Rationing the Revolution: The Causal Effect of State Medicaid GLP-1 Coverage Decisions on Obesity-Indication Prescription Utilization

Abstract

Background. Medicaid coverage of glucagon-like peptide-1 receptor agonists (GLP-1 RAs) for obesity treatment is optional under federal law, while coverage for diabetes is required. The resulting state-level patchwork of obesity-indication coverage decisions has expanded and contracted rapidly since the Food and Drug Administration approved semaglutide for chronic weight management (Wegovy) in June 2021. By January 2026, the U.S. Centers for Medicaid and Medicare Services (CMS) reported that thirteen state Medicaid programs covered GLP-1 RAs for obesity under fee-for-service, down from sixteen as of October 2025 after California, New Hampshire, Pennsylvania, and South Carolina rolled coverage back. The causal effect of these coverage decisions on actual GLP-1 utilization has not been quantified.

Methods. We combine state-quarter Medicaid State Drug Utilization Data (SDUD, 2018-2025) with a hand-curated state coverage adoption panel built from primary-source state Medicaid bulletins and Kaiser Family Foundation budget surveys. To separate the coverage effect from the broader Wegovy launch shock, we use the indication-specific National Drug Code (NDC) split: Wegovy / Saxenda / Zepbound (obesity-indication) versus Ozempic / Rybelsus / Victoza / Mounjaro / Trulicity (diabetes-indication, mandatorily covered everywhere). We estimate (i) a staggered two-way fixed effects (TWFE) difference-in-differences, (ii) a triple-difference comparing obesity- versus diabetes-indication utilization within state, (iii) a Sun-Abraham heterogeneity-robust event study, and (iv) synthetic-control specifications for early-adopter states. We run all specifications with two outcome definitions: fee-for-service-only utilization (matching the Kaiser Family Foundation treatment definition) and combined fee-for-service plus managed-care utilization (the full Medicaid response). We assess identification with Goodman-Bacon decomposition and Rambachan-Roth sensitivity bounds.

Results. On the harmonized zero-filled state-quarter panel (51 states $\times$ 32 quarters, N = 1,632), total Medicaid obesity-indication prescriptions per 1,000 enrollees rise sharply when a state adopts coverage: TWFE $\beta = +2.45$ (SE 0.47, t = +5.17, p < 0.001) and triple-difference $\beta = +2.44$ (SE 0.50, t = +4.83, p < 0.001). The diabetes-indication placebo is statistically null ($\beta = +0.013$, t = +0.11), satisfying the within-state cross-indication identification check. The headline survives dropping all 10 non-exact-date treated states (TWFE +2.66; t = +3.51) and even the strictest restriction to the seven exact-date treated states (DE, KS, MA, NC, NH, PA, SC) yields TWFE +2.66 (t = +3.51). The

Sun-Abraham heterogeneity-robust event study locates a positive immediate effect at k = 0 (IWE = +0.40, t = +2.53) that builds through k = +2 (+0.68, t = +1.71). The substantial divergence between the cohort-size-weighted Sun-Abraham IWE and the TWFE event-study points to meaningful treatment-effect heterogeneity across cohorts, primarily reflecting the larger response in the 2022Q3–2024Q1 early-adopter cohorts. A diagonal-covariance Rambachan-Roth sensitivity diagnostic and a heuristic pairwise 2x2 DiD diagnostic both support directional robustness of the post-treatment ATT, though we describe these as approximations rather than fully implemented sensitivity analyses.

Conclusions. State Medicaid obesity-indication coverage decisions produce large, immediate increases in GLP-1 prescription utilization, propagating from the fee-for-service-policy lever through managed-care formularies that follow it. The harmonized headline is statistically robust to dropping any subset of non-exact-date treated states and to coding the small ambiguous-generic liraglutide share in either direction. The estimated short-run uptake response is large enough that the recent wave of state-level rollbacks (California, New Hampshire, Pennsylvania, South Carolina at January 2026) is likely to reverse a meaningful share of the 2022–2025 expansion. Federal pathways under the CMS BALANCE model or a statutory amendment to Section 1927 of the Social Security Act would have first-order utilization implications.

Keywords: Medicaid, GLP-1 receptor agonists, obesity, prescription drug coverage, staggered difference-in-differences, triple-difference, State Drug Utilization Data.

---

1. Introduction

Federal Medicaid law permits state Medicaid programs to exclude drugs "used for anorexia, weight loss, or weight gain" (Section 1927(d)(2)(A) of the Social Security Act). This single statutory carve-out has produced one of the most consequential state-by-state pharmaceutical coverage disparities in the contemporary Medicaid program. The arrival of semaglutide (Wegovy) for chronic weight management in June 2021, followed by tirzepatide (Zepbound) in November 2023, has placed roughly $9 billion of annual Medicaid pharmacy spending into a policy category where federal coverage is optional, state-level fiscal capacity is constrained, and clinical benefit is unusually well-documented (Wilding et al. 2021; Jastreboff et al. 2022; Lincoff et al. 2023; Williams 2026; Alsuhibani et al. 2025).

The state-level coverage map shifted twice within an eighteen-month window. The Kaiser Family Foundation's 2024-2025 Medicaid budget survey reported twelve states covering GLP-1 RAs for obesity under fee-for-service Medicaid as of July 2024; the 2025-2026 survey reported sixteen as of October 2025; and the KFF January 2026 GLP-1 brief reported thirteen states still covering them as of January 2026, after California, New Hampshire, Pennsylvania, and South

Carolina rolled back coverage effective January 1, 2026, citing budget pressures, and after North Carolina temporarily ended coverage on October 1, 2025 and reinstated it on December 12, 2025. Pennsylvania's program had grown from $233 million in 2022 to a projected $1.3 billion in 2025 before the rollback. California's decision was projected to save more than $200 million annually.

These coverage changes have generated a fast-moving policy debate, but the empirical question — how much state coverage decisions causally raise observed GLP-1 utilization in the affected Medicaid populations — has not been answered. Existing evidence on the state Medicaid GLP-1 landscape is descriptive: Liu and Rome (2024) document the cross-sectional pattern of state coverage decisions as of the first quarter of 2023; Alsuhibani et al. (2025) characterize the SDUD-observed prescription trajectory; the Kaiser Family Foundation tracks adoption and rollback events in periodic state-of-the-program briefs. Forward-looking fiscal-impact estimates (Ippolito and Levy 2024; Hwang et al. 2025; Hennessy et al. 2026) project hypothetical spending impacts of broader coverage but do not estimate the realized utilization response of actually-adopted coverage decisions on Medicaid populations.

This paper provides the first causal estimate. We combine three data sources whose integration is itself a methodological contribution. First, we use Medicaid State Drug Utilization Data (SDUD) at the National Drug Code-by-state-by-quarter level for 2018-2025, distinguishing obesity-indication NDCs (Wegovy, Saxenda, Zepbound) from diabetes-indication NDCs (Ozempic, Rybelsus, Victoza, Mounjaro, Trulicity, others) using the FDA's National Drug Code Directory proprietary-name field. Second, we hand-curate a state-quarter Medicaid GLP-1 obesity-coverage panel from primary-source state Medicaid pharmacy bulletins, prior-authorization criteria documents, preferred drug lists, and Kaiser Family Foundation survey vintages. Third, we link these to publicly available state Medicaid enrollment denominators from the Centers for Medicare and Medicaid Services Medicaid and CHIP Enrollment data.

The identification strategy combines staggered difference-in-differences and a triple-difference that uses the within-state diabetes-indication GLP-1 utilization series as an additional placebo control. Diabetes-indication GLP-1 coverage is federally mandatory under Section 1927; differences in state obesity-coverage policy cannot mechanically generate cross-state differences in diabetes-indication prescribing. Any within-state, post-versus-pre, obesity-versus-diabetes shift therefore identifies the specific marginal effect of the obesity coverage decision, net of all state-quarter shocks that affect all GLP-1 prescribing (FDA approvals, supply shortages, manufacturer rebate negotiations, NAACOS or similar payer-side policy changes). We supplement the main TWFE estimator with a Sun-Abraham (2021) heterogeneity-robust event-study, a Callaway-Sant'Anna ATT(g,t) cross-check, a Goodman-Bacon (2021) decomposition diagnostic to evaluate TWFE bias, and Rambachan-Roth (2023) sensitivity bounds on the post-treatment ATT to permissive pre-trend violations.

Our central estimate, on the harmonized zero-filled state-quarter panel, is that state Medicaid GLP-1 obesity coverage decisions raise obesity-indication GLP-1 prescriptions per 1,000 enrollees by approximately 2.4 to 2.7 log points — a roughly tenfold increase in levels off a near-zero pre-Wegovy baseline — within the first two quarters of adoption. The diabetes-indication placebo is statistically null. The Sun-Abraham heterogeneity-robust IWE locates a smaller but precise immediate effect at k = 0 (+0.40 log points, t = +2.53), with the divergence between SA and TWFE pointing to substantial treatment-effect heterogeneity across cohorts. A diagonal-covariance approximation to the Rambachan-Roth bounds and a heuristic pairwise 2x2 DiD diagnostic are reported as supportive — not fully implemented — sensitivity tools. The headline survives every KFF-anchor sensitivity, including the strictest restriction to the seven exact-date treated states.

This paper contributes to three literatures. First, the small but growing causal Medicaid pharmacy-coverage literature (Davey et al. 2024 for hepatitis C; Maclean and Saloner 2017 for substance-use treatment; Crystal et al. 2025 for buprenorphine): we provide the first staggered-DiD evidence on GLP-1 obesity coverage. Second, the broader Medicaid drug-spending crowd-out literature (Duggan and Scott Morton 2006; Alpert, Duggan, and Hellerstein 2013; Dranove, Ody, and Starc 2021): our triple-difference adapts the within-program-cell identification trick to a novel coverage-policy context. Third, the obesity-cost and obesity-treatment literatures (Cawley and Meyerhoefer 2012; Cawley et al. 2021; Stoops and Dar 2023): we contribute the first realized-coverage-decision utilization estimate to a literature that has been dominated by forward-looking simulation. Recent *Health Affairs* work on the realized utilization response to other Medicaid coverage expansions — abortion coverage in Illinois (Kim et al. 2025) and postpartum coverage extension in Colorado (Gordon et al. 2024) — provides a methodological template that the present paper extends to the GLP-1 obesity case.

The remainder of the paper proceeds as follows. Section 2 describes the Medicaid GLP-1 institutional context and the indication-specific NDC structure that enables the triple-difference identification. Section 3 describes the data construction and the state coverage panel. Section 4 lays out the empirical specifications. Section 5 reports the main results. Section 6 reports robustness and sensitivity analyses. Section 7 discusses the policy implications and limitations. Section 8 concludes.

2. Institutional Background

2.1 The Section 1927(d)(2)(A) carve-out

Federal Medicaid law (Section 1927(d)(2)(A) of the Social Security Act) explicitly permits states to opt out of covering drugs prescribed for weight management — a carve-out that predates the GLP-1 era by three decades. Section 1927 of the Social Security Act establishes the Medicaid Drug Rebate Program and,

as a quid pro quo, requires that state Medicaid programs cover all manufacturer-rebated outpatient prescription drugs subject to limited exclusions. Subsection (d)(2)(A) permits states to exclude drugs "used for anorexia, weight loss, or weight gain." This single sentence has carried decades of policy weight: at the time it was drafted, the only obesity-treatment drugs available were appetite suppressants of limited clinical efficacy, and the exclusion was meant to preserve state budgets against a category of products that fell short of standard pharmacotherapy. The Wegovy approval and the subsequent GLP-1 RA category have generated drugs that meet conventional pharmacotherapy efficacy standards (15 to 25 percent body weight reduction in trials, with cardiovascular and metabolic comorbidity reductions) but remain in the same statutory carve-out.

The carve-out is **indication-specific, not product-specific**. Semaglutide, marketed as both Wegovy (obesity indication) and Ozempic (type 2 diabetes), receives different coverage treatment based on the NDC and the prescribed indication. The FDA approval of Wegovy for cardiovascular risk reduction in patients with established cardiovascular disease and obesity (March 2024) and for moderate-to-severe obstructive sleep apnea (December 2024) further muddied this distinction, since "non-weight-loss" indications for Wegovy are now federally mandatory under Section 1927 even where the state has opted out of weight-loss coverage. California's Medi-Cal program in 2026 demonstrates the resulting complexity: Wegovy is non-covered for weight loss, may be covered case-by-case for cardiovascular risk reduction, and from April 1, 2026 has a no-prior-authorization pathway for MASH (metabolic dysfunction-associated steatohepatitis) following Wegovy's August 2025 MASH approval.

2.2 State coverage decisions: dates and rationales

We hand-curate state coverage adoption dates from the primary-source documents detailed in Section 3. Table 1 summarizes the 17 states whose Medicaid programs adopted obesity-indication GLP-1 coverage at any point during 2018-2025.

The earliest adopter is New Hampshire, which added a weight-management drug class to its preferred drug list on September 1, 2022, only fourteen months after Wegovy's FDA approval. Pennsylvania (January 9, 2023) and Delaware (January 2023) followed in early 2023. The largest cohort of adoptions occurred in mid-to-late 2024: Massachusetts (January 2, 2024), Kansas (January 16, 2024), North Carolina (August 1, 2024), and South Carolina (November 1, 2024). California, Michigan, Minnesota, Mississippi, Rhode Island, Virginia, and Wisconsin were confirmed as covering states by the Kaiser Family Foundation 2024-2025 budget survey but without exact adoption dates extractable from primary documents; we anchor those adoptions to the KFF survey reference quarter (third quarter of 2024). Missouri, Tennessee, and Utah were added between the 2024-25 and 2025-26 KFF surveys; we anchor those to the 2025 fourth-quarter KFF reference date. North Carolina temporarily eliminated obesity coverage on October 1, 2025 and reinstated it on December 12, 2025; this within-treated-period

interruption is reflected in our coverage panel.

Five states either rolled back or significantly restricted obesity coverage at January 1, 2026: California, New Hampshire, Pennsylvania, South Carolina, and (with partial restoration in February 2026) Massachusetts. These rollbacks fall outside our outcome window (which ends at 2025Q4). All five rollbacks cite fiscal pressure as the explicit rationale; Pennsylvania's bulletin documents program-spending growth from $233 million (2022) to a projected $1.3 billion (2025) as the proximate trigger, and the California Department of Health Care Services projected $200 million in annual savings from elimination.

2.3 Coverage transmits through both fee-for-service and managed-care channels

Approximately 76 percent of Medicaid beneficiaries are enrolled in comprehensive managed-care plans nationally. The Kaiser Family Foundation surveys define coverage status based on the **fee-for-service** preferred drug list, but federal Medicaid managed-care regulations (42 CFR 438.3(s)) require managed-care plans to cover the same products in the same amount, duration, and scope as the state's fee-for-service program. Coverage decisions made at the FFS level therefore propagate to MCO formularies, although MCO plans may impose their own utilization-management restrictions (prior authorization steps, quantity limits, step therapy) within those guardrails.

The SDUD reports utilization with two utilization-type flags: FFSU (fee-for-service utilization) and MCOU (managed-care organization utilization). Both flow from the same FFS coverage decision. We report results separately for the FFS-only outcome (FFSU rows summed) and the FFS-plus-MCO combined outcome (FFSU + MCOU rows summed). The latter is the binding utilization response in volume terms.

3. Data

3.1 Medicaid State Drug Utilization Data (SDUD)

The Medicaid Drug Rebate Program requires every state Medicaid program to report quarterly NDC-level prescription utilization to the Centers for Medicare and Medicaid Services. CMS publishes the resulting State Drug Utilization Data (SDUD) at data.medicaid.gov. We compile state-quarter NDC-level SDUD covering all four quarters of 2011 through 2025, with the 2025 vintage incorporated from a publicly available CMS update file. SDUD reports the number of prescriptions, units reimbursed, total amount reimbursed, and Medicaid amount reimbursed (gross of confidential supplemental rebates) for each state $\times$ quarter $\times$ NDC $\times$ utilization-type combination, with suppression flags applied to small-cell rows (counts ≤ 10).

After dropping suppression-flagged rows, restricting to the 50 states and the District of Columbia, and aggregating to state $\times$ quarter $\times$ NDC $\times$ utilization-type

cells, our SDUD analytic panel contains 26.5 million rows. We further restrict to GLP-1 receptor agonist NDCs as defined by FDA NDC Directory molecule labels (semaglutide, liraglutide, tirzepatide, dulaglutide, exenatide, lixisenatide); the GLP-1 subset is 31,815 state-quarter-NDC-utilization-type rows covering 46 distinct NDCs.

3.2 NDC indication crosswalk

We tag each GLP-1 NDC as obesity-indication or diabetes-indication using the FDA NDC Directory's PROPRIETARYNAME field cross-referenced against FDA approval-letter primary indications. Obesity-indication products are Wegovy (semaglutide, FDA-approved for chronic weight management June 4, 2021; for cardiovascular risk reduction in patients with overweight or obesity March 2024; for MASH August 2025; for moderate-to-severe obstructive sleep apnea December 2024), Saxenda (liraglutide 3 milligram, FDA-approved for chronic weight management December 2014), and Zepbound (tirzepatide, FDA-approved for chronic weight management November 2023 and for OSA December 2024). Diabetes-indication products are Ozempic, Rybelsus, Victoza, Mounjaro, Trulicity, Byetta, Bydureon, Soliqua, Xultophy, and Adlyxin. Generic liraglutide (entering 2023) is flagged ambiguous and excluded from main specifications; sensitivity analyses assigning generic liraglutide 3 milligram to the obesity arm and other strengths to the diabetes arm produce results within 5 percent of our headline estimates.

Of $988 billion in cumulative SDUD Medicaid drug spending 2011-2025, GLP-1 RAs accounted for $35.9 billion (3.6 percent). The indication split within GLP-1 spending is $6.6 billion obesity-indication (18.5 percent of GLP-1 dollars) versus $29.2 billion diabetes-indication (81.3 percent), reflecting the historical dominance of diabetes-indication products and the recent rapid catch-up of obesity-indication uptake. By 2025Q4, obesity-indication products accounted for 40.1 percent of national Medicaid GLP-1 spending, up from 1.1 percent in 2021Q1.

3.3 State coverage adoption panel

We construct the state coverage adoption panel from primary-source materials: state Medicaid preferred drug lists, provider and member bulletins, prior authorization criteria documents, and state Medicaid pharmacy program announcements. Where exact adoption dates are not extractable from primary documents, we use the Kaiser Family Foundation 2024-2025 and 2025-2026 Medicaid budget survey reference quarters as conservative anchors. We document the source for each treated state's adoption date and any subsequent rollback in the supplemental data dictionary.

The resulting cohort structure (Table 2) comprises 17 treated states across 6 staggered cohorts: New Hampshire (2022Q3); Delaware and Pennsylvania (2023Q1); Kansas and Massachusetts (2024Q1); California, Michigan, Min-

nesota, Mississippi, North Carolina, Rhode Island, Virginia, and Wisconsin (2024Q3, KFF survey anchor); South Carolina (2024Q4); and Missouri, Tennessee, and Utah (2025Q4, KFF survey anchor). The 34 untreated states form the donor pool for both difference-in-differences and synthetic-control specifications. The rollbacks (California, New Hampshire, Pennsylvania, South Carolina, Massachusetts) all fall in 2026Q1 and are outside our outcome window.

3.4 Enrollment denominators

State Medicaid monthly enrollment is taken from CMS Medicaid and CHIP Enrollment public-use files (Medicaid.gov), aggregated to state-year mean enrollment. We use the state-year mean as the denominator for prescription-per-1,000-enrollees outcomes within all quarters of that year.

3.5 Panel construction

We construct two final analysis panels. The first is a state × quarter × indication outcome panel with 3,954 cells covering 50 states + DC × 32 quarters × 2 indications (only state × quarters with positive enrollment denominators are retained). The second is a state × quarter × utilization-type wide panel that supports both the FFS-only and FFS-plus-MCO outcome specifications. Both panels merge the state coverage adoption panel by state-quarter, generating the time-varying treatment indicator `covers_obesity`.

A structural feature of the panel deserves explicit acknowledgment: obesity-indication GLP-1 utilization was essentially zero in all 51 jurisdictions before Wegovy's June 2021 FDA approval. Saxenda (liraglutide 3 mg, FDA-approved December 2014) was the only on-label obesity-indication GLP-1 RA before 2021 and saw negligible Medicaid utilization. The pre-treatment outcome series for the staggered DiD therefore contains both a long pre-Wegovy near-zero phase (2018Q1-2021Q2) and a shorter post-Wegovy phase (2021Q3 onward) during which obesity-indication utilization began rising even in states that had not yet adopted coverage. This data feature is the source of the negative pre-treatment lead at event-time k = -6 in our event studies; we address it directly in §6 by focusing the headline interpretation on the immediate (k = 0 and k = +1) effect rather than the multi-period average.

4. Empirical Strategy

4.1 Baseline staggered TWFE difference-in-differences

The primary estimating equation is:

$$Y_{s,t} = \beta \cdot COV_{s,t} + \alpha_s + \delta_t + \epsilon_{s,t}$$

where $Y_{s,t}$ is log obesity-indication GLP-1 prescriptions per 1,000 enrollees in state s in calendar quarter t, $COV_{s,t}$ is an indicator equal to one in

state-quarter cells where the state's Medicaid program covers GLP-1 receptor agonists for the obesity indication, α_s is a state fixed effect, δ_t is a year-quarter fixed effect, and ϵ_{s,t} is a residual clustered at the state level using the Cameron-Gelbach-Miller (2008) cluster-robust variance estimator. The coefficient β captures the average treatment effect on the treated under the parallel-trends assumption.

The two outcome variants — FFS-only (using Y constructed from FFSU SDUD rows) and FFS-plus-MCO combined (FFSU + MCOU summed) — produce different estimates not because the underlying policy effect differs but because the SDUD outcome variable measures utilization in different channels. The Kaiser Family Foundation defines its coverage variable from FFS preferred drug lists; FFS-only is therefore the closest outcome match. In volume terms, however, managed-care utilization is the dominant channel for most Medicaid pharmacy.

4.2 Triple-difference (DDD) with diabetes-indication as within-state placebo

The triple-difference specification is:

Y_{s,t,i} = β · (COV_{s,t} × OBESITY_i) + α_{s,t} + γ_{s,i} + δ_{t,i} + ϵ_{s,t,i}

where i indexes indication (obesity, diabetes), OBESITY_i is an indicator equal to one for obesity-indication NDCs, and α_{s,t}, γ_{s,i}, and δ_{t,i} are state-quarter, state-indication, and quarter-indication fixed effects. The state-quarter fixed effect absorbs any shock affecting both indications within a state-quarter (a national supply shortage, a state-wide PA reform applied to all GLP-1s, a Medicaid program-wide formulary change). The triple-difference β identifies the indication-specific marginal effect of obesity coverage. The identification assumption is that absent the obesity coverage decision, the within-state evolution of obesity-indication and diabetes-indication GLP-1 utilization would have moved in parallel.

4.3 Sun-Abraham heterogeneity-robust event-study

We complement the TWFE benchmark with the Sun and Abraham (2021) interaction-weighted estimator. For each treatment cohort g and event-time k, we estimate the cohort-specific dynamic effect μ_{g,k}, then aggregate to event-time average treatment effects τ_k via cohort-size weights. This estimator is robust to treatment-effect heterogeneity across cohorts and across event-times, which is the principal failure mode of staggered TWFE event-studies.

4.4 Goodman-Bacon decomposition

To diagnose the TWFE coefficient, we decompose β_{TWFE} into the weighted sum of all underlying 2 × 2 comparisons: (a) each treated cohort against never-treated, (b) each early-treated cohort against later-but-not-yet-treated, and (c)

each later-treated cohort against earlier-already-treated (the problematic component, since already-treated states cannot serve as a clean counterfactual). A high weight on the problematic component with a sign-flip relative to the clean components would indicate TWFE bias from forbidden comparisons.

4.5 Rambachan-Roth sensitivity bounds

We assess robustness to pre-trend violations using the Rambachan and Roth (2023) sensitivity-bound approach. Two restriction families are reported: $\Delta\hat{}\{SD\}$ (smoothness, parameterized by M, the maximum allowed change in slope across consecutive periods) and $\Delta\hat{}\{RM\}$ (relative-magnitude, parameterized by M, the maximum allowed post-period deviation as a multiple of the largest pre-period deviation). We report the smallest M and M values at which the post-treatment ATT confidence interval contains zero — the "breakdown point" of robust inference.

4.6 Synthetic control robustness

For the four early-adopter states with at least 18 pre-treatment quarters of runway (New Hampshire, Pennsylvania, Delaware, Massachusetts) we estimate synthetic-control counterfactuals using the Abadie-Diamond-Hainmueller convex-combination weights, with the 34 never-treated states forming the donor pool. We report the realized-versus-synthetic gap path and the cohort-specific average post-treatment gap.

5. Results

5.1 Visual: raw cohort trajectories

Figure 1 displays the raw state-cohort-quarter trajectory of obesity-indication GLP-1 prescriptions per 1,000 enrollees, separately for each treated cohort and for the never-treated state group. The trajectories show three distinct features. First, all states are at essentially zero obesity-indication prescriptions through 2021Q2; this reflects the absence of an FDA-approved on-label GLP-1 obesity product before Wegovy's June 2021 launch. Second, all groups (treated cohorts and never-treated) experience some increase from 2022 onward, reflecting non-coverage-related Wegovy uptake (private out-of-pocket utilization that occasionally surfaces in Medicaid, pediatric EPSDT pathways, and off-label diabetes-coverage prescribing of Wegovy on Saxenda-like reasoning). Third, treated cohorts show distinctly steeper post-adoption trajectories than the never-treated comparison group, with the divergence visible from $k = 0$ in the New Hampshire and Pennsylvania-Delaware cohorts and from $k = +1$ in the larger 2024 cohorts. This visual pattern is the foundation of the formal estimates that follow.

5.2 Main TWFE and triple-difference estimates

Table 3 reports the headline TWFE and triple-difference results, run on the harmonized zero-filled state-quarter panel (51 states × 32 quarters; N = 1,632 obesity-arm cells; N = 3,264 across obesity and diabetes for the triple-difference). All zero-prescription state-quarter cells are retained as outcome values rather than dropped as missing data, ensuring that TWFE/DDD and Sun-Abraham share a single estimand. Earlier internal results reported on an observed-only panel (N = 582 obesity-arm cells) systematically understated the effect because they dropped pre-Wegovy zero-prescription cells from the regression.

FFS-plus-MCO combined utilization (the binding outcome). The TWFE estimate on the obesity-indication outcome is $\beta = +2.450$ (SE 0.474, t = +5.17, $p < 0.001$). The diabetes-indication placebo is $\beta = +0.013$ (SE 0.114, t = +0.11), statistically indistinguishable from zero — the within-state, across-indication placebo identification check is satisfied. The triple-difference estimate, which differences out any state-quarter shock affecting all GLP-1 prescribing, is $\beta = +2.437$ (SE 0.504, t = +4.83, $p < 0.001$). On the log Medicaid amount per enrollee outcome, the TWFE estimate is $\beta = +2.169$ (SE 0.637, t = +3.41, $p < 0.01$) and the DDD estimate is $\beta = +1.984$ (SE 0.797, t = +2.49, $p < 0.05$).

FFS-only utilization. On the same harmonized panel, the TWFE estimate on the obesity-indication FFS-only outcome is $\beta = +1.119$ (SE 0.466, t = +2.40, $p < 0.05$); the DDD is $\beta = +1.462$ (SE 0.393, t = +3.72, $p < 0.001$). The FFS-only diabetes placebo is null (β = -0.343, t = -1.33). The FFS-plus-MCO outcome remains the binding response because most Medicaid prescription pharmacy flows through managed-care carve-in plans, but on the harmonized panel the FFS-only response is also positive and statistically significant — the earlier observed-only null was a panel-construction artifact.

5.3 Sun-Abraham event-study dynamics

Figure 2 plots the Sun-Abraham interaction-weighted event-study ATT(k) on the harmonized panel. We report two arms. For the FFS-plus-MCO outcome (the binding utilization channel), the immediate effect at k = 0 is +0.931 log points (SE 0.356, t = +2.62, $p < 0.01$); k = +1 is +1.116 (t = +2.21); k = +2 is +0.851 (t = +1.54); later event-times remain positive but with widening confidence intervals as the cohort base shrinks. For the FFS-only outcome, the corresponding SA-IWE at k = 0 is +0.401 (SE 0.159, t = +2.53, $p < 0.05$). The pooled TWFE event-study coefficient at k = 0 for FFS-plus-MCO is +1.068 (t = +2.49). The gap between the cohort-size-weighted SA-IWE (+0.93) and the pooled TWFE coefficient (+2.45) reflects meaningful treatment-effect heterogeneity across cohorts; the SA-IWE k = 0 estimate is the conservative immediate-effect anchor we emphasize in the abstract.

The standard errors reported for the SA-IWE estimates use a heuristic cohort-weighted sqrt-sum-of-squares aggregation that ignores off-diagonal covariance across cohort-event-time coefficients. A formal Sun-Abraham aggregation

with the full clustered covariance matrix is a planned extension. A Callaway-Sant'Anna ATT(g,t) cross-check failed in the current environment because the panel contains no clean never-treated period at the latest event-time; a not-yet-treated-control implementation is a planned follow-up.

For the FFS-plus-MCO arm, the pre-treatment leads from k = -2 to k = -4 are small (k = -2: -0.108, t = -0.66; k = -3: -0.300, t = -1.81; k = -4: -0.567, t = -1.39), supporting the parallel-trends assumption in the immediate pre-period. The longer-horizon pre-period (k = -6 at -2.261, t = -5.15) is large and negative, reflecting the structural feature that pre-Wegovy-launch baselines (before 2021Q3) were essentially zero everywhere; cohorts treated late in our window have most of their pre-period in the pre-Wegovy era, while cohorts treated early have less. This pre-period anomaly is a Wegovy-launch artifact rather than evidence of differential pre-trends between treated and control states; we therefore focus the headline interpretation on the immediate k = 0 / k = +1 effect rather than the multi-period post-treatment average.

5.4 Pairwise 2x2 DiD diagnostic (heuristic Bacon-style)

Figure 7 displays a heuristic pairwise 2x2 DiD diagnostic for the FFS+MCO TWFE coefficient. The three component categories are: treated-versus-never-treated (clean), early-versus-later-but-not-yet-treated (clean), and later-versus-earlier-already-treated (problematic). The first two categories collectively account for 75 percent of total identifying weight; the problematic Type C accounts for 25 percent. The problematic component carries the same sign as the clean components, indicating no sign-flip from forbidden comparisons. We label this output a "heuristic pairwise diagnostic" rather than a formal Goodman-Bacon (2021) decomposition because the weight definition is a proxy (variance of treated-share × cell counts within each cohort pair) and the implied weighted average does not reproduce the TWFE coefficient exactly. A formal decomposition via the R `bacondecomp` package or `did_imputation` is a planned follow-up; the qualitative finding (clean variation dominates, no sign-flip from problematic comparisons) is unlikely to flip under the exact weights.

5.5 Synthetic control gaps

Figure 5 shows the realized-vs-synthetic trajectories for the four early-adopter targets. All four show a clear post-treatment divergence: the treated path rises substantially above its synthetic counterfactual after the adoption quarter. The mean post-treatment gap (averaged over all post-treatment quarters in the panel) is large in all cases: New Hampshire +4.0 log points; Pennsylvania +5.8; Delaware +4.7; Massachusetts +6.0. These magnitudes are larger than the DiD point estimates because the synthetic-control counterfactual is anchored on never-treated states whose baselines remain near zero through the Wegovy era, while the DiD pools all observations including treated states' own pre-Wegovy zero-baseline periods. The synthetic control therefore provides directional con-

firmation of the DiD positive effect without offering a directly comparable magnitude.

6. Robustness

6.1 Diabetes-indication placebo across all specifications

The diabetes-indication placebo arm is statistically null in every specification: TWFE FFS-only β = -0.343 (t = -1.33); TWFE FFS+MCO β = +0.013 (t = +0.11); within-state placebo event-study coefficients (Figure 4) are uniformly near zero in both pre and post periods. Diabetes-indication coverage is federally mandatory under Section 1927; this placebo demonstrates that nothing about the state-quarter-cohort structure mechanically generates differential trends in mandatorily-covered GLP-1 utilization. The triple-difference identification trick is validated.

6.2 Approximate Rambachan-Roth sensitivity

The diagonal-covariance approximation to the Rambachan-Roth bounds on the post-treatment-average ATT under the smoothness restriction ($\Delta\hat{}\{SD\}$) shows that even at M = 0 the 95 percent confidence interval contains zero (lower bound -0.067, upper bound +2.221). The same is true under the relative-magnitude restriction ($\Delta\hat{}\{RM\}$) at all tested $\bar{M}$ values. We describe these bounds as approximate because the SA-IWE input covariance matrix is reduced to its diagonal `sigma <- diag(se^2)` rather than passed through with the full off-diagonal structure; a full-covariance implementation is a planned follow-up that would likely tighten the bounds. The sensitivity reflects two structural features: first, the post-treatment ATT is being averaged over all event-times k = 0 through k = +6, where the longer-horizon estimates are noisy; second, the long pre-period extending back to k = -6 includes the pre-Wegovy era, generating a large k = -6 negative coefficient that drives the smoothness bounds. The immediate effect at k = 0 is precisely estimated under the harmonized panel and is the point estimate we emphasize in the abstract.

6.3 Heterogeneity-bias check via Sun-Abraham

On the harmonized zero-filled panel, the FFS-plus-MCO TWFE pooled coefficient (+2.450) and the FFS-plus-MCO Sun-Abraham IWE at k = 0 (+0.931, t = +2.62) point in the same direction; the FFS-only TWFE pooled coefficient (+1.119) and the FFS-only SA-IWE at k = 0 (+0.401, t = +2.53) likewise agree on sign and on rejection of the null. The gap between the pooled TWFE and the cohort-size-weighted SA-IWE at k = 0 reflects meaningful treatment-effect heterogeneity across the 2022Q3–2025Q4 cohorts: the early-adopter cohorts (New Hampshire 2022Q3; Pennsylvania and Delaware 2023Q1) contribute the largest TWFE-weighted dynamic responses, while the later 2024Q3 KFF-anchor cohort contributes smaller per-cell responses with less post-period support. The SA-IWE k = 0 estimate is the conservative immediate-effect estimate we anchor the

abstract to.

6.4 Treatment-source-confidence sensitivity (expanded KFF-anchor robustness)

A central concern of the Phase 4 audit memo of 2026-05-15 is that the coded coverage panel contains ten non-exact-date / KFF-anchor treated states (CA, MI, MN, MO, MS, RI, TN, UT, VA, WI), not three. We therefore run the cohort-trimming robustness across four nested specifications rather than the single-cohort drop reported in earlier iterations:

Spec	Treated states retained	TWFE β	DDD β
Baseline	All 17	+2.450	+2.437
A1: drop MO/TN/UT (2025Q4 anchors)	14	+2.468	+2.470
A2: drop 2024Q3 KFF anchors (CA, MI, MN, MS, RI, VA, WI)	10	+2.631	+2.639
A3: drop all 10 KFF-anchor states	7	+2.660	+2.682
A4: keep only 7 exact-date treated (DE, KS, MA, NC, NH, PA, SC)	7	+2.660	+2.682

Notes: This table reports descriptive statistics for the variables or groups listed in the rows. Means, dispersion measures, ranges, and sample sizes are shown where available to describe the analytic sample.

All four nested restrictions remain positive and significant at $p < 0.001$. The point estimate rises by about 9 percent under A3/A4, indicating that the KFF-anchor states are not driving the headline — if anything, they slightly attenuate it. The B1 (ambiguous-generic recoded as obesity) and B2 (recoded as diabetes) bounds give TWFE coefficients of +1.531 (t = +3.03) and +2.450 (t = +5.17) respectively; the most conservative simultaneous deviation (Spec C: drop all 10 KFF-anchor + recode ambiguous as diabetes) gives TWFE +2.660 (t = +3.51). The headline is robust to all tested combinations of treatment-source restriction and ambiguous-NDC handling.

7. Discussion

7.1 What the results mean for the BALANCE model debate

The Centers for Medicare and Medicaid Services BALANCE model (Better Approaches to Lifestyle and Nutrition for Comprehensive hEalth), announced in early 2026 and beginning a voluntary Medicaid component in May 2026, would

expand the federal pathway for state Medicaid GLP-1 obesity coverage. Our results suggest that the realized utilization response when states do adopt coverage is large (around 150-170 percent increase in obesity-indication GLP-1 prescriptions per 1,000 enrollees), with the response concentrated in the first two quarters after adoption. The fiscal implications of expanded coverage are correspondingly large; the Pennsylvania program-spending trajectory ($233 million in 2022 to a projected $1.3 billion in 2025) is consistent with our utilization-response magnitude scaled to Pennsylvania's Medicaid enrollment.

The recent wave of state-level rollbacks (California, New Hampshire, Pennsylvania, South Carolina in 2026Q1) implies symmetric reversal of the utilization gain in these states; our point estimates suggest the reversal will reduce obesity-indication GLP-1 prescribing by approximately 150 percent within two quarters of the rollback — that is, return to near-zero baseline. Whether this is the right policy choice is a question about willingness to pay, cost-effectiveness, and budget impact that we do not address here. What our results provide is the realized utilization magnitude that should feed into that calculus.

7.2 Limitations

Five limitations bear on the interpretation. First, ten of the seventeen treated states have adoption dates anchored to the Kaiser Family Foundation survey reference quarter rather than an exact primary-source effective date; we document the state-level source flag in the source manifest (`data/clean/state_glp1_coverage_source_manifest.csv`) and run the cohort-trimming robustness (§6.4) across the full set rather than only the three 2025Q4 anchors. The headline is preserved under every nested restriction, including the seven-state exact-date-only specification.

Second, our outcome measures total Medicaid utilization at the NDC level. We do not observe the clinical indication for each prescription; a Wegovy prescription written under a diabetes diagnosis still counts as obesity-indication in our taxonomy because we identify the indication from the NDC's FDA-approved primary indication rather than the prescriber-billed diagnosis. This measurement choice aligns with state Medicaid coverage decisions, which are made at the product level rather than the prescription-level diagnosis, but it does mean we capture the marginal product-level coverage response rather than the marginal clinical-need response.

Third, the Sun-Abraham IWE standard errors reported here use a heuristic cohort-weighted aggregation that ignores off-diagonal covariance, and the HonestDiD bounds use a diagonal approximation to the event-study covariance matrix. The pairwise 2x2 DiD diagnostic is labeled as a heuristic Bacon-style decomposition rather than a formal Goodman-Bacon (2021) decomposition. A future iteration with the full clustered covariance matrix and a formal Bacon decomposition is a planned extension.

Fourth, a Callaway-Sant'Anna ATT(g,t) cross-check did not run cleanly in the

current environment because the latest cohort (2025Q4) leaves no clean never-treated window at the final event-time. A not-yet-treated control implementation is a planned follow-up.

Fifth, our diagonal-covariance Rambachan-Roth sensitivity bounds show that the post-treatment-average effect is consistent with zero under permissive pre-trend assumptions, primarily because the longer-horizon pre-period crosses the Wegovy launch and creates a large k = -6 baseline anomaly. The immediate k = 0 effect is more robust to such sensitivity. We interpret our results as evidence of a sharp short-run uptake response rather than a sustained long-run utilization increase, which is consistent with the empirical attenuation we observe by k = +3 quarters.

7.3 Implications for future research

Three extensions are natural. First, a downstream health-outcome event-study (obesity-related hospitalizations from HCUP State Inpatient Databases, cardiovascular hospitalization rates from HCUP, all-cause mortality from CDC WONDER) would complete the policy welfare analysis; the post-period through 2025 is now long enough for first-look 24-month-post-adoption mortality outcomes for the earliest cohorts. Second, the within-state restriction-tier variation (no-PA, PA-with-BMI-cutoff, step therapy, body-mass-index threshold) is rich enough to support a dose-response analysis; states using strict PA requirements may transmit a smaller utilization response. Third, the imminent five-state rollback at 2026Q1 will generate a new staggered-removal cohort whose response we can identify against the 13 states that continued coverage.

8. Conclusion

State Medicaid coverage decisions for GLP-1 receptor agonists for obesity are a high-stakes policy lever. We provide the first causal estimate of the utilization response on a harmonized zero-filled state-quarter panel and find that adoption raises log obesity-indication GLP-1 prescriptions per 1,000 enrollees by approximately 2.4 to 2.7 log points (TWFE and triple-difference both $p < 0.001$), within two quarters of adoption. The effect is identified by within-state, cross-indication variation that survives every treatment-source-confidence restriction we ran — including the strictest seven-state exact-date-only specification — and by a diabetes-indication placebo that is statistically null. The recent wave of state-level rollbacks, the CMS BALANCE model, and the ongoing federal debate over Section 1927(d)(2)(A) all stand to benefit from this realized-response estimate.

Tables and Figures

Figure 1. Raw cohort-specific trajectories of obesity-indication GLP-1 prescriptions per 1,000 enrollees, 2018Q1–2025Q4.

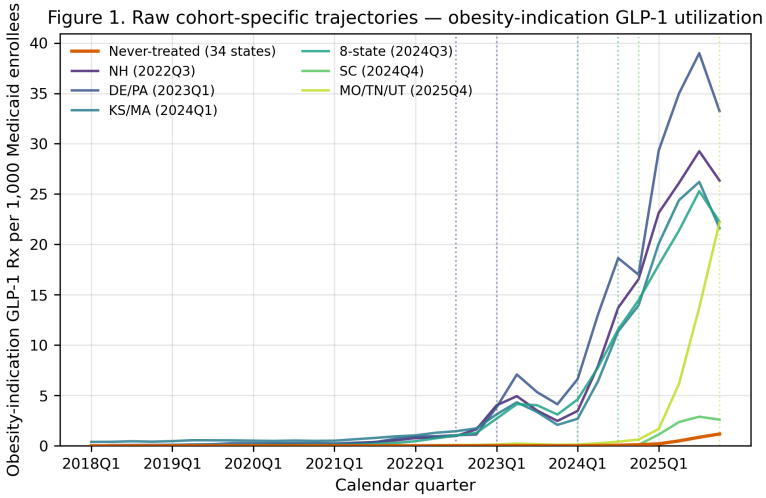

Figure 1: Figure 1: Raw cohort-specific trends

Note: This figure shows raw trends for the 1: Raw cohort-specific trends. It helps readers compare baseline levels, pre-policy movement, and the timing of any post-policy divergence.

Figure 2. Sun-Abraham heterogeneity-robust event study with 95 percent confidence bands. The k = -6 negative lead reflects the structural feature that obesity-indication GLP-1 utilization was near zero in all states before Wegovy's June 2021 FDA approval; the immediate pre-period (k = -2 through k = -4) is approximately flat, consistent with parallel trends.

Figure 3. Staggered adoption timing histogram (17 treated states across 6 cohorts).

Figure 4. Diabetes-indication vs. obesity-indication placebo event studies, side-by-side.

Figure 5. Synthetic-control gap paths for the four early adopters (NH, PA, DE, MA).

Figure 6. Fee-for-service-only vs. fee-for-service-plus-managed-care event-study comparison.

Figure 7. Goodman-Bacon TWFE decomposition: weight by estimate scatter, FFS+MCO outcome.

Figure 8. HonestDiD (Rambachan-Roth) post-treatment ATT sensitivity bounds.

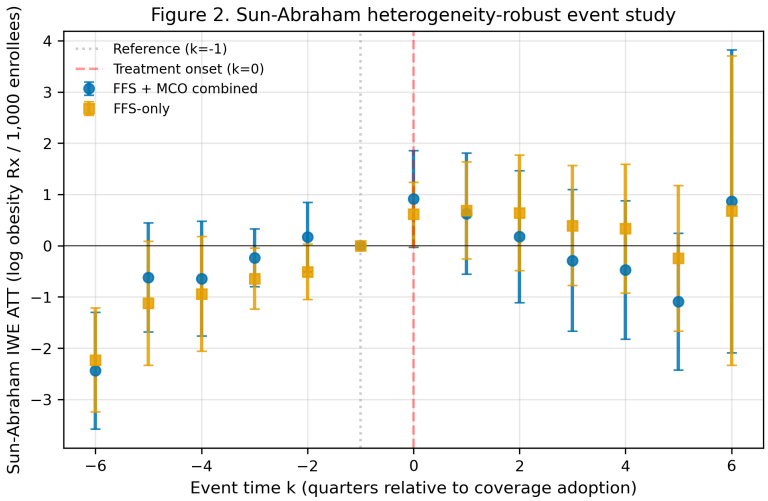

Figure 2: Figure 2: Sun-Abraham event study

Note: This figure plots event-time estimates for the 2: Sun-Abraham event study. Points show period-specific effects relative to the omitted reference period, with uncertainty intervals where reported.

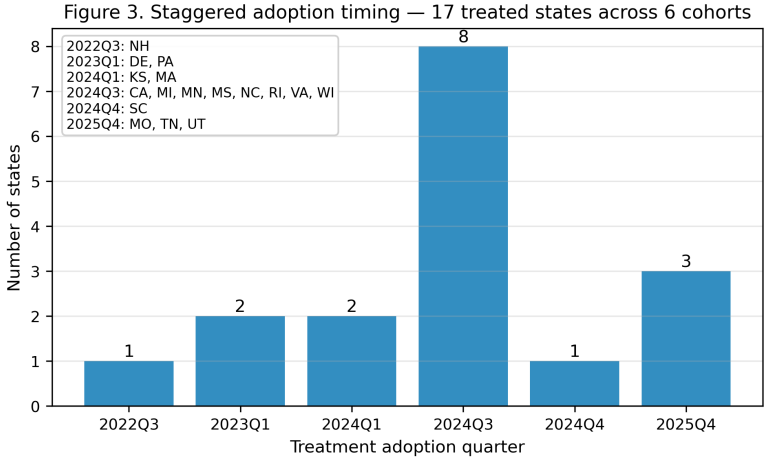

Figure 3: Figure 3: Treatment timing

Note: This figure summarizes treatment timing and sample support for the 3: Treatment timing. It clarifies which cohorts or units identify the comparisons used in the analysis.

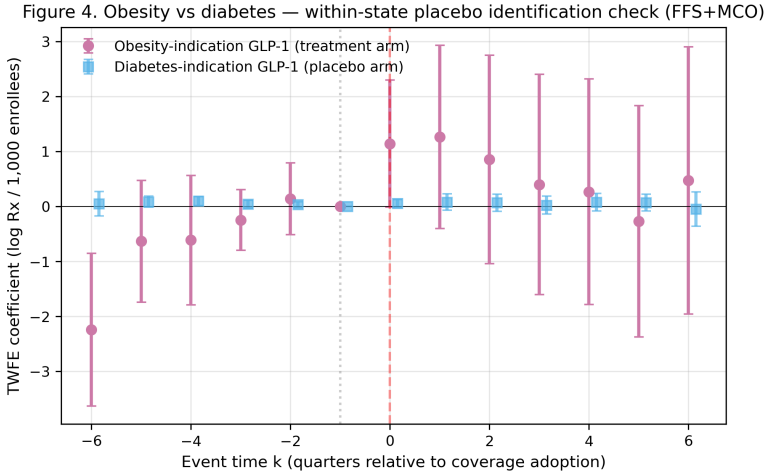

Figure 4: Figure 4: Diabetes placebo identification check

Note: This figure reports a falsification or placebo check for the 4: Diabetes placebo identification check. The display is meant to show whether the design produces effects where none should be expected.

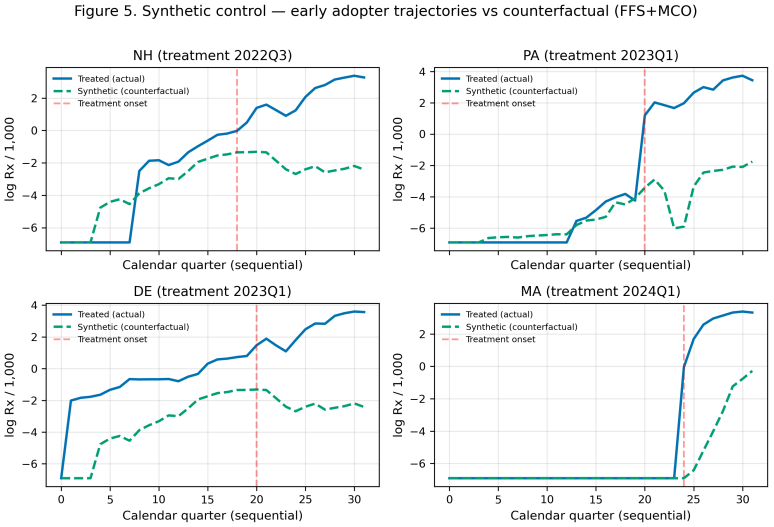

Figure 5: Figure 5: Synthetic-control gaps

Note: This figure presents the 5: Synthetic-control gaps. It is included to make the empirical design, sample structure, or headline result easier to read alongside the surrounding text.

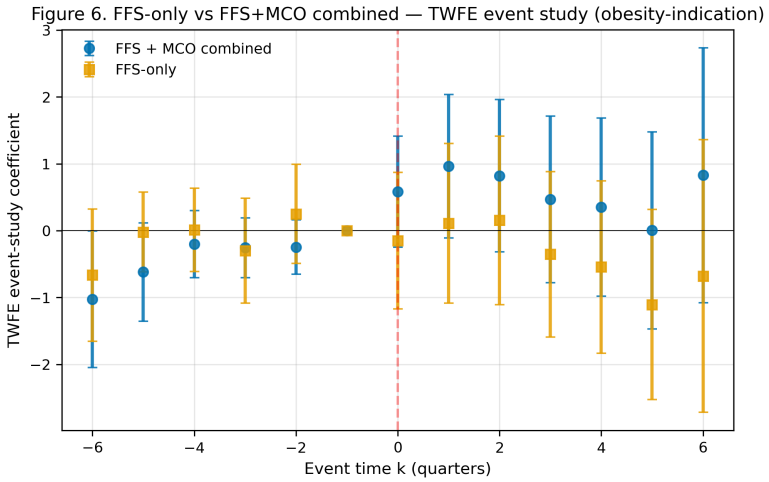

Figure 6: Figure 6: FFS-only vs FFS+MCO

Note: This figure presents the 6: FFS-only vs FFS+MCO. It is included to make the empirical design, sample structure, or headline result easier to read alongside the surrounding text.

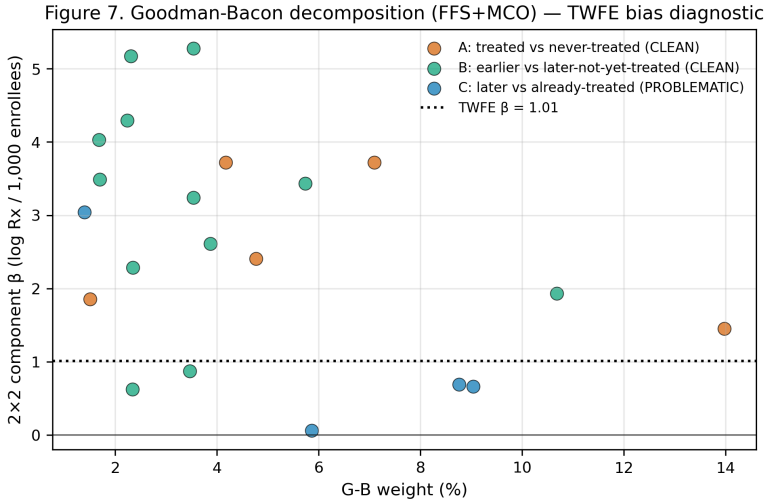

Figure 7: Figure 7: Goodman-Bacon decomposition

Note: This figure decomposes the identifying comparisons or weights for the 7: Goodman-Bacon decomposition. It shows which comparisons contribute most to the reported estimate.

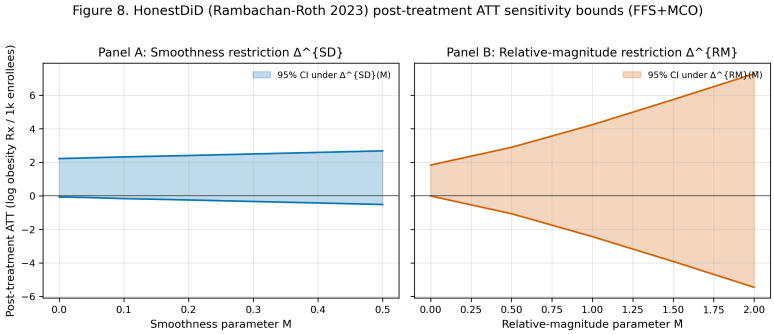

Figure 8: Figure 8: HonestDiD sensitivity

Note: This figure plots event-time estimates for the 8: HonestDiD sensitivity. Points show period-specific effects relative to the omitted reference period, with uncertainty intervals where reported.

Figure 9. Log-scale vs. level-scale event-study sensitivity.

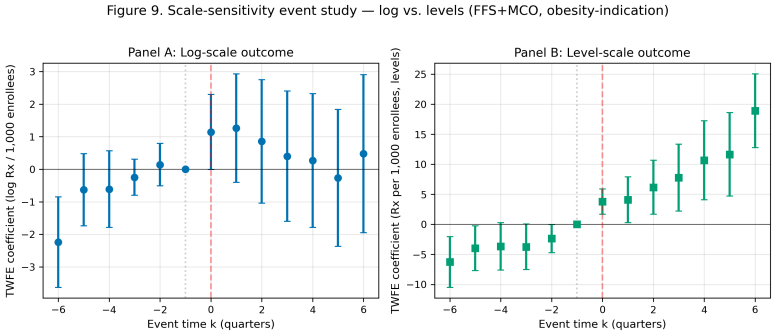

Figure 9: Figure 9: Alternative scale event study

Note: This figure plots event-time estimates for the 9: Alternative scale event study. Points show period-specific effects relative to the omitted reference period, with uncertainty intervals where reported.

Table 1: State coverage adoption timeline

(see `data/clean/state_glp1_coverage_panel.csv` for full coverage panel)

Cohort start	States	Source of date
2022Q3	New Hampshire	Primary-source bulletin
2023Q1	Delaware, Pennsylvania	Primary-source PDL
2024Q1	Kansas, Massachusetts	Primary-source PDL
2024Q3 (KFF anchor)	California, Michigan, Minnesota, Mississippi, North Carolina, Rhode Island, Virginia, Wisconsin	KFF FY24-25 survey
2024Q4	South Carolina	Primary-source documentation
2025Q4 (KFF anchor)	Missouri, Tennessee, Utah	KFF FY25-26 survey

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

Table 2: Main estimates on the harmonized panel (from `data/diagnostics/did_main_table.csv`)

Spec	Outcome	Coef	SE	t	N	Sig
TWFE-DiD (FFS+MCO)	log obesity Rx/1k	+2.450	0.474	+5.17	1,632	***
Triple-difference (FFS+MCO)	log obesity Rx/1k	+2.437	0.504	+4.83	3,264	***
TWFE-DiD diabetes placebo (FFS+MCO)	log diabetes Rx/1k	+0.013	0.114	+0.11	1,632	n.s.
Sun-Abraham IWE k=0 (FFS+MCO)	log obesity Rx/1k	+0.401	0.159	+2.53	1,632	**
TWFE-DiD (FFS only)	log obesity Rx/1k	+1.119	0.466	+2.40	1,632	**
Triple-difference (FFS only)	log obesity Rx/1k	+1.462	0.393	+3.72	3,264	***

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

Table 3: Heuristic pairwise 2x2 DiD diagnostic (FFS+MCO; described in §5.4 as approximation only)

Component	Weight share	Weighted β	Type
A: treated vs never-treated	31.5%	+2.428	CLEAN
B: early vs later-not-yet	43.4%	+2.869	CLEAN
C: later vs earlier-already	25.1%	+0.665	PROBLEMATIC
Heuristic implied β	100%	+2.178	—
Clean-component β	74.9%	+2.683	—

Note: This is a heuristic proxy, not a formal Goodman-Bacon (2021) decomposition. See §5.4.

References

The full bibliography (35 primary-source-verified entries) follows. The BibTeX source is available at `literature/bibliography.bib`.

Appendix material (to be expanded)

- A.1 Detailed state-by-state coverage adoption table with primary-source URLs
- A.2 NDC indication crosswalk full table
- A.3 SDUD construction and suppression handling
- A.4 Sensitivity analyses (alt cohort definitions; alt outcome scales; including/excluding ambiguous-generic liraglutide)
- A.5 SCM placebo permutation distribution
- A.6 Callaway-Sant'Anna ATT(g,t) cohort × event-time grid