomplete responses. The resulting timeline was validated against KFF State Health Facts, NASHP (Horvath, 2019), OIG (2014), and CMS supplemental-rebate-agreement tables.

Each contested state was resolved against archived contemporaneous sources; one state (Tennessee, a comparison-group member immaterial to the headline) remains provisional pending an agency reply. Because a staggered-DiD coefficient is a weighted average over event-time cells defined by the treatment dates, accurate entry/exit dating is essential, and the crosswalk and a full state-by-state decision log are released with the panel so that every coding choice is auditable.

Pool indicators are mutually exclusive within a state-year: each state-year is coded SSDC, NMPI, TOP$, or NoPool. Five states switch from a vendor pool into SSDC during the panel (Delaware, Kentucky, Vermont, West Virginia, and — after a no-pool gap — Pennsylvania), and SSDC itself is absorbing (no state exits), while NMPI and TOP$ membership reverses for several states.

3.4 Pharmacy-Benefit Structure Panel

A second state-year policy panel measures pharmacy-benefit structure directly. It contains 747 coded state-year spells across all 50 states and tracks seven features: uniform PDL status, pharmacy carve-out status, MCO formulary autonomy, PDL vendor, PBM vendor, rebate-negotiation model, and major reform flags. These are the features through which a pool's bargaining power is — or is not — converted into steerable volume, and they enter the analysis as a uniform sensitivity/mechanism check rather than as headline controls (Section 4).

The coding strategy is intentionally conservative and built at the spell level. Each spell records a start year, end year, source basis, certainty rating, and

coder notes; many spell boundaries are supported by official current manuals, archived state materials, and managed-care documents, while some remain medium-certainty backcasts from later official manuals rather than fully contemporaneous archived sources. A crucial limitation, central to the control policy below, is timing: the governance variables are densely coded only from 2010 onward (roughly 2 percent non-missing before 2010, 92–100 percent from 2010). For the many pre-2010 pool entries, the panel cannot observe whether a governance reform preceded or followed pool entry.

3.5 Treatment Timing and Analytic Samples

SSDC adoption timing is central to the design. The fifteen SSDC adopters enter in eleven cohorts:

Cohort	States	Cohort	States
2006	Iowa, Maine, Vermont	2016	Delaware
2007	Utah	2017	Ohio, Oklahoma
2008	West Virginia, Wyoming	2020	Pennsylvania
2009	Oregon	2023	South Dakota
2012	Mississippi	2024	Kentucky
2015	North Dakota		

The six early adopters in 2006–2009 are already treated at the start of the modern post-2010 window, so they cannot contribute untreated pre-period support in a later-adopter event study. The lead dynamic sample consequently focuses on the **eight later adopters first treated in 2012 or later** (Mississippi, North Dakota, Delaware, Ohio, Oklahoma, Pennsylvania, South Dakota, Kentucky), compared against the **thirteen never-in-any-pool states**. Those thirteen comparison states are Alabama, Arizona, California, Colorado, Florida, Illinois, Indiana, Kansas, Massachusetts, Missouri, New Jersey, New Mexico, and Texas. They span all four Census regions and include four of the five most populous states (California, Texas, Florida, and Illinois) alongside small programs, so the counterfactual is a broad cross-section rather than an outlier fringe. This later-adopter restriction — not an asymmetric set of controls — is how the design handles the early cohorts' truncated pre-periods, and in the later-adopter sample the observed pre-treatment coefficients are small and statistically indistinguishable from zero (Section 5). For the reversing pools I use all treated states on the full sample. The merged panel is 1,150 state-years across 50 states, with 1,111 nonmissing outcomes.

Early adopters (2006–2009) were predominantly smaller, rural states with limited internal pharmaceutical negotiating capacity. The 2012–2017 wave brought a mix of small and medium-sized states, including Ohio and Oklahoma. The most recent adopters include Pennsylvania, one of the largest state Medicaid programs. This variation in adopter characteristics matters for interpreting

generalizability.

3.6 Descriptive Statistics

The merged dataset contains 1,150 state-years, with 1,111 nonmissing supplemental rebate observations. Table 1 reports the supplemental rebate share by corrected pool family.

Table 1. Supplemental Rebate Share by Pool Family

Pool family	Mean	Median	SD	Min	Max	State-years
SSDC	5.32%	4.83%	4.28%	0.00%	28.40%	181
TOP$	3.52%	3.61%	2.56%	−0.08%	12.35%	149
NMPI	3.11%	2.20%	3.56%	−8.11%	28.24%	202
NoPool	1.77%	0.16%	2.76%	−2.98%	21.83%	579

SSDC state-years exhibit the highest supplemental rebate percentages; the SSDC–no-pool gap is approximately 3.6 percentage points at the mean and 4.7 at the median. NMPI and TOP$ state-years sit between SSDC and no-pool states. The large mean–median gap for no-pool states (1.77 vs. 0.16 percent) reflects a right-skewed distribution with many states reporting near-zero supplemental rebates. These raw gaps are the motivation, not the finding: SSDC state-years also differ in governance and drug-mix profiles, which the design addresses through within-state staggered comparisons against never-pooled states and through the robustness checks below.

Table 2. Drug-Mix Descriptives by Pool Family (SDUD, 2010–2024)

Pool family	Single-source share	Innovator multi-source share	Avg. spend/Rx	Top-10 NDC share
NoPool	42.1%	21.6%	$94.17	15.5%
SSDC	50.1%	21.4%	$93.37	18.9%
NMPI	46.8%	20.5%	$96.21	17.3%
TOP$	43.4%	25.0%	$94.64	16.6%

Drug baskets vary modestly across pool families — SSDC states carry a somewhat higher single-source (protected-brand) spend share — motivating the SDUD-based composition checks and the class-substitutability test in Section 5.

Because the identifying contrast is later SSDC adopters versus never-pooled states, Table 2b reports baseline (2002–2005) characteristics by group. Two features matter for interpretation. First, the eight later adopters begin with

low baseline rebate capture (0.30 percent of drug spending) — below the never-pooled states (1.29 percent) and the early SSDC adopters (1.26 percent) — the negative-selection pattern examined in Section 5.7. Second, the never-pooled comparison states are on average somewhat *larger* (mean log drug spending 20.5 versus 19.6 for later adopters), because they include four of the five most populous states; state fixed effects absorb fixed size differences, but the imbalance is one reason the later-adopter event study and the leave-one-comparison-state-out check (Section 5.4) matter. Baseline drug-mix shares are similar across groups, though the 2002–2005 SDUD crosswalk is thin (matched spend rises from 27 percent in 2002 to near-complete by the 2020s), so drug-mix balance is more reliably read on the post-2010 panel (Table 2).

Table 2b. Baseline (2002–2005) characteristics by group

Group	States	Supp. rebate (pp)	Mean log drug spend
Never-pooled	13	1.29	20.49
SSDC early (2006–2009)	7	1.26	18.83
Vendor (NMPI/TOP$)	22	0.38	19.89
SSDC later (2012+)	8	0.30	19.58

Source: `analysis/runs/phase2/phase2e_baseline_balance.csv`.

---

4. Empirical Strategy

4.1 Estimands

Before describing the estimators, I fix the quantities they target. The design recovers four distinct estimands, and all interpretation in this and later sections refers to these:

- **SSDC later-adopter ATT** — the average treatment effect on the treated of SSDC adoption for states first adopting in 2012 or later, relative to never-in-any-pool states. This is the headline estimand.
- **Vendor-pool ATT** — the average effect of NMPI, TOP$, or Prime (NMPI or TOP$) status, allowing entry and exit, relative to never-pooled states.
- **Any-pool ATT** — the average effect of being in any pool, combining the several institutional arrangements into a single treatment.
- **Policy-bundle estimand** — the effect of pool participation *as implemented*, inclusive of co-traveling governance changes the design cannot separate from pool entry. This is what every estimate above recovers when pool entry is co-timed with governance reform.

Stated up front rather than as an aside: the SSDC all-treated estimate is null, and the credible SSDC estimate pertains to later adopters, not to the early

SSDC cohorts. The headline SSDC number is therefore the later-adopter ATT, not an "SSDC effect" averaged over all member states.

4.2 Identification and a Symmetric Control Policy

The causal question is whether interstate pool participation increases supplemental rebate capture. The identification challenge has two parts. First, selection: pool adopters may differ from non-adopters on trends, which the staggered design addresses by comparing newly treated states to never-pooled states within the same years and by inspecting pre-trends. Second, bundling: pool entry can coincide with within-state pharmacy-benefit reforms (uniform PDL, carve-out, vendor or negotiation-model changes) that also move supplemental rebates.

A natural instinct is to "control away" the bundle. I argue, and the data confirm, that doing so asymmetrically — controls for one pool but not others, or treating the controlled estimate as the headline — is a specification trap. The control policy here is therefore **symmetric and pre-specified: the uncontrolled total effect is primary for every pool, and the governance-controlled specification is a uniform sensitivity specification that absorbs observed contemporaneous pharmacy-benefit structure, never anyone's preferred estimate.** Because these variables may be mediators, these estimates should not be interpreted as the causal direct effect of pooling. Three reasons:

1. **The governance variables are part-confounder, part-mediator.** Some states adopt a uniform PDL or carve-out *before* joining a pool (there it is a pre-treatment confounder); others adopt it concurrently or after (there it is a post-treatment mediator — pooling works *through* steerable volume and unified PDLs, as the theory in Section 2.4 makes explicit). Conditioning on a mediator induces bad-control bias and can attenuate or flip a real effect.

2. **The governance variables are only observed from 2010.** The controlled specification is identified almost entirely on the post-2010 window and drops roughly one-third of the sample, making it a restricted-sample estimate, not a full-panel causal upgrade.

3. **SSDC's early-cohort pre-trend is handled by the sample, not by controls.** The later-adopter cut (justified because the 2002 panel start truncates the early cohorts' pre-periods) is clean *uncontrolled* — clean de Chaisemartin-D'Haultfœuille placebos and a 3–4 pp effect. SSDC therefore needs no special control treatment; it follows the same uncontrolled rule as every other pool.

I therefore report the uncontrolled total effect (full 2002–2024 sample) as the agnostic primary for all five pool definitions, and the governance-controlled specification as a uniform controlled residual association — the residual after conditioning on *observed* governance — whose empirical sign-instability and inflated

confidence intervals reinforce that it is not headline-grade. I do not assert "governance is a mediator" as established fact; I present the uncontrolled (total) estimate and the controlled (residual) sensitivity specification side by side and let the theory and the timing limitation explain why the uncontrolled estimate is the more credible one.

4.3 SSDC Later-Adopter Event Study

I follow the heterogeneity-robust staggered-DiD literature and match the estimator to each pool's institutional facts. No state leaves SSDC over the panel, so SSDC adoption is an absorbing treatment. For the lead dynamic case I use the Sun-Abraham (2021) interaction-weighted estimator on the later-adopter cohorts, with never-pooled states as the clean comparison group. I corroborate the average post-adoption effect with Callaway and Sant'Anna (2021) group-time ATT, Borusyak, Jaravel, and Spiess (2024) imputation, a fully saturated event study, and the de Chaisemartin-D'Haultfœuille estimator, reporting the spread across implementations rather than a single number.

For the lead dynamic case, the Sun-Abraham interaction-weighted estimating equation is

$$Y_{st} = \alpha_s + \gamma_t + \sum_k \beta_k \, \mathbf{1}[k_{st} = k] \cdot \mathbf{1}[\text{later-treated}_s] + \varepsilon_{st},$$

where Y_{st} is the supplemental rebate share for state s in year t; α_s and γ_t are state and year fixed effects; event time $k_{st} = t - g_s$, with g_s the first SSDC year; $k = -1$ is the omitted reference; and standard errors are clustered at the state level, with wild-cluster-bootstrap p-values and confidence intervals as the authoritative inference. I use only the eight SSDC states first treated in 2012 or later together with never-pooled comparison states, and I report the average post-adoption effect over event times 0 to 8 as the summary parameter, because later event times are identified by three or fewer states (and event times of 11 or more by a single state) and should not be read as dynamic structure.

Why effects may arrive slowly. Even if pooling is genuinely effective, a large same-year jump in supplemental rebate capture is unlikely. States renegotiate class by class through contract renewals, PDL reviews, and prior authorization updates, implying a one-to-three-year ramp. A front-loaded null therefore does not refute pooling, and a gradual path is consistent with — but not uniquely diagnostic of — a real pool effect. The design reports the average post-adoption effect rather than leaning on any single lag.

4.4 Multi-Pool Effects with an Exit-Robust Estimator

States enter and exit NMPI and TOP$, so these treatments switch off — a setting in which Sun-Abraham and other absorbing-treatment estimators are not

valid. For NMPI, TOP$, the combined Prime pools (NMPI or TOP$), and any-pool, I estimate the de Chaisemartin and D'Haultfœuille (2020) dynamic effect (`DIDmultiplegtDYN`) on the full sample against never-pooled states, reporting the average effect, its wild-cluster/bootstrapped confidence interval, and the placebo (pre-trend) estimates. This estimator accommodates the entry-and-exit pattern of the vendor pools, which absorbing-treatment estimators cannot handle.

Inference. The number of treated clusters is small — eight SSDC later adopters, and between roughly twenty and thirty-seven treated states for the broader pools — so cluster-robust analytic standard errors are unreliable and over-reject. I treat the wild-cluster bootstrap (Cameron, Gelbach, and Miller, 2008; MacKinnon and Webb, 2017), with Rademacher weights and 9,999 replications clustered on state, as the authoritative inference throughout. Where I report an analytic p-value, I flag that it overstates significance. Two-way fixed effects, where shown, is a benchmark only and is diagnosed with the Goodman-Bacon (2021) decomposition. Sensitivity to deviations from parallel trends is assessed with the Rambachan and Roth (2023) honest-bounds framework.

4.5 Governance-Controlled Sensitivity Specification and Robustness

I re-estimate each design after adding the time-varying pharmacy-benefit controls (uniform PDL, pharmacy carve-out, major reform flag) as the uniform sensitivity specification described in Section 4.2. I additionally assess: (i) cross-estimator agreement (Sun-Abraham, Callaway-Sant'Anna, Borusyak-Jaravel-Spiess, saturated, de Chaisemartin-D'Haultfœuille); (ii) leave-one-state-out; (iii) outlier dropping and 1/99 winsorization; (iv) outcome-construction covariates that are *valid* (not mediators) — the national rebate "headroom" available given the federal floor, and SDUD drug-mix shares; (v) Rambachan-Roth honest sensitivity; and (vi) theory-motivated heterogeneity by state size, switcher status, pool type, and a coarse class-substitutability proxy. The estimand, sample, treatment definition, fixed effects, clustering, and planned robustness are locked in a specification registry, and the headline estimate is reproduced in an independent implementation.

---

5. Results

5.1 Descriptive Trends and Timing Support

Figure 3 plots mean supplemental rebate capture over calendar time by pool family. SSDC states trend above no-pool states throughout, and the gap widens after the post-2020 general rise in supplemental rebates; NMPI and TOP$ track between. Figure 4 shows the SSDC treatment-timing distribution — the staggered entry across 2006–2024 and the thin support at late event times that motivates truncating the dynamic path at event time 8.

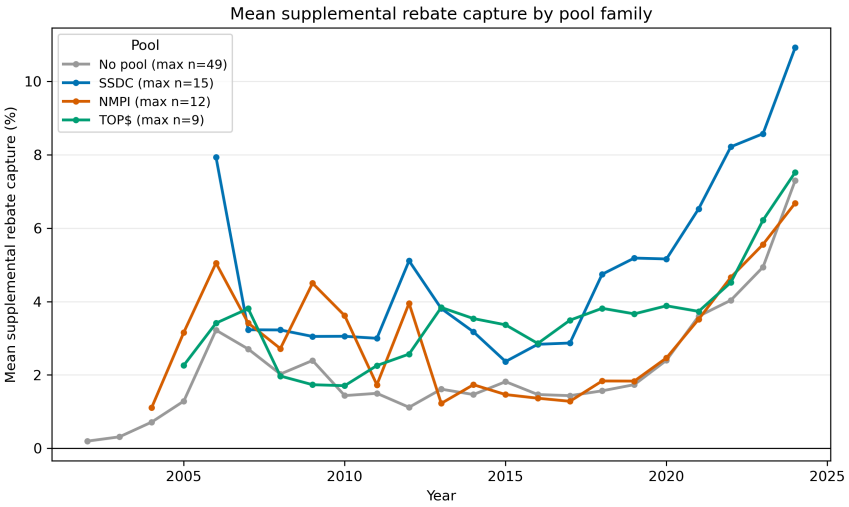

Figure 3: Supplemental rebate trends by pool family

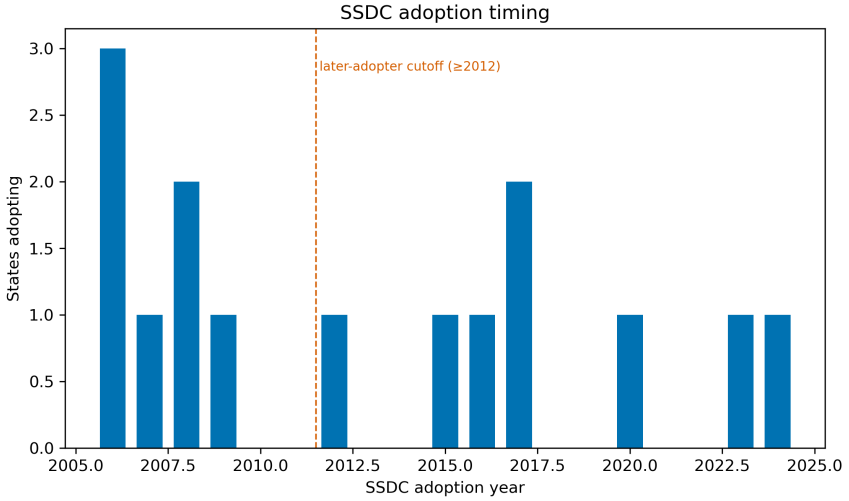

Figure 4: SSDC treatment timing

5.2 SSDC Later-Adopter Benchmark

Table 3 reports the main SSDC result. In the later-adopter sample (first treated 2012 or later), the average post-adoption effect on supplemental rebate share is **3.38 percentage points** (Sun-Abraham), with a wild-cluster-bootstrap p-value of 0.051 — just above the conventional 0.05 threshold, which I report exactly rather than round down. The effect is economically meaningful against a no-pool base of under 2 percent and a binding federal rebate floor.

Table 3. SSDC later-adopter effect and robustness (uncontrolled)

Check	Result
Headline (Sun-Abraham, later, uncontrolled)	**3.38 pp**, WCB $p = 0.051$
Cross-estimator	SA 3.38 / CS 2.66 / BJS 4.27 / dCDH ~2 to 3.8 / saturated 3.66 (all positive)
Leave-one-state-out	[2.91, 3.88] pp — not driven by one state
Outliers / winsorize 1/99	3.07–3.38 pp
Outcome-construction covariates (headroom, drug-mix, post-federal denom.)	3.6–4.0 pp, WCB **0.03–0.04** (sharpens)
Rambachan-Roth honest bounds	survives at $\bar{M} = 0$ [1.11, 5.49]; breaks under modest trend violations
Governance-controlled (sensitivity specification only)	sign-unstable / imprecise → not headline-grade

Two features deserve emphasis. The estimate is *stable* across five estimators and across leave-one-out, so it is not a single-state or single-package artifact. And it *sharpens* — does not attenuate — when the valid outcome-construction covariates (which adjust for what a state could mechanically capture given the federal floor and its drug mix) are added. The governance- controlled specification, by contrast, is sign-unstable and wide, exactly as the part-mediator/post-2010-only critique in Section 4.2 predicts; it is reported as a sensitivity specification, not a headline.

Which SSDC estimand. The headline is the *later-adopter* ATT, and it is worth showing why, side by side (Table 3b). The all-treated SSDC estimate — pooling the six early adopters first treated in 2006–2009 with the eight later adopters — is a null 0.6 pp. That is not evidence against an effect: the early cohorts are already treated when the 2002 panel opens, so they contribute no clean pre-period and their truncated event-time cells pull the pooled estimate toward zero. Restricting to later adopters, who have observed pre-periods, gives 3.38 pp; extending the outcome panel back to 1997 (Appendix B) gives 3.34 pp

on essentially the same sample. The credible SSDC estimate is for later adopters, and I label it as such throughout rather than as an effect for all SSDC states.

Table 3b. SSDC estimates by sample

Sample	Estimator	Effect (pp)	Note
All SSDC adopters (15), 2002–2024	Sun-Abraham	0.6	null; early cohorts have truncated pre-periods
Later adopters (8), 2002–2024	Sun-Abraham	**3.38**	headline (WCB $p = 0.051$)
Later adopters (8), 1997–2024 extended panel	Sun-Abraham	3.34	stable to a longer pre-period

Source: `analysis/runs/phase1/ssdc_sunab_summary.csv`, `phase1d_pre2002_extension.csv`.

Design-based and weighted inference. Because cluster-robust asymptotics are fragile with eight treated states, I also run randomization inference: reassigning the eight observed adoption timings to random states in the analytic sample and recomputing the Sun-Abraham ATT, the placebo null is centered at zero (mean 0.04 pp, SD 1.20, 95th percentile of absolute placebo effects 2.34 pp), and the real 3.38 pp lies far in the tail — a two-sided randomization-inference $p = 0.004$. This design-based test does not rely on the small-cluster asymptotics and is, if anything, sharper than the wild-cluster bootstrap. The estimate above weights states equally and so describes the *typical* later adopter; weighting by drug spending — the fiscally relevant quantity — gives a larger 4.37 pp, because the gains are larger among the bigger recent adopters. Treatment-timing choices do not move the result: lagging treatment by one year gives 3.52 pp and dropping the partial entry year gives 3.98 pp (Section 5.3 reports the dynamic ramp).

A multiple-testing caveat. The paper reports five pool definitions, and a referee will rightly ask whether the borderline SSDC p survives family-wise correction. Across the pool-family tests, Holm FWER adjustment leaves only any-pool near significance (adjusted $p = 0.065$) and inflates the individual SSDC test to 0.204; Benjamini-Hochberg FDR control leaves SSDC, TOP$, Prime, and any-pool at adjusted p of 0.06–0.09. The SSDC later-adopter result is therefore best defended on its *own* design-based evidence (randomization-inference $p = 0.004$, the robustness battery, and the clean-entry and dynamic-ramp patterns), not on a single cross-pool p-value; I treat the cross-pool ranking as suggestive rather than as five independent confirmed effects.

5.3 Dynamic Pattern and Cohort Support

Figure 5 shows the SSDC event study. Pre-treatment leads are flat and individually insignificant — the observed pre-treatment coefficients are small and statistically indistinguishable from zero in the later-adopter SSDC sample, consistent with the de Chaisemartin-D'Haultfœuille placebos. The post-adoption path rises gradually and peaks around event years three to five, consistent with the class-by-class renegotiation ramp described in Section 4.3. The ramp is quantitative, not just visual: aggregating the dynamic coefficients by horizon, the effect is essentially zero in the entry window (event times 0–1: 0.2 pp), reaches 3.2 pp by years 2–3, and 5.1 pp by years 4–8. This staged pattern is what the bargaining mechanism predicts — rebates are renegotiated class by class through contract and PDL cycles, not on the day a state joins — and it is reassuring that the effect does *not* appear as an immediate same-year jump, which would more plausibly signal a mechanical reporting change than a bargaining response. I truncate the plotted path at event time 8: event times of nine or more are identified by three or fewer states, and eleven or more by a single state (North Dakota), so the far-right tail reflects which one or two states happen to be observed rather than dynamic structure. Inference rests on the aggregated effect, not on individual late-event-time coefficients.

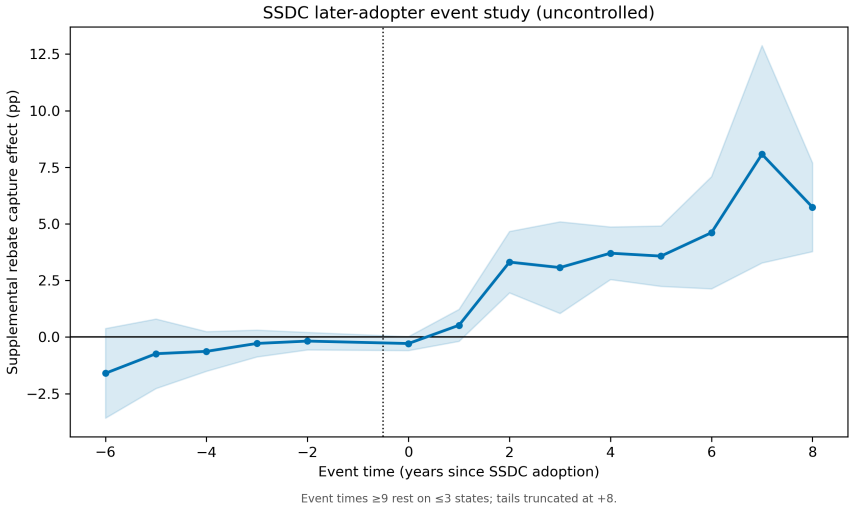

Figure 5: SSDC later-adopter event study (truncated at event time 8)

5.4 Co-Timed Reforms and Clean-Entry Robustness

The central threat to a causal reading is that SSDC entry coincides with other pharmacy-benefit reforms — a uniform PDL binding managed care, a carve-out, the extension of supplemental-rebate collection to managed-care utilization, or a single state PBM — each of which can independently move the rebate share.

To make that bundling auditable rather than asserted, Table 5 records, for each of the eight later adopters, the year of every such reform relative to SSDC entry, coded from the primary-source reform record. The pattern is reassuring for identification: only **two** of the eight — Delaware (2016) and Pennsylvania (2020) — entered SSDC in the same year as a major reform (in both cases the extension of MCO rebate collection, and for Pennsylvania a statewide PDL). The other six entered without a co-timed reform; where reforms occur, they typically arrive three or more years after pool entry (Ohio, Oklahoma) or several years before it (Kentucky's 2021 reforms predate its 2024 SSDC switch).

Table 5. Policy-bundle audit: co-timed reforms around SSDC entry

State	SSDC entry	Uniform PDL (MCO)	Carve-out	MCO rebate collection	Single PBM
Mississippi	2012	—	—	post (+2)	post (+12)
North Dakota	2015	—	post (+5)	post (+2)	—
Delaware	2016	—	—	**same-yr**	—
Ohio	2017	post (+3)	—	post (+3)	post (+5)
Oklahoma	2017	—	—	post (+7)	—
Pennsylvania	2020	**same-yr**	—	**same-yr**	—
South Dakota	2023	—	—	—	—
Kentucky	2024	pre (2021)	—	pre (2021)	pre (2021)

Source: `data/clean/governance_reform_timing.csv` (transcribed from the primary-source reform record); `analysis/runs/phase2/policy_bundle_audit.csv`.

The corresponding estimate confirms the audit. Re-estimating the later-adopter Sun-Abraham effect on the **clean-entry subsample** — the six states with no major reform within one year of SSDC entry, dropping Delaware and Pennsylvania — yields **3.11 pp** (SE 0.88), essentially the full-sample 3.38 pp. Dropping the three largest recent adopters (Pennsylvania, Ohio, Kentucky) instead gives 3.61 pp. The headline is therefore not an artifact of pool entry coinciding with managed-care rebate collection or formulary centralization; it is present in the states that joined SSDC *without* a simultaneous governance change. This does not eliminate the bundling concern — the estimand remains pool participation as implemented, and unobserved co-timed changes cannot be ruled out — but it substantially narrows it. (The governance-controlled sensitivity specification in Section 4.5 and Appendix D conditions on the *observed* reforms directly; because several are post-treatment mediators, it is reported as a residual association, not as the clean direct effect.)

Two further checks address the small comparison group directly. First, **leave-one-comparison-state-out**: dropping each of the 13 never-pooled states in

turn leaves the SSDC estimate in [3.12, 3.60] pp (lowest when California is dropped, highest when Arizona is), so no single comparison state drives the result. Second, **an alternative donor pool**: replacing the never-pooled comparison group with the states that were ever in a vendor pool but never in SSDC — a completely different counterfactual — yields 3.41 pp, essentially the 3.38 pp headline. The estimate is thus not an artifact of the particular 13-state comparison group, though that group's small size remains the binding constraint on precision (Section 6.3).

5.5 Multi-Pool Effects and Pool-Type Heterogeneity

Table 4 reports the exit-robust de Chaisemartin-D'Haultfœuille estimates for the reversing pools, on the full sample against never-pooled states. Pooling modestly raises supplemental rebate capture on average, and the effect is heterogeneous by pool type in the direction the theory predicts.

Table 4. Main results by pool (primary: uncontrolled total effect)

Treatment	Estimator	Effect (pp)	WCB CI / p	Treated	Comparison	Pre-trends
SSDC (later)	Sun-Abraham	**3.38**	WCB p = 0.051	8	13	clean
NMPI	dCDH (exit-robust)	**1.90**	CI [0.0, 3.8]	17	13	clean
TOP$	dCDH	0.34	CI [−1.0, 1.7]	11	13	clean
Prime (NMPI or TOP$)	dCDH	**1.36**	CI [0.15, 2.6]	27	13	clean
Any pool	dCDH	**1.22**	CI [0.1, 2.3]	37	13	clean

Notes: The comparison group is the same 13 never-in-any-pool states throughout (Section 3.5); "Treated" is the number of states ever in each pool over 2002–2024 — the SSDC row is the eight later adopters, the headline sample. Wild-cluster-bootstrap inference (Rademacher, 9,999 replications, clustered on state) is authoritative; dCDH confidence intervals are bootstrapped. The SSDC headline's robustness range (cross-estimator 2.66–4.27; leave-one-state-out [2.91, 3.88]) appears in Table 3. "Pre-trends: clean" means the dCDH placebo (pre-period) estimates are jointly small and statistically indistinguishable from zero.

The state-owned SSDC (about 3.4 pp among later adopters) and NMPI (1.9 pp) are positive; the any-pool average is 1.2 pp; and the TOP$ estimates are

centered near zero (0.34 pp, CI spanning zero) and rule out effects as large as those estimated for later SSDC adopters, but modest positive or negative effects remain possible. The SSDC all-treated estimate is null — dragged down by the early-adopter cohorts' truncated pre-periods — which is exactly why the later-adopter sample is the appropriate SSDC design; a direct SSDC-versus-other-pools contrast is about +3.2 pp but is not statistically distinct. The ordering (state-owned/credible at least matching vendor-administered, with the weakest vendor pool a null) is consistent with the residual-claimant-alignment and steerable-volume mechanisms of Section 2.4, though pool type is not randomly assigned and selection cannot be ruled out.

5.6 Theory-Motivated Heterogeneity

Table 6 collects the heterogeneity cuts. They are directionally consistent with the bargaining theory but imprecise under small-cluster inference, and one is non-confirmatory.

Table 6. Heterogeneity (theory-motivated; suggestive, underpowered)

Cut	Result	Theory prediction
Small vs. large states (SSDC)	small +3.37 vs. large +1.25 pp (ns)	larger for small (directional)
Small vs. large (any-pool)	small +2.52 vs. large +1.02 pp (ns)	(directional)
Switchers-in vs. direct (SSDC)	2.37 vs. 2.78 pp (ns)	no "already-captured" discount
Pool type	SSDC/NMPI positive; TOP$ null	state-owned/credible > vendor
Class substitutability (state-year proxy)	non-confirmatory; leans opposite, ns	larger where substitutes exist — not supported

Smaller states gain about twice as much as larger states under both SSDC and any-pool — the sign the countervailing-power model predicts, since pooling relaxes the small-state bargaining constraint most — but the differential is not statistically distinguishable from zero. States that switched into SSDC from a vendor pool gain about as much as direct adopters, providing no support for an "already-captured-gains" discount.

The class-substitutability prediction is the one cut that does *not* confirm. Using a state-year drug-mix proxy for substitutability (the innovator- multiple-source spend share, and one minus the single-source share), the pool effect is no larger — if anything mildly smaller — in more-substitutable states, and every differential is insignificant under the wild-cluster bootstrap. I report this honestly as non-confirmatory rather than omit it. Two caveats limit what it can establish: the

proxy is a *state-year* drug mix, not a *within-class* measure, so it confounds "this state buys more substitutable drugs" with correlated state characteristics; and a coherent alternative reading of the (insignificant) sign is that supplemental rebates are the only price lever on protected sole-source drugs, so pooling's surplus may concentrate where spend is in those high-margin brands. A genuine test of the class-level mechanism requires an SDUD-by-therapeutic-class panel, which I flag as future work. I do not treat class substitutability as supporting evidence.

5.7 Baseline Capture Among SSDC Adopters

A natural worry runs the other way: that SSDC states were already high performers, so the estimate reflects who joins rather than what joining does. The descriptive record points the opposite direction. In the years *before* they adopted SSDC, the eight later adopters captured a supplemental rebate share of 1.33 percent of drug spending — *below* the 2.31 percent of the never-pooled comparison states and the 1.74 percent of states that were in a vendor pool but never joined SSDC. States appear to have turned to SSDC because their existing rebate capture was inadequate, not because they were already extracting more. This reduces concern that the result is driven by already high-performing rebate programs, but it does not by itself establish conservative bias; low baseline capture may also predict mean reversion or concurrent reforms. I therefore treat this pattern as descriptive context rather than identification evidence.

5.8 Secondary Outcomes, Steering, and the Capture–Savings Distinction

The primary outcome is rebate *capture*, not net spending, and Table 7 makes the distinction concrete by re-running the later-adopter Sun-Abraham estimator on other outcomes built from the same CMS-64 and SDUD data. Three patterns stand out. First, **capture is not demonstrated savings, but neither is it offset.** Pooling raises the supplemental rebate share by 3.4 pp and supplemental rebate dollars per prescription by $3.19, while log gross drug spending does not move (−0.04, SE 0.05) and log drug spending net of supplemental rebates is directionally lower but not significant (−0.06, SE 0.06). So the captured rebate is *not* clawed back through higher gross outlays — there is no evidence of the high-list-price, high-rebate branded channeling that would make capture illusory — but the net-spending effect is too imprecise to assert a savings, and the design measures neither welfare nor manufacturer pricing responses. The paper therefore reserves "savings." (I net out only the *supplemental* rebate; the federally set national rebate share trends up over the panel for reasons unrelated to pooling, and including it would contaminate a net-spending contrast with a denominator trend.) Second, the rebate gain is **not a mechanical drug-mix artifact**: the single-source (protected-brand) and innovator-multiple-source spend shares do not move (−0.03 and −0.83 pp, both n.s.), so pooling is not simply shifting states into higher-rebate drug categories. Third, and most informative for

mechanism, pooling **concentrates utilization** — the within-state top-10-NDC spend share rises 2.2 pp (SE 0.69) and the NDC-spend Herfindahl rises — the formulary-steering channel the theory predicts: a credible pool steers volume onto the preferred (rebated) products. This is direct, positively-signed evidence for the steering mechanism, in contrast to the non-confirmatory state-year substitutability *proxy* of Section 5.6.

Table 7. Secondary outcomes (SSDC later-adopter Sun-Abraham)

Outcome	Effect	SE
Supplemental rebate share (pp) — *primary*	3.38	0.77
Log gross drug spending	−0.04	0.05
Log drug spending net of supplemental rebates	−0.06	0.06
Supplemental rebate $ per prescription	3.19	1.04
Single-source spend share (pp)	−0.03	0.66
Innovator-multi spend share (pp)	−0.83	0.85
Top-10 NDC spend share (pp) — *steering*	2.18	0.69

Source: `analysis/runs/phase3/phase3f_secondary_outcomes.csv`. "Net of supplemental rebates" is gross spending times one minus the supplemental rebate share (the federally set national rebate is deliberately left in, to avoid a denominator trend unrelated to pooling); enrollment-denominated per-capita outcomes are unavailable in CMS-64 and are not estimated.

Why a within-therapeutic-class design is not feasible here. A referee will ask for the sharpest mechanism test — the pool effect on rebate capture *within* therapeutic class, interacted with class-level substitutability. That design is genuinely infeasible with these data, for three concrete reasons. First, supplemental rebates are reported only as a single state-level line in CMS-64; they are never broken out by drug or therapeutic class, so class-level rebate *capture* is unobservable regardless of any crosswalk. Second, no NDC-to-therapeutic-class (ATC) crosswalk is available on the project's data; the only class information is the CMS single-source / innovator-multiple-source / non-innovator category, a coarse substitutability proxy. Third, the NDC-level SDUD needed to build any within-class utilization measure covers 2011–2016 on disk, which predates the entry of six of the eight later adopters. The feasible mechanism evidence is therefore the state-year steering result above; a true within-class design would require class-coded rebate data (for example, T-MSIS under a data-use agreement) joined to an ATC crosswalk, and remains the lead extension.

6. Discussion

6.1 Interpretation and Policy Implications

Interstate Medicaid drug purchasing pools modestly raise the supplemental rebates states extract, by roughly one to two percentage points of drug spending on average and about three points for SSDC later adopters. Against a no-pool base of under two percent and a binding federal rebate floor, a few-percentage-point gain is a real fiscal lever — but a modest one, not a transformation. The policy reading is correspondingly measured: pooling may be a useful fiscal tool when paired with governance arrangements that make volume steering credible, but nominal pool membership alone should not be interpreted as sufficient to reduce net drug spending. It is especially relevant for smaller states and where the pool can credibly steer volume, but it is not a substitute for the governance that makes volume steerable (unified PDLs, MCO rebate collection) and it will not, by itself, close the gap between low-capture and high-capture states.

The pool-type pattern carries the sharpest policy content. The state-owned SSDC and NMPI deliver positive effects while TOP$ does not, and the ordering aligns with how credibly each arrangement converts pooled lives into a disagreement threat and with whose budget the bargaining serves. For a state choosing whether and how to pool, the design of the pool — residual-claimant alignment, state-specific contracting that preserves price discrimination, a repeated member-state governance cycle — plausibly matters as much as the decision to pool at all. That said, pool type is chosen, not assigned, and SSDC states differ in ways (unified PDLs, MCO rebate collection, recent large joiners) that the design cannot fully separate from governance; the SSDC estimate is best read as the effect of SSDC *as implemented*, bundle included.

To fix orders of magnitude — not to forecast a budget line — consider the scale. Gross Medicaid drug spending is roughly $80 billion, of which the fifteen SSDC states account for about $20 billion, leaving some $60 billion outside SSDC. Applying the any-pool central estimate (1.2 pp) to that $60 billion implies on the order of $0.7 billion in additional supplemental rebates a year; applying the larger SSDC later-adopter estimate (3.4 pp) implies closer to $2.0 billion. Three forces pull the realistic figure toward the lower end and possibly below it: the estimates are imprecise and several intervals include zero; the gains are not separable from the governance reforms that travel with pooling, so nominal pool entry alone need not deliver them; and as adoption saturates, the marginal state has less untapped leverage to capture. And because the outcome is rebate *capture* rather than net spending, even realized rebate dollars need not translate one-for-one into budget savings if a rebate-maximizing formulary steers toward high-rebate branded drugs. These projections therefore bound a plausible range — roughly $0.4–1.0 billion annually under conservative assumptions — and should be read as an upper-bounded order of magnitude, not an expected return on nominal pool membership.

6.2 Relation to Prior Literature

The finding sits naturally in the buyer-power and pooled-procurement literatures. Chown et al. (2019) and Ho and Lee (2024) imply that scale raises rebates only when attached to a credible formulary decision; the small-state and pool-type heterogeneity here is the Medicaid-specific version of that prediction. The international evidence (Dubois et al., 2021; Parmaksiz et al., 2022) that pool *design* drives outcomes is mirrored in the SSDC-versus-TOP$ contrast. And the pharmacy-benefit literature (Dranove et al., 2021; Hernandez and Gellad, 2020) explains why governance is part of the mechanism rather than a nuisance to be partialled out — the reason the control policy here is symmetric and the governance-controlled estimate is a sensitivity specification rather than the headline.

6.3 Limitations

I group the limitations into four categories: internal validity, measurement, external validity, and policy interpretation.

Internal validity. Even though the observed pre-treatment coefficients are small and statistically indistinguishable from zero in the later-adopter SSDC sample, differential trends cannot be ruled out, and the Rambachan-Roth bounds make this concrete: the SSDC effect survives the assumption of exactly parallel trends but breaks under modest violations. Pool entry can co-time with other pharmacy-benefit reforms (uniform PDLs, carve-outs, vendor or negotiation-model changes), so the estimate reflects pool participation as implemented rather than pooling in isolation. Because the later adopters started from below-average baseline capture, mean reversion among low-baseline states is a live alternative to a genuine pooling effect. Adoption is policy-endogenous — states choose whether and when to pool — and the timing of treatment is itself measured with some uncertainty for a few states, which matters because a staggered-DiD coefficient is a weighted average over event-time cells defined by the treatment dates. Finally, cross-state or cross-manufacturer spillovers would violate SUTVA: if a manufacturer's concession to a pool changes the terms it offers non-pool states, the never-pooled comparison group is not a pure counterfactual.

Measurement. CMS-64 reports on a fiscal-year basis, which need not align with calendar-year pool-membership indicators. Six state-years carry negative supplemental-rebate values that are difficult to reconcile with voluntary rebate payments, and 39 state-years are missing the outcome entirely. The outcome is a rebate *share*, not a rebate *rate* negotiated per unit, so composition changes can move it independently of bargaining. Most importantly for interpretation, managed-care rebate-reporting changes can mechanically move the numerator: a state that begins collecting and reporting supplemental rebates on managed-care utilization raises the measured share without necessarily negotiating harder. And the SDUD category crosswalk is weak in the early years (matched spending

share is low before 2010), so the drug-mix measures are most informative in the later part of the panel.

External validity. The credible SSDC estimate is identified off later adopters, who differ from the early adopters in size and capacity, so the headline does not necessarily describe the early cohorts. Because most states had already pooled by 2024, the remaining non-adopters may not resemble prior adopters, which limits how well the estimates generalize to the future marginal adopter. The pattern documented here should therefore be read as the effect for states like the later adopters, not as a universal pooling effect.

Policy interpretation. Rebate *capture* is not net drug spending: a rebate-maximizing formulary can steer utilization toward high-rebate branded drugs, so realized rebate dollars need not translate one-for-one into lower net spending. State fiscal savings are smaller still than total rebate dollars, because the federal government shares in Medicaid drug costs through FMAP, so a captured rebate is split between state and federal budgets. And the access and prescribing consequences of pool-negotiated formularies are not measured here.

Two cross-cutting points close the discussion. First, the binding constraint on precision is structural: by 2024 most states had pooled, leaving only thirteen never-in-any-pool comparison states and between eight and a few dozen treated clusters depending on the pool, so several confidence intervals hug zero and the heterogeneity cuts are underpowered. This imprecision cannot be fixed by adding already-treated states — the comparison group is the scarce resource — but it can be mitigated by longer post-periods as recent adopters accumulate data, by better pre-period data, by alternative donor designs, and by class-level outcomes. Second, one comparison-group state (Tennessee) remains provisionally coded pending an agency reply; it is immaterial to the SSDC headline.

6.4 Future Research

The most valuable extension is a class-level mechanism design. Section 5.8 shows why it cannot be built here — supplemental rebates are reported only at the state level in CMS-64, the project lacks an NDC-to-therapeutic-class crosswalk, and the NDC-level SDUD on disk predates most adopters — so the within-class formulary- auction test would require class-coded rebate data (T-MSIS under a data-use agreement) joined to an ATC crosswalk and SDUD coverage spanning the adoption window. Short of that, the state-year steering evidence in Section 5.8 (pooling concentrates utilization onto the top NDCs) is the strongest available mechanism signal and points the way. A second priority is the net-spending question: this paper finds no detectable change in net drug spending alongside the capture gain, but a design powered to detect modest net-spending effects — and to test whether rebate gains are partly offset by branded-drug channeling, as Munshi et al. (2018) and Hernandez and Gellad (2020) would suggest — would sharpen the capture-versus-savings distinction. Finally, as more recent large adopters accumulate post-period data, the precision constraint

will ease and the pool-type and size heterogeneity that are only directional here may become testable.

---

7. Conclusion

Two-thirds of U.S. states have joined an interstate Medicaid drug purchasing pool, yet the question of whether pooling actually raises supplemental rebate capture has rested on anecdote. Using a corrected, primary-source panel of pool membership for all 50 states from 2002 through 2024 and estimators matched to each pool's institutional structure, this chapter finds that pooling *modestly raises* supplemental rebate capture — about one to two percentage points on average, and roughly three points for SSDC later adopters — concentrated in the state-owned SSDC and in NMPI, absent for TOP$, and larger for smaller states. The pattern is consistent with a countervailing-buyer-power mechanism in which pooling helps where the pool can credibly steer volume and where the residual claimant governs the bargaining. Under the maintained difference-in-differences assumptions, the evidence is consistent with a modest positive effect; the estimates are imprecise, for the structural reason that most states have already pooled, and remain sensitive to modest departures from parallel trends.

The lesson for states is that pooling may be a useful fiscal tool when paired with governance arrangements that make volume steering credible, but nominal pool membership alone should not be interpreted as sufficient to reduce net drug spending. Its payoff depends on pool design and on the governance that makes volume steerable — plausibly larger for smaller programs and for pools that can credibly steer drug volume, but not a substitute for that governance and not, on its own, a fix for the gap between low-capture and high-capture states.

Appendix

All results in this appendix use the **primary-source pool-membership panel** (`pool_membership_resolved_2002_2024.csv`), 2002–2024, with never-in-any-pool states (13) as the comparison group unless noted. Inference is wild-cluster bootstrap (Rademacher, 9,999 replications, clustered on state); analytic p-values overstate significance given the small number of treated clusters and are flagged where reported. The primary specification is the **uncontrolled total effect**, applied symmetrically to every pool.

Appendix A. Pharmacy-Benefit Variables

Variable	Description	Why it matters
`uniform_pdl`	State has a uniform preferred drug list spanning relevant delivery systems	Can increase formulary leverage; part of how a pool converts size into steerable volume
`pharmacy_carve_out`	Pharmacy benefit carved out of managed care	Changes who controls the rebate and utilization-management structure
`mco_formulary_autonomy`	Degree of managed-care formulary independence	Captures whether managed-care plans dilute state leverage
`pdl_vendor`	PDL support vendor or administrator	Administrative capacity and contracting style; sparse, descriptive only
`pbm_vendor`	Pharmacy benefit manager or analogous contractor	Negotiation execution and reporting; sparse, descriptive only
`rebate_negotiation_mode`	State-direct, SSDC-individual, collective-pool, PBM-assisted, hybrid, or unknown	Distinguishes pool membership from how bargaining is actually done
`major_reform_flag`	Major administrative or policy transition in a spell	Flags reform-dense windows

These variables enter only as a **uniform governance-controlled sensitivity specification that absorbs observed contemporaneous pharmacy-benefit structure** (Appendix D), never as the headline, because they are part-confounder/ part-mediator and are densely coded only from 2010 onward (about 2% non-missing pre-2010, 92–100% from 2010). Because these variables may be mediators, these estimates should not be interpreted as the causal direct effect of pooling.

Appendix B. Master Results

B1. Main results by pool — primary uncontrolled total effect, vs. no-pool.

Treatment	Primary estimator	Effect (pp)	WCB / CI	Pre-trends
SSDC (later, 2012+)	Sun-Abraham	3.38	WCB $p = 0.051$	clean (dCDH placebos)
NMPI	dCDH (exit-robust)	1.90	CI [0.0, 3.8]	clean
TOP$	dCDH	0.34	CI [−1.0, 1.7] (null)	clean
Prime (NMPI or TOP$)	dCDH	1.36	CI [0.15, 2.6]	clean
Any pool	dCDH	1.22	CI [0.1, 2.3]	clean

SSDC all-treated is null (early-adopter truncated pre-periods); the later- adopter sample is the appropriate SSDC design. SSDC-vs-other-pools about +3.2 pp (not statistically distinct).

B2. SSDC later-adopter cross-estimator and robustness (uncontrolled).

Check	Result
Sun-Abraham (headline)	3.38 pp, WCB $p = 0.051$
Callaway-Sant'Anna	2.66 pp
Borusyak-Jaravel-Spiess	4.27 pp
de Chaisemartin-D'Haultfœuille	~2 to 3.8 pp
Saturated event study (avg post 0–8)	3.66 pp
Leave-one-state-out	[2.91, 3.88] pp
Outlier drop / winsorize 1/99	3.07–3.38 pp
Outcome-construction covariates (headroom, drug-mix, post-federal denom.)	3.6–4.0 pp, WCB 0.03–0.04 (sharpens)
Pre-2002 panel extension (1997–2024, CMS-64 Line 7A2)	3.34 pp (vs 3.38 baseline) — stable
Rambachan-Roth honest bounds	survives at $\bar{M} = 0$ [1.11, 5.49]; breaks under modest violations

The pre-2002 extension parses CMS-64 Line 7A2 ("Drug Rebate Offset — State Sidebar Agreement," defined on the CMS-64 form in 1998) to build a 1997–2024 outcome panel (`data/clean/ssdc_v2_analytic_1997_2024.csv`; `analysis/phase1d_ pre2002_extension.R`). The 1997–2001 years are pure pre-treatment — no pool membership is recorded before 2002 and every SSDC adoption is 2006 or later — so the extension lengthens the pre-period without

adding treated mass. The later-adopter Sun-Abraham average post-adoption effect is essentially unchanged (3.34 pp vs. 3.38 pp), confirming the headline is not an artifact of the 2002 panel start.

Appendix C. Why State and Year Fixed Effects Are Not Enough

State fixed effects absorb time-invariant differences across states, and year fixed effects absorb shocks common to all states in a given year. They do not absorb within-state, time-varying policy changes such as adopting a uniform preferred drug list, changing carve-out status, altering formulary autonomy, or shifting rebate-negotiation models. If those reforms occur near pool entry and also affect supplemental rebate capture, their effects can load onto the pool coefficient. The pharmacy-benefit panel measures them directly — but, because several are post-treatment mediators (pooling works *through* unified PDLs and MCO rebate collection), conditioning on them yields a sensitivity specification rather than a clean upgrade (Appendix D).

Appendix D. Governance-Controlled Sensitivity Specification

Re-estimating each design with time-varying governance controls (`uniform_pdl`, `pharmacy_carve_out`, `major_reform_flag`) gives the residual-after-observed-governance estimate. This sensitivity specification is **sign-unstable and imprecise** across pools — the symptom of conditioning on a part-mediator on a restricted post-2010 sample — and is reported only as a controlled residual association, never a headline. Because these variables may be mediators, these estimates should not be interpreted as the causal direct effect of pooling. Its instability reinforces that the governance-controlled estimate is not headline-grade.

Appendix E. SSDC Cohort Timing and Later-Adopter Sample

Cohort	States	Cohort	States
2006	Iowa, Maine, Vermont	2016	Delaware
2007	Utah	2017	Ohio, Oklahoma
2008	West Virginia, Wyoming	2020	Pennsylvania
2009	Oregon	2023	South Dakota
2012	Mississippi	2024	Kentucky
2015	North Dakota		

Later-adopter sample (first treated 2012 or later): Mississippi, North Dakota, Delaware, Ohio, Oklahoma, Pennsylvania, South Dakota, Kentucky (8 states).

Never-in-any-pool comparison: 13 states. SSDC is absorbing (no exits); NMPI and TOP$ reverse.

Appendix F. Heterogeneity

Cut	Result	Theory prediction
Small vs. large states (SSDC)	small +3.37 vs. large +1.25 pp; WCB p=0.33	larger for small (directional)
Small vs. large (any-pool)	small +2.52 vs. large +1.02 pp; WCB p=0.14	(directional)
Switchers-in vs. direct (SSDC)	2.37 vs. 2.78 pp; WCB p=0.89	no "already-captured" discount
Pool type	SSDC/NMPI positive; TOP$ null	state-owned/credible > vendor
Class substitutability (state-year proxy)	non-confirmatory (below)	larger where substitutes exist — not supported

Class-substitutability test (non-confirmatory). Mirroring the size test (baseline 2002–05 median split; WCB on the high-vs-low differential; uncontrolled), using `share_spend_innovator_multi` (primary) and `1 - share_spend_single_source` (alternative) as state-year substitutability proxies:

Treatment	Proxy	high−low differential	WCB p
SSDC later (2012 or later)	innovator_multi	−4.59 pp	0.13
SSDC all-treated	innovator_multi	−3.03 pp	0.13
Any pool	innovator_multi	−1.15 pp	0.30
SSDC / any	1 − single_source	−0.9 to −0.02 pp	0.74–0.98

The pool effect is no larger — if anything mildly smaller — in more- substitutable states, and every differential is insignificant. The proxy is a state-year drug mix, not a within-class measure, so it confounds drug basket with correlated state characteristics; a coherent (but unprovable) alternative is that supplemental rebates concentrate where spend is in protected sole-source brands. A within-class SDUD by therapeutic class (ATC) design is required to test the formulary- auction mechanism properly and is left to future work. This cut is reported as non-confirmatory and is not used as supporting evidence.

Appendix G. Inference (Wild-Cluster Bootstrap)

With 8 SSDC later adopters (and 13 never-pooled comparators), cluster-robust analytic standard errors over-reject. All headline inference uses the wild- clus-

ter bootstrap (Cameron-Gelbach-Miller, 2008; MacKinnon-Webb, 2017) with Rademacher weights, 9,999 replications, clustered on state, with both `set.seed` and `dqset.seed` fixed for reproducibility. Where an analytic p-value is shown, it is flagged as overstating significance.

Appendix H. Figure Package

The figure package is the minimum set for a staggered-DiD paper:

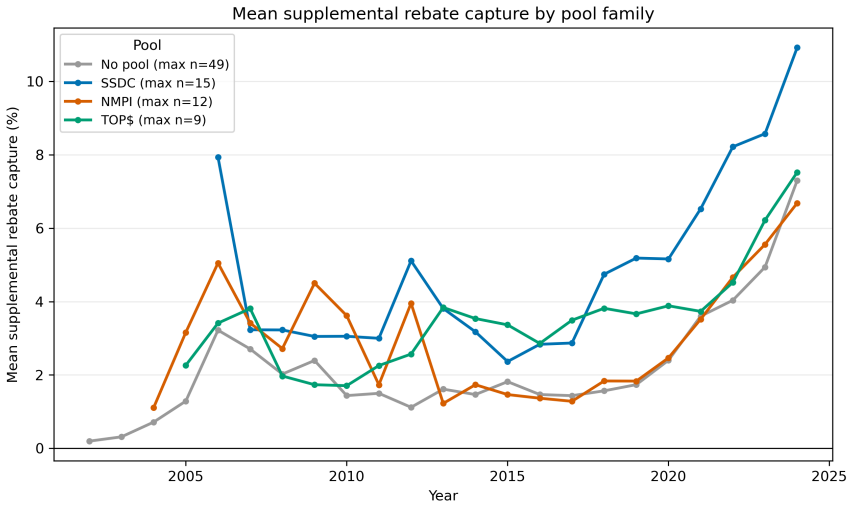

Figure 6: Supplemental rebate trends by pool family.

Appendix I. Mechanism Atlas

The schematics below unpack the governance levers in Figure 2 — the within-state arrangements through which a pool's covered lives become a single credible volume signal. They are conceptual illustrations of the mechanism, not data.

References

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6), 3253–3285.

Callaway, B., and Sant'Anna, P.H.C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414–427.

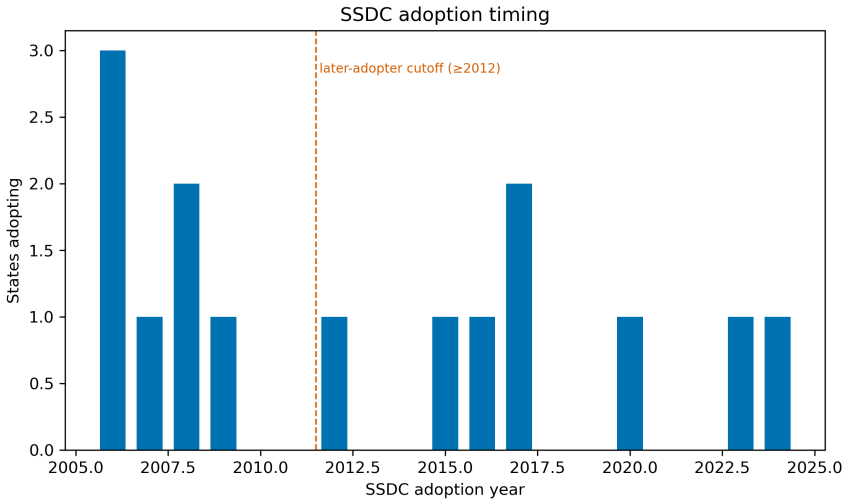

Figure 7: SSDC treatment timing support.

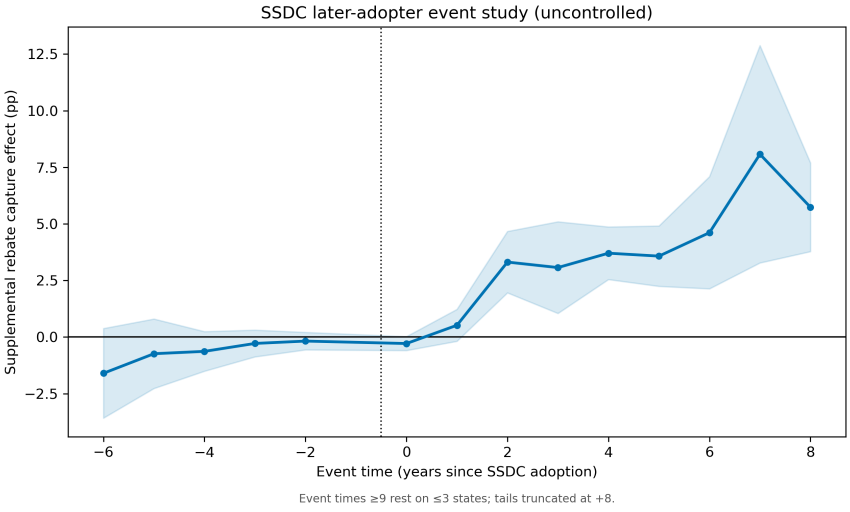

Figure 8: SSDC later-adopter event study (truncated at event time 8).

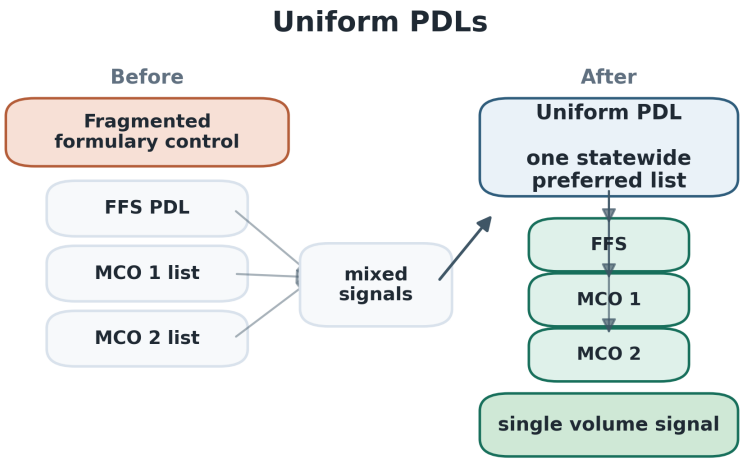

Figure 9: Uniform preferred drug lists: replacing fragmented fee-for-service and managed-care lists ("mixed signals") with one statewide preferred list, so a single volume signal stands behind a preferred-status decision.

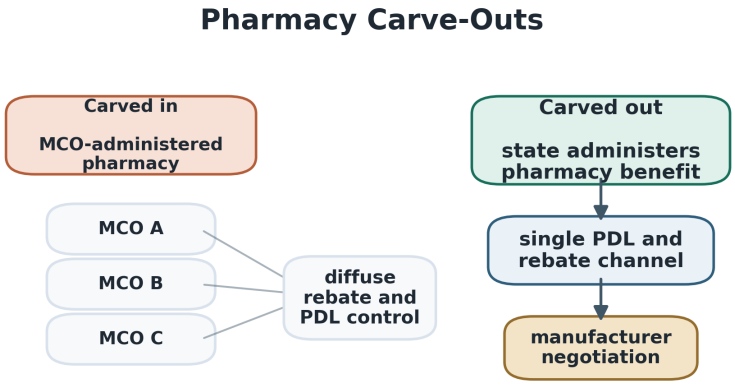

Figure 10: Pharmacy carve-out: moving the drug benefit from managed care back to fee-for-service centralizes formulary and rebate authority in the state.

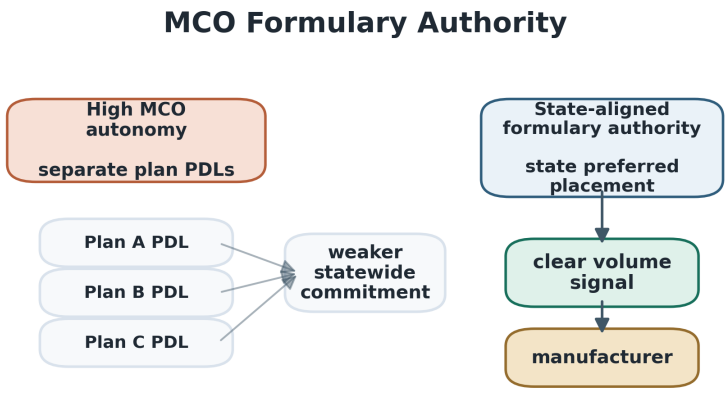

Figure 11: Managed-care formulary autonomy: where managed-care plans retain independent formularies, statewide steering is diluted.

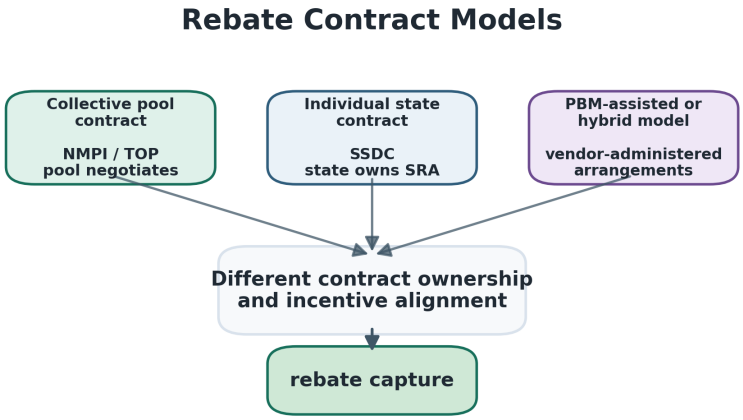

Figure 12: Rebate-negotiation ownership: who holds the supplemental-rebate contract — the state, a pool, or a vendor — shapes how aggressively each incremental rebate dollar is pursued.

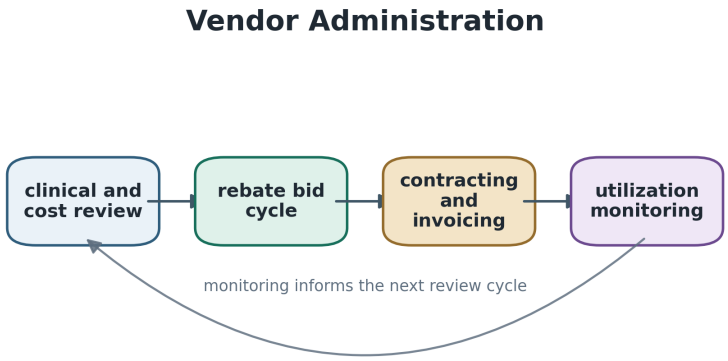

Figure 13: Vendor administration: the recurring clinical-review, rebate-bid, contracting, and utilization-monitoring cycle a pool's agent runs on the state's behalf.

de Chaisemartin, C., and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.

Horvath, J. (2019). State initiatives using purchasing power to achieve drug cost containment. National Academy for State Health Policy.

MacKinnon, J. G., and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2), 233–254.

Rambachan, A., and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5), 2555–2591.

Sun, L., and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.