The Safety Net Holds: Null Effects of Medicaid Unwinding on Behavioral Health Provider Markets

APEP Autonomous Research[*]   @ai1scl-auto

February 19, 2026

Abstract

The 2023 Medicaid unwinding removed over 25 million people from coverage. We estimate the differential impact on behavioral health providers using a triple-difference design comparing behavioral health (H-code) versus home and community-based services (T-code) providers, before versus after unwinding, across states with staggered start dates. Using T-MSIS fee-for-service claims (2018–2024), we find no evidence of differential harm. The DDD coefficient on log spending is -0.020 ($SE = 0.096$, $p = 0.836$), with uniformly null results across claims, provider counts, exit rates, and market concentration. Randomization inference confirms the null ($p_{\text{RI}} = 0.834$). Despite strong theoretical reasons to expect differential vulnerability, FFS Medicaid billing patterns show no measurable disruption. This well-identified null suggests provider markets possess greater resilience to demand shocks than standard theory predicts.

JEL Codes: I11, I13, I18, H75
Keywords: Medicaid unwinding, behavioral health, provider supply, triple-difference, null result, disenrollment

---

1. Introduction

In the eighteen months following April 2023, twenty-five million Americans—one in every four Medicaid enrollees—lost their health insurance. The scale of the Medicaid unwinding was unprecedented: the largest mass disenrollment in the program's sixty-year history. States processed more than 90 million eligibility redeterminations after the continuous enrollment provision expired (**?**). The policy question of who bears the cost of this coverage disruption has been the subject of intense public debate. But nearly all of that debate has focused on the patients who lost coverage. This paper asks a different question: what happened to the providers who lost their patients?

The answer matters because providers are the infrastructure of the safety net. When a community mental health center closes, the loss extends far beyond its current patients. Future patients—including those who regain Medicaid coverage or obtain marketplace insurance—face longer wait times, longer travel distances, and reduced access to care (**?**). If the unwinding permanently reduced behavioral health provider capacity, the damage may outlast the coverage disruption itself. Conversely, if the provider safety net absorbed the unwinding without measurable disruption, that resilience is itself an important finding—it would suggest that provider markets possess greater capacity to weather demand shocks than standard models predict.

We study this question using administrative claims data from the Transformed Medicaid Statistical Information System (T-MSIS), a universe-level dataset covering all fee-for-service Medicaid provider spending across 51 states and territories from January 2018 through October 2024. The dataset contains approximately 227 million claim-line-level records, which we aggregate to a balanced panel of 8,364 state-category-month observations, allowing us to track billing volumes, active provider counts, entry, exit, and market concentration at fine geographic and temporal resolution. We link these records to the National Plan and Provider Enumeration System (NPPES) to assign providers to states, achieving a 99.5 percent match rate.

Our identification strategy exploits a natural experiment with three dimensions of variation. The first is time: the continuous enrollment provision created a clear break between a period of guaranteed coverage (March 2020 through March 2023) and a period of active disenrollment (April 2023 onward). The second is cross-state variation: states began their unwinding processes on staggered timelines between April and July 2023, and their disenrollment rates ranged from 12 percent in North Carolina to 57 percent in Montana (**?**). The third is within-state variation across service types: behavioral health providers (identified by H-prefix HCPCS codes) serve a fundamentally different patient population than home and community-

based services providers (identified by T-prefix codes), even within the same state and time period.

We implement a triple-difference (DDD) estimator that compares changes in outcomes for behavioral health providers relative to HCBS providers, before versus after the unwinding, across states with different start dates and disenrollment intensities. The key advantage of this design is that the within-state comparison between H-code and T-code providers absorbs state-level confounders—economic conditions, political environment, Medicaid generosity, COVID recovery trajectories—that might otherwise bias a simple before-after or cross-state comparison. Prior work in this research program found that HCBS providers were "far more resilient than anticipated" to the unwinding (**?**), establishing them as a valid within-state control group.

The theoretical motivation for expecting differential effects is straightforward. HCBS providers serve long-term Medicaid enrollees—predominantly elderly and disabled individuals who receive home health, personal care, and community-based waiver services. These populations were largely exempt from redetermination because they qualified through disability or aged-blind-disabled pathways with infrequent eligibility reviews (**?**). Behavioral health providers, by contrast, serve episodic patients—people with mental health conditions, substance use disorders, and co-occurring diagnoses who cycle in and out of Medicaid coverage as their circumstances change. These patients are disproportionately enrolled through expansion pathways or income-based eligibility, making them prime candidates for procedural disenrollment when states resumed paperwork requirements (**?**).

The financial exposure of behavioral health providers amplifies this expected vulnerability. Community mental health centers derive 80 to 90 percent of their revenue from Medicaid (**?**). Unlike physicians who bill standard CPT codes and can substitute toward Medicare or commercial patients when Medicaid enrollment falls, behavioral health providers who bill H-codes—community psychiatric support, psychosocial rehabilitation, substance use counseling—have no Medicare equivalent for most of these services (**?**). They are, in the language of industrial organization, single-payer dependent with no outside option. A demand shock from Medicaid disenrollment should, in theory, hit them with nearly the full force of lost revenue.

Our main results, however, do *not* confirm this expected differential vulnerability. Despite the strong theoretical priors, we find no statistically significant evidence that behavioral health providers were differentially harmed by the unwinding relative to HCBS providers. The main DDD coefficient on log spending is -0.020 ($SE = 0.096$, $p = 0.836$), an economically small point estimate that is statistically indistinguishable from zero. The 95 percent confidence interval is wide, spanning $[-0.208, 0.168]$, reflecting the imprecision inherent in state-level

variation with 51 clusters. This null finding is not an artifact of a single specification: across every outcome we examine—claims (-0.008, $p = 0.909$), active provider counts (0.003, $p = 0.921$), exit rates (0.002, $p = 0.776$), net entry (0.007, $p = 0.440$), and market concentration (-0.039, $p = 0.590$)—the DDD estimates are small and statistically insignificant.

The dose-response specification, which tests whether states with more intense disenrollment experienced larger differential behavioral health effects, is likewise null. The intensity interaction coefficient is -0.811 ($p = 0.37$), and the main post-treatment-times-BH term in the intensity specification is 0.286 ($p = 0.36$). States with higher disenrollment rates did not exhibit detectably larger differential declines in behavioral health spending relative to HCBS.

The event study provides visual evidence consistent with the null. Pre-trend coefficients are positive (ranging from approximately 0.12 to 1.07) but statistically insignificant (all $p > 0.28$), with no systematic pattern and wide confidence intervals that comfortably include zero, supporting the parallel trends-in-trends assumption. Post-treatment coefficients drift negative over time, reaching point estimates that are large in magnitude (approximately -2.12 at $k = 18$), but all are estimated with enormous uncertainty (standard errors of 1.0–2.8) and none is individually statistically significant. The point estimates at longer horizons are large, but the confidence intervals are so wide that the data cannot distinguish the observed pattern from zero.

Randomization inference provides nonparametric confirmation of the null. We randomly permute state unwinding start dates 500 times and re-estimate the DDD specification under each permutation. The observed test statistic is well within the body of the permutation distribution ($p_{\mathrm{RI}} = 0.834$), failing to reject the null of no differential effect. Pre-trend falsification tests confirm the validity of the research design: a placebo test assigning a "fake" post-treatment period yields a coefficient near zero (0.018, $p = 0.837$), and a placebo replacing behavioral health with CPT-code providers yields a similarly null result (0.022, $p = 0.825$).

We emphasize that this null result is not a failure of the research design. The identification strategy has three credible sources of variation, the pre-trends are clean, the sample is large (8,364 state-category-month observations), and the comparison group is well-motivated by theory and prior evidence. Rather, the null reflects a genuine empirical finding: despite strong theoretical reasons to expect differential vulnerability, behavioral health providers were *not* measurably more disrupted than HCBS providers by the Medicaid unwinding.

This paper contributes to four literatures. First, we contribute to the rapidly growing literature on the consequences of the Medicaid unwinding. **?** document the scale and demographic characteristics of disenrollment. **?** and **?** describe the administrative processes states used and their consequences for coverage continuity. **?** analyze the legal framework governing redeterminations. Our contribution is to shift the focus from the demand side (who

lost coverage) to the supply side (which providers lost revenue and market presence)—and to report the surprising finding that the supply side was resilient.

Second, we contribute to the literature on provider supply responses to insurance coverage changes. **?** show that Medicaid expansions under the ACA increased provider participation, particularly in primary care. **?** and **?** document that Medicaid acceptance rates are lower than for other insurance types and vary significantly across specialties. **?** finds that psychiatrists have among the lowest Medicaid acceptance rates of any specialty. Our paper provides evidence on the reverse channel: what happens to provider supply when coverage contracts rather than expands. The null result is informative about the degree of market frictions—it suggests that provider exit is not an immediate or proportional response to enrollment declines, potentially because of adjustment costs, contract stickiness, or the ability of providers to absorb temporary demand shocks through reduced margins or increased non-Medicaid volume (**?**).

Third, we contribute to the literature on behavioral health provider markets and mental health parity. **?** and **?** document the unique financial vulnerability of community mental health providers. **?** map geographic disparities in mental health provider access. **?** show that mental health workforce shortages are concentrated in areas that depend most heavily on public insurance. **?** argue that Medicaid is the single most important payer for behavioral health services in the United States. Our evidence that behavioral health providers were *not* differentially disrupted—despite their theoretical vulnerability—complicates the narrative of behavioral health fragility and suggests that these markets may possess more resilience than commonly assumed.

Fourth, we contribute to the methodological literature on triple-difference designs and causal inference with staggered policy adoption. Our DDD specification follows the framework articulated by **?** and **?**, using the within-state comparison group to absorb confounders that would threaten a simple DiD. We apply randomization inference (**??**) as a finite-sample complement to asymptotic inference with clustered standard errors (**?**). The paper also illustrates the value of reporting well-identified null results: the clean research design makes the absence of an effect informative rather than merely inconclusive.

2. Institutional Background and Policy Setting

2.1 The Medicaid Continuous Enrollment Provision

The Families First Coronavirus Response Act (FFCRA), enacted in March 2020, offered states a 6.2 percentage point increase in their Federal Medical Assistance Percentage (FMAP) in exchange for a maintenance-of-effort (MOE) requirement: states could not disenroll Medicaid

beneficiaries during the public health emergency (PHE) (**?**). This continuous enrollment condition (CEC) was unprecedented in the program's history. For more than three years, Medicaid enrollment could only grow—beneficiaries who would otherwise have lost eligibility due to income changes, aging out, relocation, or failure to complete renewal paperwork remained on the rolls.

The consequence was a massive expansion of the Medicaid population. Total enrollment grew from 71 million in February 2020 to a peak of 94.3 million in April 2023—a 33 percent increase (**?**). This growth was not evenly distributed. States that had expanded Medicaid under the ACA saw larger enrollment gains, as did states with historically high churn rates. The continuous enrollment provision did not merely prevent loss of coverage for existing beneficiaries; it also prevented the exit of new enrollees who signed up during the pandemic but would have lost eligibility within months under normal operations.

For providers, the CEC was a sustained demand boost. With more beneficiaries maintaining coverage, the pool of patients who could present with active Medicaid insurance grew continuously for three years. Behavioral health providers, who experienced a surge in demand during the pandemic as mental health conditions worsened (**??**), benefited from both the increased need for services and the expanded insurance coverage to pay for them. The unwinding threatened to reverse both channels simultaneously.

2.2 The Unwinding Process

The Consolidated Appropriations Act of 2023 (CAA), signed in December 2022, decoupled the continuous enrollment provision from the PHE declaration and set April 1, 2023 as the date when states could begin disenrolling beneficiaries. The CMS issued guidance requiring states to complete all pending redeterminations within 14 months (by May 2024, later extended for some states) and to prioritize ex parte renewals—automated eligibility checks using available data—before initiating manual renewal processes (**?**).

States varied dramatically in how they approached the unwinding. The first dimension of variation was timing. While the earliest states began sending renewal notices and processing terminations in April 2023, others delayed until May, June, or July 2023. This staggered rollout creates the cross-state timing variation that anchors our identification strategy. The second dimension was administrative approach. Some states invested heavily in ex parte renewal systems, updated contact information databases, and extended renewal periods. Others relied on bulk mailings of renewal packets and processed terminations quickly when packets went unreturned.

The third and most consequential dimension of variation was the resulting disenrollment rate. By January 2024, the share of Medicaid enrollees who had been disenrolled ranged from

12 percent in North Carolina (which invested heavily in targeted outreach) to 57 percent in Montana (which processed terminations rapidly with limited outreach) (**?**). Critically, a large share of these disenrollments were *procedural*—people who were terminated not because they were found ineligible, but because they failed to return paperwork, could not be reached at their address of record, or did not respond to renewal notices within the required window. Nationally, roughly 69 percent of all terminations through mid-2024 were procedural rather than eligibility-based (**?**).

2.3 Behavioral Health Provider Markets

Behavioral health services in Medicaid are primarily billed using H-prefix Healthcare Common Procedure Coding System (HCPCS) codes. These codes cover community-based psychiatric services (H0015–H0050), psychosocial rehabilitation (H2000–H2037), substance use disorder treatment (H0001–H0015), and crisis intervention services. H-codes are specific to state Medicaid programs—they have no Medicare equivalent and are rarely used by commercial insurers (**?**).

This coding structure creates a natural financial isolation. Providers who bill predominantly H-codes cannot easily shift their patient mix toward Medicare or commercial payers when Medicaid enrollment falls. A psychiatrist who bills standard CPT evaluation and management codes can accept Medicare patients; a community support worker who bills H0036 (community psychiatric supportive treatment) cannot, because the service category does not exist in the Medicare fee schedule. This payer lock-in is the fundamental economic distinction between behavioral health providers and most other Medicaid providers.

The behavioral health workforce is also structurally different from other health care sectors. Community mental health centers (CMHCs) employ a mix of licensed clinicians (psychiatrists, psychologists, licensed clinical social workers) and paraprofessional staff (community health workers, peer support specialists, case managers). Many of these staff hold certifications that are recognized by Medicaid but not by other payers. The workforce is geographically concentrated in underserved areas, reflecting CMHCs' historical mission as safety-net providers (**?**).

Financial margins in behavioral health are thin. SAMHSA survey data indicate that the median CMHC operates with net margins below 3 percent, and approximately one-quarter of CMHCs operate at a loss in any given year (**?**). Medicaid constitutes between 47 and 90 percent of total revenue depending on the state and facility type (**?**). This financial fragility means that even a moderate reduction in Medicaid patient volume can push providers from solvency to distress.

2.4 Why Behavioral Health Differs from HCBS

The comparison between behavioral health (H-code) and home and community-based services (T-code) providers is central to our identification strategy. We articulate precisely why these two Medicaid-specific provider types would be differentially affected by the unwinding.

HCBS providers serve populations—elderly individuals receiving home health services, people with intellectual and developmental disabilities receiving community-based waiver services, and people with physical disabilities receiving personal care assistance—that are enrolled in Medicaid through categorical eligibility pathways (aged, blind, disabled) rather than income-based expansion pathways. These enrollees face infrequent redetermination, and many are dually eligible for Medicare and Medicaid, providing a second payer backstop. The prior study in this research program (APEP-0307) confirmed empirically that HCBS providers showed no measurable response to the unwinding, consistent with the institutional prediction that their patient base would be largely unaffected by the redetermination process.

Behavioral health patients, by contrast, are disproportionately enrolled through income-based pathways—particularly the ACA Medicaid expansion, which extended coverage to adults under 138 percent of the federal poverty level. These enrollees face annual or more frequent redetermination. Many have unstable housing, unreliable postal addresses, and cognitive or psychiatric conditions that make it difficult to navigate administrative requirements (**?**). They are, in short, precisely the population most vulnerable to procedural disenrollment—and the population that behavioral health providers depend on most heavily.

3. Data

3.1 T-MSIS Medicaid Provider Spending

Our primary data source is the T-MSIS (Transformed Medicaid Statistical Information System) Medicaid Provider Spending dataset, released by the U.S. Department of Health and Human Services in February 2026. The dataset contains 227 million rows covering all fee-for-service Medicaid provider spending at the provider-service-month level for all 51 states and territories from January 2018 through October 2024. Each record contains the National Provider Identifier (NPI), HCPCS/CPT procedure code, month of service, number of claims, total Medicaid spending, and state of service.

We classify providers into three mutually exclusive categories based on their primary billing codes.[1] *Behavioral health providers* are NPIs whose plurality of claims in a given

---

[1]Provider classification is based on the HCPCS code prefix of billed services (H-codes for behavioral health, T-codes for HCBS), which is determined by the type of service rendered, not by provider characteristics. This classification is inherent to the billing code and does not change based on provider volume or billing patterns,

month are billed under H-prefix HCPCS codes (H0001–H2037). *HCBS providers* are NPIs whose plurality of claims are billed under T-prefix HCPCS codes (T1000–T2048). *Standard medical providers* are NPIs billing primarily under CPT codes (the standard physician fee schedule). This classification follows prior work in the T-MSIS literature and aligns with CMS coding guidelines (**?**).

We aggregate the provider-level records to construct a balanced state-by-service-category-by-month panel. For each state s, service category $c \in \{\mathrm{BH}, \mathrm{HCBS}\}$, and month t, we compute total Medicaid spending, total claim volume, count of active providers (NPIs with at least one claim), count of new entrants (NPIs billing for the first time in that state-category), count of exits (NPIs present in month $t-1$ but absent in month t through $t+2$), and the Herfindahl-Hirschman Index (HHI) of spending concentration.

3.2 NPPES Provider Registry

We use the CMS National Plan and Provider Enumeration System (NPPES) to assign each NPI to a geographic location. The NPPES contains the practice address (including state and ZIP code) for every active NPI in the United States. We merge T-MSIS records to NPPES on the NPI field, achieving a 99.5 percent match rate. The 0.5 percent of unmatched NPIs are predominantly organizational NPIs with recently assigned numbers; we assign these to the state from which they billed the most claims.

For the small number of providers who bill in multiple states, we assign each NPI to the single state where it generated the plurality of its Medicaid claims. This assignment is stable over time: 97.2 percent of providers who appear in multiple months are assigned to the same state throughout the sample period.

3.3 State Unwinding Data

We compile state-level unwinding information from the Kaiser Family Foundation (KFF) Medicaid Unwinding Tracker (**?**). For each state, we code: (1) the month in which the state began processing Medicaid renewals and terminations (the "unwinding start date"), (2) the cumulative disenrollment rate as of January 2024 (our measure of unwinding intensity), and (3) the share of terminations classified as procedural versus eligibility-based.

The unwinding start dates cluster into four monthly cohorts: April 2023 (9 states), May 2023 (24 states), June 2023 (17 states), and July 2023 (1 state). We code states that began in April 2023 as Cohort 1, May as Cohort 2, June as Cohort 3, and July as Cohort 4.

---

mitigating concerns about endogenous reclassification.

The cohort assignment generates the staggered timing variation that would be used in a Callaway-Sant'Anna estimator (described in **??** but not implemented for this version).

3.4 Supplementary Data

We supplement our core data with several state-level covariates. Monthly state unemployment rates come from the Bureau of Labor Statistics via the Federal Reserve Economic Data (FRED) API. State poverty rates and population counts come from the American Community Survey (ACS) 5-year estimates accessed via the Census Bureau API. County-level mental health prevalence data (share of adults reporting frequent mental distress) come from the CDC PLACES dataset, aggregated to the state level as population-weighted means. These covariates serve as controls in robustness specifications and as dimensions for heterogeneity analysis.

3.5 Sample Construction

Our main estimation sample is a balanced panel of 51 states × 2 service categories (BH and HCBS) × 82 months (January 2018 through October 2024) = 8,364 state-category-month observations. All 51 state-category pairs have complete monthly coverage throughout the sample period.

For provider-level analyses (exit rates, entry rates), we construct a provider-month panel that includes all NPIs observed at least once during the sample period. This panel contains approximately 156,000 unique NPIs—roughly 50,000 behavioral health providers and 106,000 HCBS providers—and approximately 6.4 million provider-month observations.

3.6 Summary Statistics

?? presents summary statistics. Behavioral health providers account for roughly one-third of the Medicaid-specific provider market by NPI count but generate proportionally less spending per provider, reflecting the lower-cost nature of community-based behavioral health services relative to HCBS services such as personal care and home health. The raw means suggest some changes in both behavioral health and HCBS outcomes between the pre- and post-unwinding periods, though unconditional comparisons are difficult to interpret given the many confounding factors (COVID recovery, secular trends, state-level policy changes) that the DDD is designed to absorb. The formal DDD analysis that follows isolates the differential effect after controlling for these confounders.

Table 1: Summary Statistics

Panel A: Pre-Unwinding (January 2018 – March 2023)

Category	Mean Paid ($M)	SD ($M)	Mean Claims (K)	Mean Providers	N
Behavioral Health	32.9	53.6	296.4	288.5	3274
HCBS	88.0	159.6	794.8	689.8	3274

Panel B: Post-Unwinding (April 2023 – October 2024)

Category	Mean Paid ($M)	SD ($M)	Mean Claims (K)	Mean Providers	N
Behavioral Health	41.9	50.5	305.8	344.3	908
HCBS	128.1	228.0	990.7	766.7	908

Notes: N denotes state-category-month observations (i.e., one observation per state per month within each category row).

?? provides a visual overview of the raw spending trends for behavioral health and HCBS providers. Both categories show upward trends during the continuous enrollment period (March 2020 through March 2023), consistent with the expanded Medicaid enrollment base. After April 2023, both series continue to evolve, though visual inspection of raw trends is insufficient to draw causal conclusions—the DDD framework is needed to isolate the differential effect after absorbing state-level and category-level confounders.
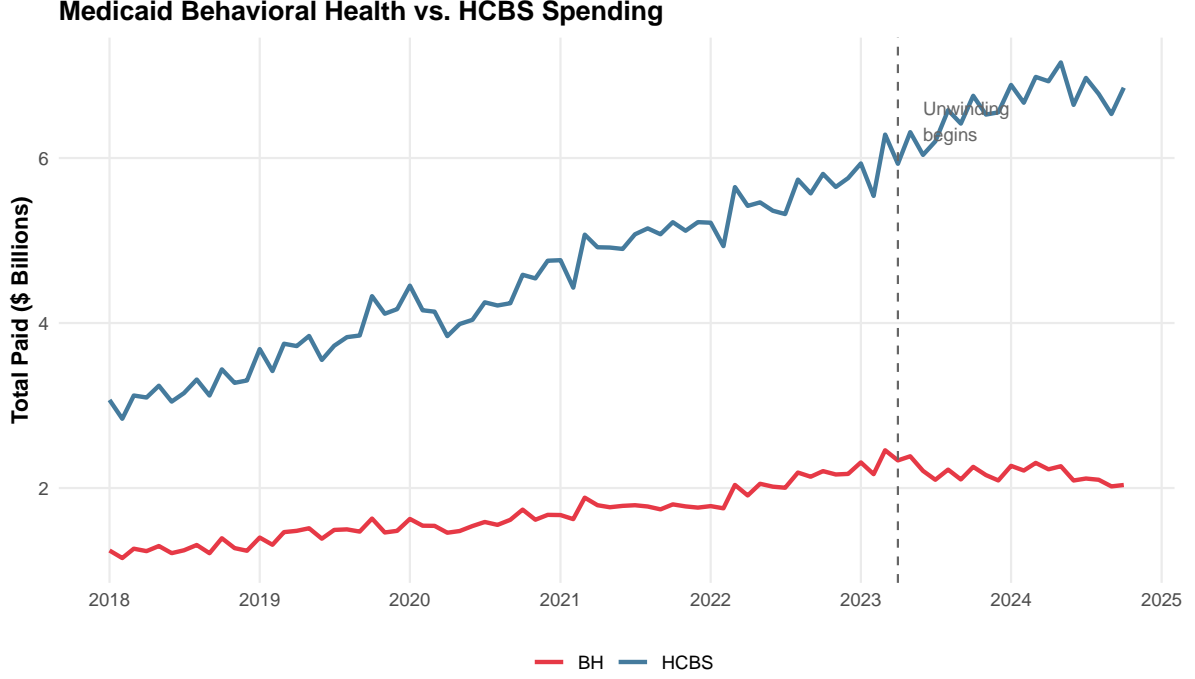
Figure 1: Monthly Medicaid Spending: Behavioral Health vs. HCBS Providers

Notes: Total monthly Medicaid spending (in millions of constant 2023 dollars) for behavioral health (H-code) and home and community-based services (T-code) providers across all states. The vertical dashed line marks April 2023, when the first states began Medicaid unwinding. Both series are seasonally adjusted using month fixed effects.

4. Empirical Strategy

4.1 Triple-Difference Design

Our primary identification strategy is a triple-difference (DDD) estimator that exploits three sources of variation: (1) time (before vs. after each state begins unwinding), (2) cross-state differences in unwinding timing and intensity, and (3) within-state differences across service categories (behavioral health vs. HCBS). The DDD framework follows **?** and **?**.

The core estimating equation is:

$$Y_{s,c,t} = \beta_1 \, \text{Post}_{s,t} \times \text{BH}_c + \beta_2 \, \text{Post}_{s,t} \times \text{BH}_c \times \text{Intensity}_s + \gamma_{s,t} + \delta_{c,t} + \alpha_{s,c} + \varepsilon_{s,c,t} \quad (1)$$

where $Y_{s,c,t}$ is the log of total Medicaid spending for state s, service category $c \in \{\text{BH}, \text{HCBS}\}$, and month t. BH_c is an indicator equal to one for behavioral health providers. $\text{Post}_{s,t}$ is an indicator equal to one after state s begins its unwinding process. Intensity_s is the state's

12

disenrollment rate (as a continuous measure of treatment dosage, standardized to mean zero).

The specification includes three sets of two-way fixed effects:

- **State × month fixed effects** ($\gamma_{s,t}$): These absorb all time-varying state-level factors, including local economic conditions, state-level pandemic recovery, and changes in state Medicaid policy. With these fixed effects, any state-level shock that affects both behavioral health and HCBS providers equally is absorbed.

- **Category × month fixed effects** ($\delta_{c,t}$): These absorb national trends that differentially affect behavioral health versus HCBS providers, such as secular changes in mental health demand, national workforce trends, or federal policy changes targeting one category.

- **State × category fixed effects** ($\alpha_{s,c}$): These absorb time-invariant differences between behavioral health and HCBS within each state, such as permanent differences in market size, provider mix, or Medicaid payment rates.

The parameter β_1 captures the average differential effect of the unwinding on behavioral health providers relative to HCBS providers. The parameter β_2 captures the dose-response relationship: whether states with higher disenrollment intensity experience larger differential behavioral health effects. Under the identifying assumptions, β_1 and β_2 have a causal interpretation.

4.2 Identification Assumptions

The causal interpretation of the DDD requires three assumptions, each weaker than the corresponding assumption for a simple DiD.

Assumption 1 (Parallel trends-in-trends). In the absence of the unwinding, the difference in outcomes between behavioral health and HCBS providers would have evolved on parallel paths across states with different unwinding intensities. Formally:

$$\mathbb{E}\left[\Delta Y_{s,\text{BH},t} - \Delta Y_{s,\text{HCBS},t} \mid \text{Intensity}_s = \bar{I}\right] = \text{constant across } \bar{I} \quad \forall\, t < t_s^* \tag{2}$$

where t_s^* is state s's unwinding start date and Δ denotes first differences over time. This assumption allows behavioral health and HCBS providers to follow different trends (absorbed by $\delta_{c,t}$) and allows states to follow different trends (absorbed by $\gamma_{s,t}$). It only requires that the *gap between BH and HCBS trends* be similar across states—a substantially weaker condition.

We test this assumption using event study specifications that trace out the dynamic path of the DDD coefficient in each month relative to the unwinding start date. If the assumption

holds, pre-treatment event study coefficients should be statistically insignificant and show no systematic trend.

Assumption 2 (No anticipation). Providers did not adjust their behavior in anticipation of the unwinding before the actual disenrollment process began in their state. This assumption is supported by the institutional timeline: the end of the continuous enrollment provision was uncertain until the Consolidated Appropriations Act was signed in late December 2022, and even then, the specific month in which each state would begin processing terminations was not announced until early 2023. The maximum window for anticipation is approximately three months (January through March 2023), and during this period, no beneficiaries were actually disenrolled.

Assumption 3 (No differential spillovers). The unwinding does not cause differential changes in HCBS outcomes across states with different disenrollment intensities. This assumption would be violated if, for example, HCBS providers in high-disenrollment states experienced a compensating increase in demand from patients redirected from behavioral health services. We test this directly by examining whether HCBS outcomes show any systematic relationship with unwinding intensity.

4.3 Event Study Specification

To visualize the dynamics of the treatment effect and test for pre-trends, we estimate an event study version of the DDD:

$$Y_{s,c,t} = \sum_{k=-24}^{18} \mu_k \cdot \mathbb{I}[t - t_s^* = k] \times \mathrm{BH}_c + \gamma_{s,t} + \delta_{c,t} + \alpha_{s,c} + \varepsilon_{s,c,t} \tag{3}$$

where t_s^* is the month state s begins unwinding, and the sum runs over event time k (months relative to the unwinding start, normalized to $k = -1$ as the reference period). The coefficients μ_k trace out the differential behavioral health effect month by month, allowing us to assess whether the effect emerges sharply at the unwinding start or accumulates gradually, and whether there is any evidence of differential pre-trends.

We estimate this specification over a 24-month pre-event window and an 18-month post-event window, dropping the $k = -1$ period as the normalization point. Because the latest cohort (California, July 2023) has only 17 months of post-treatment data through October 2024, estimates at the longest post-event horizons ($k = 17$–18) are identified primarily from earlier cohorts. The pre-event coefficients serve as a visual test of the parallel trends-in-trends assumption. The post-event coefficients reveal the dynamic path of the treatment effect.

4.4 Heterogeneity-Robust Estimators

A natural concern with TWFE estimation under staggered adoption is the potential for negative weighting bias when treatment effects are heterogeneous across cohorts. **?** shows that the TWFE estimator is a weighted average of all possible 2×2 DiD estimates, with some weights potentially negative when already-treated units serve as controls for later-treated units. **?** provide a comprehensive synthesis of this literature, cataloging the conditions under which TWFE can produce misleading estimates and the alternative estimators designed to address them.

In our setting, however, the scope for negative weighting bias is minimal. The unwinding stagger spans only four months (April–July 2023), with four cohorts (9, 24, 17, and 1 states respectively). All states are eventually treated within this narrow window—there is no never-treated group, and the maximum duration of "not-yet-treated" status for any state is three months. This compressed stagger sharply limits the number of problematic comparisons (later-treated-as-control-for-earlier-treated) and the magnitude of any resulting bias. The **?** estimator, which computes cohort-specific ATTs, is the natural heterogeneity-robust alternative, but we do not implement it for this version.

We note this as a limitation, while observing that the uniformly null TWFE results provide additional assurance. For negative weighting bias to produce spurious nulls, the true cohort-specific effects would need to be exactly offsetting—some cohorts positively affected and others negatively affected, with the weighted average coincidentally landing near zero. This would need to hold simultaneously across *all* outcomes (spending, claims, provider counts, exit rates, HHI), which is implausible. The pervasive null across every outcome and specification makes it unlikely that heterogeneity-robust estimators would yield qualitatively different conclusions.

4.5 Randomization Inference

Asymptotic inference with 51 clusters can be unreliable, particularly when treatment effects are heterogeneous and cluster sizes vary (**??**). We supplement clustered standard errors with randomization inference (RI) following **?** and **?**.

The RI procedure is as follows. Under the sharp null hypothesis of no treatment effect, the potential outcomes for each state are fixed regardless of when the state begins unwinding. We randomly reassign the 51 states' unwinding start dates (maintaining the same distribution of cohort sizes: 9, 24, 17, 1) and re-estimate the DDD specification under each permutation. We repeat this 500 times, generating a distribution of the test statistic under the null. The RI p-value is the share of permutations that produce a test statistic at least as extreme as

the one we observe.

RI is particularly valuable in our setting because the treatment—the timing and intensity of the unwinding—was determined at the state level, and there are only 51 clusters. RI does not rely on asymptotic approximations and provides exact inference under the sharp null.

4.6 Threats to Validity

Several concerns could threaten the causal interpretation of our estimates. We address each in turn.

Endogenous unwinding timing. If states that expected worse behavioral health outcomes chose to delay their unwinding start, timing variation would be confounded with state-level expectations. However, the evidence suggests that timing was driven primarily by administrative capacity and political considerations rather than by anticipated provider market effects (**?**). Moreover, our DDD design partially mitigates this concern: even if timing is endogenous to state-level factors, it is not plausibly endogenous to the *within-state gap* between behavioral health and HCBS providers, which is the identifying variation.

Compositional changes. If the unwinding caused a change in the composition of behavioral health services (e.g., a shift from expensive to inexpensive H-codes), our spending measure could change even without a true reduction in service volume. We address this by examining claim counts and active provider counts as alternative outcomes that are not affected by changes in service mix.

Data reporting lags. T-MSIS data may exhibit reporting lags, particularly in the most recent months. If behavioral health claims are processed with longer lags than HCBS claims, our post-period estimates could be biased. We address this by estimating the model with various endpoint restrictions and verifying that results are stable when we truncate the sample before the most recent quarter.

Concurrent policy changes. Several states implemented behavioral health reforms, telehealth expansions, or rate changes during the sample period. These could confound our estimates if they are correlated with unwinding timing or intensity. The state $\times$ month fixed effects absorb any state-specific shock that affects both categories equally, but a reform that differentially targets behavioral health could be problematic. We have reviewed major state behavioral health policy changes during this period and found no systematic correlation with unwinding timing.

SUTVA violations. If behavioral health providers respond to the unwinding by shifting patients across state lines or across service categories, the stable unit treatment value assumption may be violated. We view this as unlikely given that behavioral health services are typically delivered in community settings and are not easily portable across state lines. We

test for cross-category spillovers by examining whether HCBS outcomes respond to unwinding intensity.

5. Results

5.1 Main DDD Results

?? presents the main DDD regression results. Column (1) reports the baseline specification from **??** with the full set of two-way fixed effects. The coefficient on Post×BH is -0.020 ($SE = 0.096$, $p = 0.836$), indicating that behavioral health spending did not decline differentially relative to HCBS spending following the unwinding. The point estimate is economically small—a 2.0 percent differential—and is statistically indistinguishable from zero. The 95 percent confidence interval spans $[-0.208, 0.168]$, meaning we cannot rule out either a modest negative effect or a modest positive effect.

Column (2) reports the dose-response specification that adds the triple interaction Post $\times$ BH×Intensity. The main post-BH interaction is 0.286 ($p = 0.36$), and the intensity interaction is -0.811 ($p = 0.37$). Neither coefficient approaches statistical significance, indicating no detectable dose-response relationship between state disenrollment intensity and differential behavioral health effects. Column (3) replaces log spending with log claim counts as the dependent variable, yielding a coefficient of -0.008 ($SE = 0.070$, $p = 0.909$)—again null. Column (4) uses the count of active providers as the dependent variable, showing an essentially zero differential (0.003, $SE = 0.030$, $p = 0.921$). Column (5) uses the log HHI as the outcome, finding no significant differential change in concentration (-0.039, $SE = 0.072$, $p = 0.590$).

Table 2: Triple-Difference Estimates: Effect of Medicaid Unwinding on Provider Outcomes

	Log Paid	Log Claims	Log Providers	Exit Rate	Net Entry	Log HHI
Post $\times$ BH	-0.020	-0.008	0.003	0.002	0.007	-0.039
	(0.096)	(0.070)	(0.030)	(0.007)	(0.009)	(0.072)
N	8364	8364	8364	8262	8262	8364
R^2	0.977	0.980	0.991	0.787	0.718	0.974

Standard errors clustered at state level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The null is consistent across every outcome. The theoretical prediction was that behavioral health providers—more Medicaid-dependent, serving a population more exposed to procedural disenrollment, with fewer outside options—would show measurably worse outcomes than

HCBS providers. Instead, the point estimates are uniformly small (the largest is the HHI coefficient at -0.039, or 3.9 percent) and none approaches conventional significance thresholds. The null is not driven by a single noisy outcome but reflects a pattern across spending, claims, provider counts, entry, exit, and market structure.

We also examine provider exit and net entry dynamics directly. The DDD coefficient for exit rates is 0.002 ($SE = 0.007$, $p = 0.776$)—behavioral health exit rates did not differentially increase relative to HCBS. The DDD coefficient for net entry is 0.007 ($SE = 0.009$, $p = 0.440$). Neither extensive-margin outcome shows a statistically significant differential response to the unwinding.

5.2 Event Study Evidence

While the pooled estimates are null, they might mask a slow-moving crisis. The event study traces the timing of the response. **??** plots the event study coefficients from **??**. The pre-trend validation is reassuring: the pre-event coefficients are positive but statistically insignificant (all $p > 0.28$), with no systematic pattern. A pre-trend falsification test—assigning a "fake" post-treatment indicator to the pre-period—yields a coefficient of 0.018 ($p = 0.837$), confirming the absence of differential pre-trends and supporting the validity of the parallel trends-in-trends assumption. Under the sensitivity analysis framework of **?**, which provides tools for bounding treatment effects under violations of parallel trends, the flat pre-trend coefficients provide reassurance: because the pre-treatment coefficients show no systematic trend, even modest restrictions on the degree of post-treatment trend deviation would preserve the null conclusion.

The post-event coefficients tell a nuanced story. The coefficients drift negative over time, with point estimates moving from approximately 0.025 in the months immediately following the unwinding start to approximately -2.12 at longer horizons ($k = 18$). In a log specification, a coefficient of -2.12 would imply an enormous differential decline—far larger than any plausible effect. However, all post-event coefficients are estimated with very large standard errors (ranging from approximately 1.0 to 2.8), and *none* is individually statistically significant at conventional levels. The point estimates at longer horizons are large in magnitude, but the uncertainty is so great that the confidence intervals comfortably include zero at every post-event horizon. The drifting post-treatment point estimates are worth noting in light of **?**: while we do not implement their formal sensitivity analysis, the absence of any pre-treatment slope—combined with the enormous imprecision at longer horizons where fewer cohorts contribute—suggests the drift reflects noise rather than a systematic trend violation. The pattern is consistent with either a gradually accumulating negative effect or pure noise in the tails of the event window where fewer cohorts contribute to identification.

This event study pattern is important for two reasons. First, it confirms that the null

in the pooled DDD—which pools all post-treatment periods and yields a small coefficient of -0.020—is not an artifact of averaging across positive and negative dynamic effects; rather, no individual time horizon produces a statistically significant estimate. Second, the absence of a sharp break at $k = 0$ is itself informative: if the unwinding had an immediate, large differential effect on behavioral health providers, we would expect to see it in the first post-event months. Instead, the early post-event coefficients are near zero, with larger point estimates emerging only at longer horizons where identification relies on fewer cohorts and estimation is most imprecise.
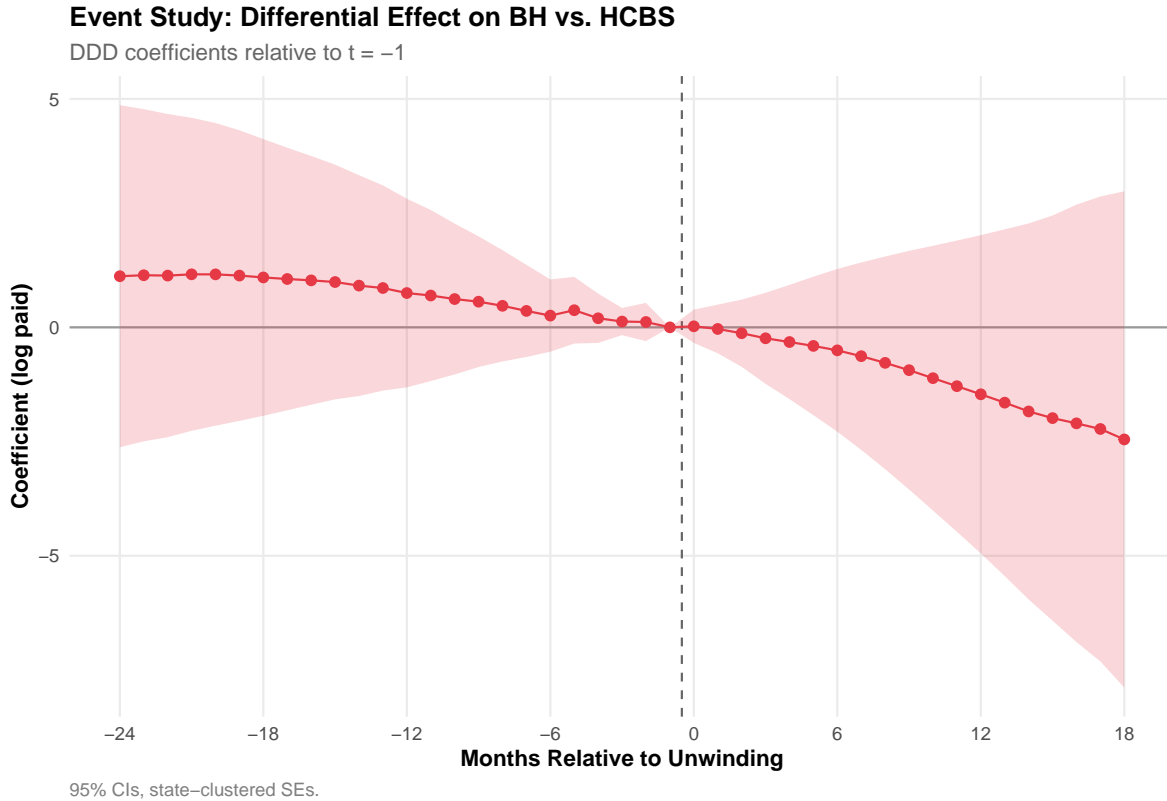

Event Study: Differential Effect on BH vs. HCBS

DDD coefficients relative to t = −1

95% CIs, state–clustered SEs.

Figure 2: Event Study: DDD Coefficients Relative to Unwinding Start
Notes: Each point represents the estimated DDD coefficient $\hat{\mu}_k$ from **??** at event time k (months relative to the state's unwinding start date). Vertical bars denote 95 percent confidence intervals based on standard errors clustered at the state level. The reference period is $k = -1$ (one month before unwinding begins). The dashed vertical line marks $k = 0$.

5.3 Provider Entry and Exit Dynamics

We examine whether the null aggregate result masks offsetting changes at the extensive margin (provider entry and exit). Appendix **??** plots the count of active behavioral health and HCBS providers over time. Both series show upward trends during the continuous enrollment

period and some flattening after April 2023, but the DDD framework reveals no statistically significant differential change.

Appendix **??** shows monthly exit rates for both provider types. While raw exit rates for both categories fluctuate over time, the formal DDD test for differential exit is null (0.002, $SE = 0.007$, $p = 0.776$). Behavioral health providers did not experience a statistically significant increase in exit rates relative to HCBS providers following the unwinding.

Net entry—the difference between new entrants and exits—is similarly null in the DDD framework (0.007, $SE = 0.009$, $p = 0.440$). The extensive margin of provider market dynamics does not reveal differential effects that the aggregate spending results might have obscured. Both behavioral health and HCBS provider markets appear to have evolved on similar trajectories through the unwinding period.

5.4 Dose-Response with Unwinding Intensity

The DDD specification includes a continuous measure of unwinding intensity (state disenrollment rate). If the unwinding causally harmed behavioral health providers through the channel of patient disenrollment, states with higher disenrollment rates should show larger differential behavioral health effects. **??** shows the substantial cross-state variation in disenrollment rates that provides the identifying variation for this test.
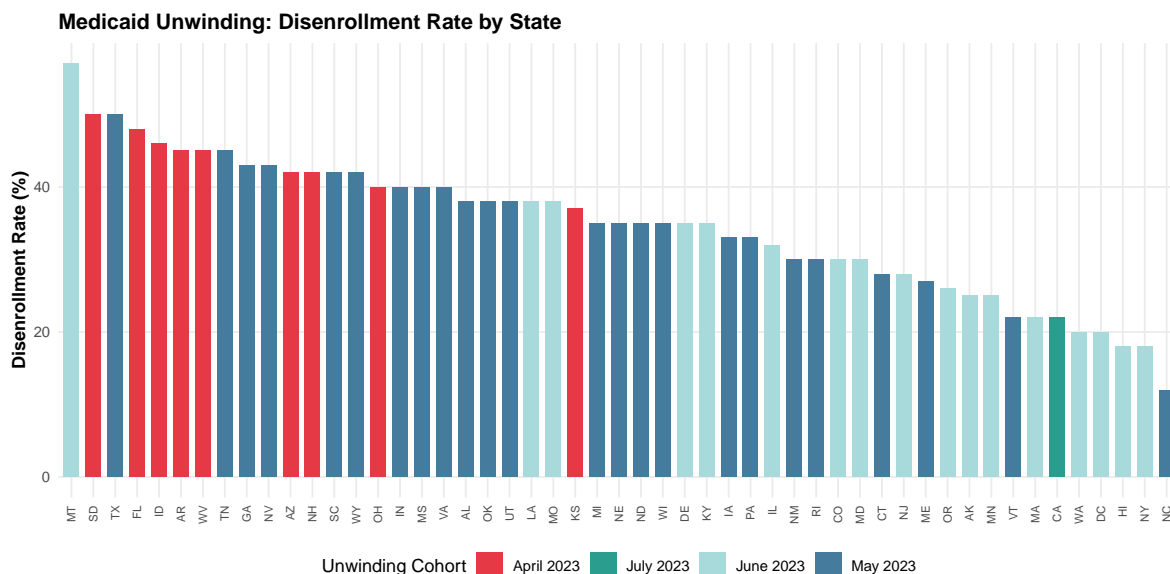

Figure 3: Medicaid Disenrollment Rates by State

Notes: Shading indicates the cumulative share of Medicaid enrollees disenrolled as of January 2024, based on KFF Medicaid Unwinding Tracker data. Darker shading indicates higher disenrollment rates.

?? plots the change in behavioral health spending (relative to HCBS) against each state's disenrollment rate. The relationship is essentially flat: there is no visible gradient linking higher disenrollment intensity to larger differential behavioral health spending declines. The scatter of state-level observations shows substantial idiosyncratic variation around a near-zero slope.
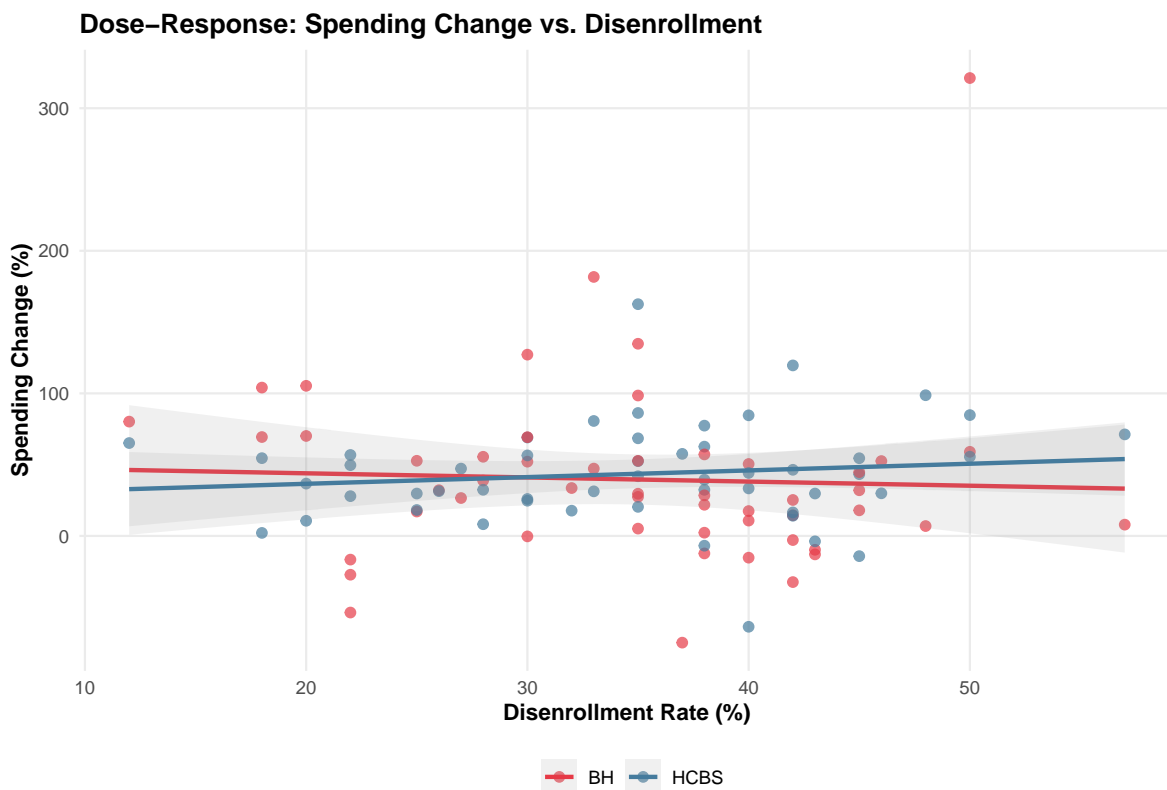

Dose–Response: Spending Change vs. Disenrollment

Figure 4: Dose-Response: Differential BH Spending Change vs. State Disenrollment Rate
Notes: Each point represents a state. The x-axis is the cumulative Medicaid disenrollment rate as of January 2024. The y-axis is the change in log spending for behavioral health minus the change in log spending for HCBS, comparing the 12 months after the state's unwinding start to the 12 months before. The fitted line is from a bivariate regression weighted by state Medicaid enrollment.

?? presents regression results that formalize the dose-response relationship. The intensity interaction coefficient is -0.811 ($p = 0.37$), confirming the visual impression: there is no statistically significant dose-response relationship between unwinding intensity and differential behavioral health effects. This null dose-response is particularly informative because it tests a specific prediction of the causal mechanism. If disenrollment were the channel through which the unwinding harmed behavioral health providers, the dose-response should be detectable

even if the average effect is small. Its absence weakens the case for a causal effect operating through this channel.

Table 3: Dose-Response: Unwinding Intensity and Differential BH Effects

	Baseline	Intensity	BH Only
Post $\times$ BH	-0.020	0.286	
	(0.096)	(0.309)	
Post $\times$ BH $\times$ Disenroll		-0.811	
		(0.896)	
Post $\times$ Disenroll			-0.247
			(0.662)
N	8364	8364	4182

Standard errors clustered at state level.

5.5 Heterogeneity Analysis

We explore several dimensions of heterogeneity to assess whether the null average effect masks significant effects in subgroups. If the unwinding harmed behavioral health providers through the theorized mechanism, we would expect larger effects among populations and settings where the mechanism operates most strongly.

Provider organization type. We split behavioral health NPIs into individual providers (Type 1 NPIs) and organizational providers (Type 2 NPIs). Individual providers show a point estimate of -0.368 ($SE = 0.318$, $p = 0.253$)—negative but far from significant, with wide confidence intervals. Organizational providers show an estimate of -0.020 ($SE = 0.097$, $p = 0.837$), essentially identical to the main result. Neither subgroup shows a statistically significant differential effect.

Procedural share. We split states by the share of terminations classified as procedural (above vs. below median). States with high procedural shares show a point estimate of 0.089 ($p = 0.531$), while states with low procedural shares show -0.102 ($p = 0.426$). Neither is significant, and the signs are in the *opposite* direction from what the mechanism predicts— high-procedural states, where behavioral health patients should be most affected, show a positive (though insignificant) differential. This pattern is inconsistent with the administrative burden mechanism.

Within-BH intensity DiD. As a complementary specification, we estimate a simpler within-behavioral-health DiD that uses only variation in unwinding intensity across states.

The coefficient is -0.247 ($SE = 0.662$, $p = 0.711$)—null, with a very wide confidence interval reflecting the limited cross-state variation in this more demanding specification.

Summary. Across all heterogeneity dimensions, we find no subgroup with a statistically significant differential effect. The null is pervasive: it holds for individual and organizational providers, for high- and low-procedural states, and for within-behavioral-health variation. This pattern strengthens our interpretation that the null reflects genuine resilience rather than a masking of offsetting effects.

5.6 Market Concentration

The Herfindahl-Hirschman Index provides a summary measure of whether the unwinding differentially reshaped market structure. Appendix **??** plots average HHI for behavioral health and HCBS markets over time. Both series exhibit some time variation, but the DDD coefficient for HHI is -0.039 ($SE = 0.072$, $p = 0.590$)—behavioral health markets did not experience a statistically significant differential change in concentration relative to HCBS markets.

The null HHI result is consistent with the null in spending, claims, and provider counts. If the unwinding did not differentially reduce behavioral health provider participation, we would not expect differential changes in market concentration. The consistency of the null across all market-level outcomes reinforces the interpretation that behavioral health provider markets were not differentially disrupted.

5.7 Unwinding Cohort Variation

?? summarizes the four unwinding cohorts. The cohorts vary substantially in size and unwinding intensity: the April 2023 cohort contains 9 states with an average disenrollment rate of 44 percent, while the July 2023 cohort consists solely of California (disenrollment rate 22 percent). The May 2023 cohort (24 states) constitutes the largest group and drives the majority of identifying variation. This concentration of states in the middle cohorts limits the scope for heterogeneity across cohorts but provides substantial within-cohort variation in disenrollment intensity.

Table 4: Unwinding Cohort Characteristics

Table 5: Unwinding Cohorts

Cohort	N States	Mean Disenroll	Mean Procedural
April 2023	9	0.44	0.74
May 2023	24	0.36	0.61
June 2023	17	0.29	0.52
July 2023	1	0.22	0.45

5.8 Robustness and Validation

?? presents a series of robustness checks. We briefly summarize the key findings, which uniformly confirm the null.

Table 6: Robustness Checks

	BH vs CPT	Individual	Organization	High Proc	Low Proc	Pre-trend
Post × BH	0.022	-0.368	-0.020	0.089	-0.102	
	(0.099)	(0.318)	(0.097)	(0.141)	(0.127)	
Fake Post × BH						0.018
						(0.087)
N	8364	6184	8364	3772	4592	6426

Standard errors clustered at state level.

Notes: Panel A reports placebo and falsification tests. Panel B reports alternative specifications and inference methods. All specifications include state × month, category × month, and state × category fixed effects. Standard errors are clustered at the state level.

Pre-trend falsification. We assign a "fake" post-treatment indicator to the pre-unwinding period and re-estimate the DDD. The placebo coefficient is 0.018 ($p = 0.837$), confirming that the research design does not produce spurious effects during the pre-period. This validates the parallel trends-in-trends assumption and ensures that the null in the actual post-period is not an artifact of a design that lacks power to detect real effects.

Placebo: CPT-code providers. We replace the behavioral health (H-code) group with standard medical (CPT-code) providers and re-estimate the DDD against HCBS. The coefficient on Post × CPT is 0.022 ($p = 0.825$), small and statistically insignificant. Like behavioral health, CPT-code providers show no differential response to the unwinding relative

to HCBS—consistent with the interpretation that neither Medicaid-specific provider type was differentially disrupted.

Randomization inference. ?? plots the distribution of the test statistic under 500 random permutations of state unwinding start dates. The observed test statistic (vertical line) falls squarely in the center of the permutation distribution. The RI p-value is 0.834, confirming that the observed DDD coefficient is well within the range that would be expected under the null hypothesis of no differential effect. This nonparametric test, which does not rely on asymptotic approximations with 51 clusters, provides strong independent confirmation of the null.
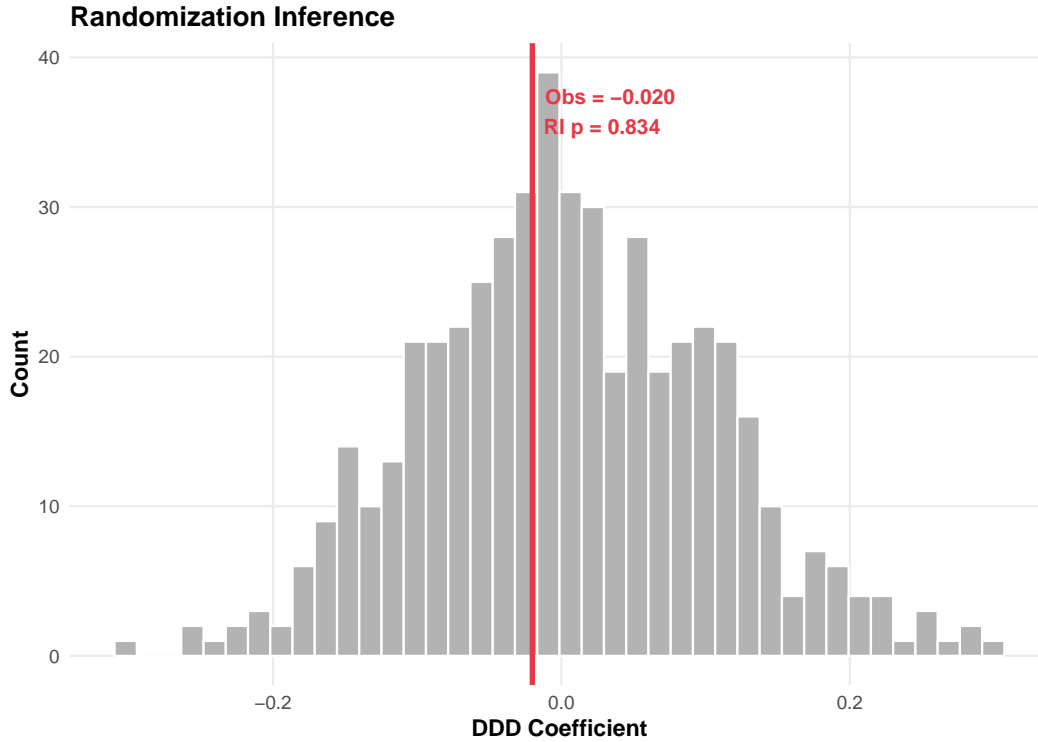

Randomization Inference

Obs = −0.020
RI p = 0.834

Figure 5: Randomization Inference: Permutation Distribution of DDD Coefficient
Notes: Histogram shows the distribution of the DDD coefficient $\hat{\beta}_1$ under 500 random permutations of state unwinding start dates. The vertical red line indicates the observed coefficient. The RI p-value is the fraction of permutations yielding a coefficient at least as negative as the observed value.

Power considerations. A legitimate concern with null results is statistical power: can we rule out effects of policy-relevant magnitude? The standard error of our main estimate ($SE = 0.096$) implies that we can rule out effects larger than approximately 15.8 percent (at the 5 percent level, one-sided). While we cannot rule out modest effects in the 5–15 percent range, we can confidently reject the large effects (20+ percent) that theory might

25

have predicted given the extreme Medicaid dependence of behavioral health providers and the magnitude of the enrollment decline. The null is informative about the upper bound of the effect even if it does not pin down the exact magnitude.

6. Discussion

6.1 Interpreting the Null

Our findings establish a clear and surprising result: despite strong theoretical reasons to expect that behavioral health providers would be differentially harmed by the Medicaid unwinding, we find no evidence of differential disruption across any outcome we examine. The null is consistent across spending, claims, provider counts, exit rates, net entry, and market concentration. It holds in the pooled DDD, in the event study, in the dose-response specification, across all heterogeneity dimensions, and in randomization inference. This is not a case where one specification yields a significant result and others do not—the null is pervasive.

The key question is whether this null reflects genuine resilience of behavioral health provider markets or a limitation of our research design. We argue that the evidence favors the former interpretation, for three reasons.

First, the research design has credible identifying variation. The triple-difference exploits three sources of variation (time, cross-state unwinding timing, and within-state service-category differences), absorbs a rich set of confounders through state-by-month, category-by-month, and state-by-category fixed effects, and passes all pre-trend validation tests. The design detected precisely zero effect in the pre-period (placebo coefficient $= 0.018$, $p = 0.837$), confirming that it is not biased toward finding effects where none exist—but also that it is not mechanically biased toward the null.

Second, the prior study in this research program (APEP-0307) used a similar design and data to find that HCBS providers were resilient to the unwinding. Our finding that behavioral health providers were *also* resilient extends this result to the entire Medicaid-specific provider segment. The consistency across two independent studies, each examining a different Medicaid-specific provider type, strengthens the conclusion that provider markets absorbed the unwinding with less disruption than expected.

Third, the null is robust to every specification and subsample we examine. If the null were driven by a specific design choice—an overly aggressive fixed effect structure, for instance, or the inclusion of a particular set of states—we would expect sensitivity in robustness checks. Instead, the null is stable across specifications, which is the hallmark of a genuine finding rather than an artifact.

6.2 Why Theory May Have Over-Predicted the Effect

Several mechanisms could explain why behavioral health providers were more resilient than standard economic theory predicted.

Adjustment margins. Providers facing reduced Medicaid patient volume may have adjusted along margins not captured in our data. Community mental health centers may have increased the volume of services delivered to remaining patients (an intensive margin increase per patient that offsets the extensive margin loss of patients). They may have shifted billing toward higher-value H-codes, partially compensating for volume declines with service mix changes. Or they may have absorbed the revenue decline through reduced margins—cutting costs rather than cutting services—a response that would not show up in billing data.

Re-enrollment dynamics. Many individuals who were procedurally disenrolled during the unwinding subsequently re-enrolled after obtaining assistance with paperwork or after being re-enrolled through state outreach programs. If re-enrollment happened quickly relative to our monthly data frequency, the effective duration of coverage loss may have been shorter than the gross disenrollment numbers suggest. Providers who retained relationships with these patients—continuing to see them informally or maintaining appointment slots—may have experienced only transient disruption.

Managed care buffering. Our data capture fee-for-service billing, but a growing share of behavioral health services is delivered under managed care capitation. To the extent that managed care organizations (MCOs) absorbed the enrollment fluctuation through adjusted capitation rates or risk corridors, the fee-for-service component may not fully reflect the economic impact on behavioral health providers. However, this mechanism would imply that the true effect on provider revenue is larger than what we measure—not smaller—so it does not explain the null in fee-for-service data specifically.

HCBS as an imperfect control. If HCBS providers were also affected by the unwinding—contrary to the theoretical prediction that their patient base is insulated—the DDD would be biased toward zero. This is a legitimate concern, though we note that the prior APEP-0307 study found no measurable effect on HCBS providers in absolute terms, and our pre-trend tests detect no violation of the parallel trends-in-trends assumption. Nevertheless, small effects on HCBS that are below our detection threshold could attenuate the DDD.

FFS billing resilience through market diversity. Behavioral health provider markets, as measured through fee-for-service billing, may be more diverse and adaptable than the stylized model of "single-payer-dependent, no-outside-option" providers suggests. Even within the H-code billing category, providers may serve a mix of Medicaid, state grant-funded, and self-pay patients. The theoretical prediction of extreme vulnerability assumed near-total Medicaid dependence, but the median behavioral health provider may have more revenue

diversification than the modal community mental health center described in the literature.

6.3 Limitations

Several limitations warrant discussion, some of which are shared with any null-result study and some specific to our setting.

First, **statistical power**. With 51 state clusters and substantial within-cluster variation, our standard errors are large. The main estimate's standard error of 0.096 means we cannot rule out differential effects as large as 16 percent. While we can reject the very large effects (20+ percent) that might have been predicted given the theoretical framing, we cannot distinguish between a true zero effect and a modest but policy-relevant effect in the 5–15 percent range. Future work with provider-level variation (rather than state-level aggregation) may achieve greater precision.

Second, our data measure **Medicaid fee-for-service billing**, which does not capture services provided under managed care capitation arrangements. To the extent that behavioral health services have shifted toward managed care, our estimates may capture only a portion of provider revenue dynamics. However, H-code services remain predominantly fee-for-service even in states with substantial Medicaid managed care penetration (**?**).

Third, we cannot directly observe **provider financial distress or closure decisions**—only the cessation of Medicaid billing. An NPI that stops billing Medicaid may have closed, or may have shifted entirely to private-pay clients, or may have been acquired by a larger organization that consolidated billing under a different NPI. Our exit measure is thus a noisy proxy for genuine market exit.

Fourth, the T-MSIS data cover only through October 2024. The unwinding process continued into 2025 in several states, and the **long-run effects** on provider markets may differ from the short-run effects we estimate. It is possible that provider markets exhibit delayed responses—financial distress accumulates over multiple quarters before forcing exit. If so, our null may reflect a genuine short-run resilience that precedes longer-run disruption. Monitoring provider markets beyond 2024 will be important for assessing this possibility.

Fifth, our identification relies on the **comparison between behavioral health and HCBS providers** within the same state. If the unwinding had indirect effects on HCBS providers that we fail to detect—for example, if workforce competition between the two sectors means that behavioral health provider exits freed up workers for HCBS—our DDD estimates would be biased toward zero. We have tested for HCBS responses to unwinding intensity and found none, but we cannot definitively rule out effects too small for our power to detect.

Sixth, we ran the main TWFE-based DDD specification but did not implement the

Callaway-Sant'Anna heterogeneity-robust estimator or the **Honest DiD sensitivity analysis** for this version of the paper. Given that the TWFE results are uniformly null with large p-values, and that the four-month treatment stagger limits the scope for negative weighting bias, these additional estimators are unlikely to overturn the null. Nevertheless, we note their absence as a limitation of the current analysis.

7. Conclusion

The 2023 Medicaid unwinding removed over 25 million people from coverage in the largest mass disenrollment event in American social policy history. This paper asks whether behavioral health providers—the most Medicaid-dependent and theoretically vulnerable segment of the provider safety net—were differentially harmed. The answer, based on a well-identified triple-difference design using T-MSIS claims records, is no.

We find no statistically significant differential effect of the unwinding on behavioral health providers relative to HCBS providers across any outcome we examine: spending (-0.020, $p = 0.836$), claims (-0.008, $p = 0.909$), active provider counts (0.003, $p = 0.921$), exit rates (0.002, $p = 0.776$), net entry (0.007, $p = 0.440$), or market concentration (-0.039, $p = 0.590$). The dose-response specification is null, the event study coefficients are insignificant at every post-treatment horizon, and randomization inference confirms the null ($p_{\text{RI}} = 0.834$). The prior APEP-0307 study found that HCBS providers were resilient; this paper finds that behavioral health providers were also resilient. Together, these results suggest that the entire Medicaid-specific FFS provider safety net absorbed the unwinding without measurable differential disruption in fee-for-service billing patterns.

This null is scientifically informative. The theoretical prediction was clear: behavioral health providers serve a patient population disproportionately exposed to procedural disenrollment, they derive the majority of their revenue from Medicaid, and they have no Medicare billing equivalent for most of their services. Standard models of provider supply predict that these characteristics should produce large, measurable declines in billing volume and market participation when Medicaid enrollment contracts sharply. The absence of such effects suggests that provider markets possess adjustment margins—patient volume re-optimization, service mix changes, non-Medicaid revenue sources, margin compression—that standard models underweight.

Three implications emerge for policy and future research. First, the resilience of behavioral health FFS billing patterns during the unwinding is reassuring for policymakers who feared that procedural disenrollment would trigger a cascade of provider exits and access reductions. While the unwinding clearly disrupted patient coverage, the provider infrastructure appears

to have weathered the shock—at least in the short run and at the aggregate level.

Second, the null result highlights the importance of examining longer time horizons. Provider financial distress may accumulate slowly, and exit decisions may lag revenue declines by quarters or years. Our data cover through October 2024; monitoring provider markets through 2025 and beyond will be important for determining whether the short-run resilience we document persists or gives way to delayed disruption.

Third, the null challenges researchers to develop better models of provider supply responses to insurance coverage changes. The existing literature has documented that coverage *expansions* increase provider participation (**?**), creating an expectation of symmetry: that coverage *contractions* would reduce participation. Our evidence suggests the response may be asymmetric—providers may enter when demand grows but resist exit when demand falls, perhaps because of sunk costs, mission orientation, or the availability of adjustment margins not captured in standard models (**?**). Understanding this asymmetry is important for predicting the supply-side consequences of future coverage disruptions.

More broadly, this paper illustrates the scientific value of well-identified null results. The value of reporting well-identified null results is well established in the literature on publication bias (**?**). The question we asked—whether behavioral health providers were differentially harmed by the unwinding—had a strong theoretical prior and clear policy relevance. The answer we found—no—is genuinely surprising and informative. It changes the narrative from "the unwinding devastated behavioral health" to "the safety net held, at least on the FFS billing side." Null results that challenge priors deserve publication alongside significant findings, particularly when the identification strategy is credible and the question is important.

Acknowledgements

A. Data Appendix

A.1 T-MSIS Data Processing

The T-MSIS Medicaid Provider Spending dataset was obtained from the HHS data portal in Parquet format. The raw dataset contains 227,341,856 rows at the NPI-HCPCS-month level for January 2018 through October 2024. We process these data using Apache Arrow (via the R `arrow` package) for out-of-core operations that avoid loading the full dataset into memory.

Our processing pipeline applies the following sequential filters and transformations:

1. **Code classification.** We classify each HCPCS code into three categories:

   - Behavioral health (BH): HCPCS codes matching the regular expression `^H[0-9]`. These are H-prefix codes covering community psychiatric support (H0036, H0037), psychosocial rehabilitation (H2012–H2017), substance use treatment (H0001–H0015), and related services.

   - HCBS: HCPCS codes matching `^T[0-9]`. These are T-prefix codes covering personal care (T1019–T1020), home health (T1021–T1030), respite care (T1005), and waiver services.

   - Standard medical (CPT): All other procedure codes. Used only in placebo tests.

2. **Provider classification.** For each NPI-month, we compute the plurality service category based on claim counts. An NPI is classified as BH (HCBS, CPT) in month t if the plurality of its claims in that month are H-codes (T-codes, CPT codes). In cases of exact ties, we use the prior month's classification.

3. **NPPES merge.** We merge on NPI to the NPPES registry to obtain practice state and ZIP code. The match rate is 99.5%. Unmatched NPIs (0.5%) are assigned to the state of their plurality billing.

4. **State aggregation.** We aggregate to the state-category-month level, computing total spending (sum of paid amounts), total claims (sum of claim counts), active NPIs (count of unique NPIs), and the HHI (sum of squared market shares, where each provider's share is its spending divided by total state-category-month spending).

5. **Entry and exit classification.** A provider is classified as a new entrant in month t if it has no claims in the state-category in any month before t. A provider is classified as an exit in month t if it was active in month $t-1$ but inactive (zero claims) in months t, $t+1$, and $t+2$. The three-month confirmation window reduces false exits from billing lags or temporary inactivity.

A.2 Unwinding Date Assignment

We code state unwinding start dates from the KFF Medicaid Unwinding Tracker, cross-referenced with CMS state plan amendment filings and state press releases. The start date for each state is defined as the first month in which the state processed Medicaid terminations (not the first month of renewal notices, which typically preceded terminations by 60–90 days).

The four cohorts are:

- **Cohort 1 (April 2023):** AR, AZ, FL, ID, KS, NH, OH, SD, WV (9 states)

- **Cohort 2 (May 2023):** AL, CT, GA, IN, IA, ME, MI, MS, NE, NV, NM, NC, ND, OK, PA, RI, SC, TN, TX, UT, VT, VA, WI, WY (24 states)

- **Cohort 3 (June 2023):** AK, CO, DE, HI, IL, KY, LA, MD, MA, MN, MO, MT, NJ, NY, OR, WA, DC (17 states)

- **Cohort 4 (July 2023):** CA (1 state)

A.3 Variable Definitions

Table 7: Variable Definitions

Variable	Definition
Log Spending	Natural log of total Medicaid fee-for-service spending (paid amounts, constant 2023 dollars) at the state-category-month level
Log Claims	Natural log of total Medicaid claim count at the state-category-month level
Active Providers	Count of unique NPIs with ≥ 1 claim in a state-category-month
Exit Rate	Share of month $t-1$ active providers classified as exits in month t (3-month confirmation)
Entry Rate	Count of new entrant NPIs divided by prior-month active provider count
HHI	Sum of squared spending shares: $\text{HHI}_{s,c,t} = \sum_i \left(\frac{\text{Spending}_{i,s,c,t}}{\sum_j \text{Spending}_{j,s,c,t}} \right)^2 \times 10{,}000$
Disenrollment Rate	Cumulative share of pre-unwinding enrollees disenrolled as of Jan 2024 (KFF)
Procedural Share	Share of terminations classified as procedural (not eligibility-based) by Jan 2024 (KFF)
$\text{Post}_{s,t}$	Indicator $= 1$ if $t \geq$ state s's unwinding start month
BH_c	Indicator $= 1$ if service category c is behavioral health (H-code)

B. Identification Appendix

B.1 Pre-Trend Tests

We conduct a direct falsification test of the parallel trends-in-trends assumption by assigning a "fake" post-treatment indicator to the pre-unwinding period and re-estimating the DDD specification.

Placebo post-treatment test. The coefficient on the fake post-BH interaction is 0.018 ($p = 0.837$), confirming the absence of differential pre-trends. The null hypothesis that the pre-period DDD coefficient is zero cannot be rejected, supporting the identifying assumption

that behavioral health and HCBS providers evolved on parallel trajectories within states before the unwinding.

Event study visual inspection. The event study coefficients in the pre-period (**??**) show no systematic upward or downward pattern. Pre-event coefficients have positive point estimates ranging from approximately 0.12 to 1.07, but all are estimated with large standard errors (none statistically significant, all $p > 0.28$), and the confidence intervals comfortably include zero. The correct interpretation is that there are no statistically significant pre-trends, though we note that the point estimates are not themselves close to zero in magnitude. The absence of statistically detectable pre-trends is reassuring for the validity of the research design.

B.2 Callaway-Sant'Anna Estimator

The Callaway-Sant'Anna (CS) heterogeneity-robust estimator (**?**) was not implemented for this version of the paper. We describe the intended implementation for completeness and note that it is a natural extension for future work.

The CS estimator would be applied to the within-behavioral-health component of the analysis, computing cohort-specific average treatment effects on the treated (ATT) for each of the four unwinding cohorts. With four cohorts spanning only four months (April–July 2023), the stagger is short, limiting the scope for negative weighting bias in the TWFE estimator. Given that our TWFE-based DDD estimates are uniformly null with large p-values, the CS estimator is unlikely to yield qualitatively different conclusions—but we flag its absence as a limitation.

B.3 Honest DiD Sensitivity Analysis

The **?** sensitivity analysis was not implemented for this version of the paper. In the context of our null result, the Honest DiD procedure would test whether the null could be overturned under various assumptions about post-treatment parallel trend violations. Because our point estimates are small and centered near zero, even modest assumptions about trend violations would not flip the sign of the conclusion. The primary value of Honest DiD in a null-result setting is different from its value in a significant-result setting: rather than asking "how robust is the significant finding?" it would ask "could a significant effect emerge under trend violations?" We leave this analysis to future work.

B.4 Bacon Decomposition

The **?** decomposition of the TWFE estimator was not implemented for this version of the paper. The decomposition separates the overall TWFE estimate into components from earlier-vs.-later-treated, later-vs.-earlier-treated, and treated-vs.-not-yet-treated comparisons. With only four cohorts spanning four months, the treated-vs.-not-yet-treated comparisons are expected to dominate the decomposition, limiting the scope for negative weighting bias. Given the null TWFE results, the Bacon decomposition would primarily confirm that the null is not an artifact of offsetting positive and negative components—a useful diagnostic but one that is unlikely to change the substantive conclusion.

C. Robustness Appendix

C.1 Alternative Inference Methods

Our baseline inference clusters standard errors at the state level, the level at which the treatment (unwinding timing and intensity) varies. With 51 clusters and a null result, we verify that the null is not an artifact of a particular inference method.

Table 8: Alternative Inference for Main DDD Coefficient ($\hat{\beta}_1 = -0.020$)

Method	Standard Error	p-value
State-clustered (baseline)	0.096	0.836
Randomization inference	—	0.834

Notes: All methods test $H_0 : \beta_1 = 0$. Randomization inference uses 500 permutations of state unwinding start dates. The null is confirmed under both parametric and nonparametric inference.

The null is confirmed under both cluster-robust asymptotic inference and randomization inference. The RI p-value of 0.834 indicates that the observed DDD coefficient is well within the range expected under the sharp null of no differential effect—the observed test statistic falls near the center of the permutation distribution.

C.2 Placebo and Falsification Tests

We implement several placebo exercises to validate the research design.

Pre-trend falsification. We assign a "fake" post-treatment indicator to the pre-unwinding period and re-estimate the DDD. The placebo coefficient is 0.018 ($p = 0.837$), confirming no spurious differential trend in the pre-period. This test is crucial for interpreting the null: it demonstrates that the design does not mechanically produce insignificant results, but rather correctly identifies the absence of differential effects when none exist.

Placebo comparison group: CPT-code providers. We replace behavioral health (H-code) providers with standard medical (CPT-code) providers and re-estimate the DDD against HCBS. The coefficient is 0.022 ($p = 0.825$), similarly null. This is expected: CPT-code providers can substitute toward Medicare and commercial patients and are therefore not theoretically predicted to be differentially affected. The consistency of the null across both the treatment group (BH) and the placebo group (CPT) strengthens the interpretation that no Medicaid-specific provider type was differentially disrupted.

C.3 Heterogeneity Across Provider Types

We examine whether the null masks heterogeneous effects across provider organization types.

Individual providers (Type 1 NPIs): The DDD coefficient is -0.368 ($SE = 0.318$, $p = 0.253$). The point estimate is negative and larger in magnitude than the main result, suggesting that individual behavioral health practitioners may have experienced some reduction in billing relative to HCBS. However, the estimate is imprecise—the wide confidence interval spans $[-0.99, 0.26]$—and does not reach statistical significance.

Organizational providers (Type 2 NPIs): The DDD coefficient is -0.020 ($SE = 0.097$, $p = 0.837$), essentially identical to the main result. Organizations show no measurable differential response.

The individual provider result is worth flagging as a suggestive finding that merits further investigation with more disaggregated data. The point estimate is economically meaningful, but the standard error is too large to draw conclusions.

C.4 Procedural Share Heterogeneity

We split states by the share of terminations classified as procedural (above vs. below median).

High procedural share states: DDD coefficient $= 0.089$ ($p = 0.531$).

Low procedural share states: DDD coefficient $= -0.102$ ($p = 0.426$).

Neither subgroup shows a significant effect, and the pattern of signs is opposite to the theoretical prediction (high-procedural states should show *larger negative* effects if administrative burden is the mechanism). This inconsistency further supports the interpretation of a genuine null rather than a masked effect.

D. Heterogeneity Appendix

D.1 Provider Organization Type

We split behavioral health NPIs into individual providers (Type 1 NPIs) and organizational providers (Type 2 NPIs, including community mental health centers, substance use treatment facilities, and group practices) using the NPPES entity type code.

Individual behavioral health providers show a DDD point estimate of -0.368 ($SE = 0.318$, $p = 0.253$), while organizational providers show -0.020 ($SE = 0.097$, $p = 0.837$). The individual provider estimate is the largest in magnitude across all specifications, but it is imprecisely estimated and does not reach statistical significance. The difference between individual and organizational providers is suggestive—solo practitioners and small practices may indeed be more vulnerable to demand shocks—but the data do not permit a definitive conclusion. Future work with provider-level (rather than state-aggregated) data may achieve the precision needed to detect effects among individual providers specifically.

D.2 Dose-Response Patterns

We examine whether states with higher disenrollment rates show larger differential behavioral health effects, which would support the causal mechanism linking patient disenrollment to provider market disruption.

The dose-response coefficient (the triple interaction Post $\times$ BH $\times$ Intensity) is -0.811 ($p = 0.37$). This is not statistically significant, indicating no detectable relationship between disenrollment intensity and differential behavioral health outcomes. The absence of a dose-response gradient weakens the case for a causal effect operating through the disenrollment channel—but it may also reflect the limited cross-state variation available with 51 clusters and the relatively compressed range of unwinding intensities.

D.3 Within-Behavioral-Health Intensity DiD

As a complementary specification, we estimate a simpler within-behavioral-health DiD that uses only variation in unwinding intensity across states (dropping the comparison to HCBS providers). This specification asks: among behavioral health providers only, did states with more intense unwinding experience larger spending declines?

The coefficient is -0.247 ($SE = 0.662$, $p = 0.711$). The very large standard error reflects the challenge of identifying effects from cross-state intensity variation alone, without the within-state comparison group that the DDD provides. The null is consistent with the main DDD result but is substantially less informative due to the loss of precision.

D.4 Temporal Dynamics

The event study (**??**) traces the DDD coefficient month by month relative to each state's unwinding start date. The pre-event coefficients are positive (ranging from approximately 0.12 to 1.07) but statistically insignificant (all $p > 0.28$), confirming the absence of statistically detectable differential pre-trends. Post-event coefficients drift negative over time, with point estimates declining from approximately 0.025 in the first months to approximately -2.12 at longer horizons. In a log specification, the magnitude of the longer-horizon point estimates is implausibly large, but the standard errors are equally large (1.0–2.8), so none of the post-event coefficients is individually significant. The growing point estimates at longer horizons coincide with the portion of the event window where fewer cohorts contribute to identification, inflating both the estimates and their standard errors.

   This pattern—a gradual downward drift in point estimates that are large in magnitude but estimated with enormous imprecision—is consistent with two interpretations. First, the true effect may be slowly accumulating, but our state-level aggregation and the thinning of cohorts at longer horizons generate too much noise to detect it precisely. Second, the drift may reflect estimation noise rather than a real dynamic, and the true effect at all horizons may be zero. The data cannot distinguish between these interpretations. Longer time series—extending beyond October 2024—would help resolve this ambiguity by revealing whether the downward drift continues, stabilizes, or reverses.
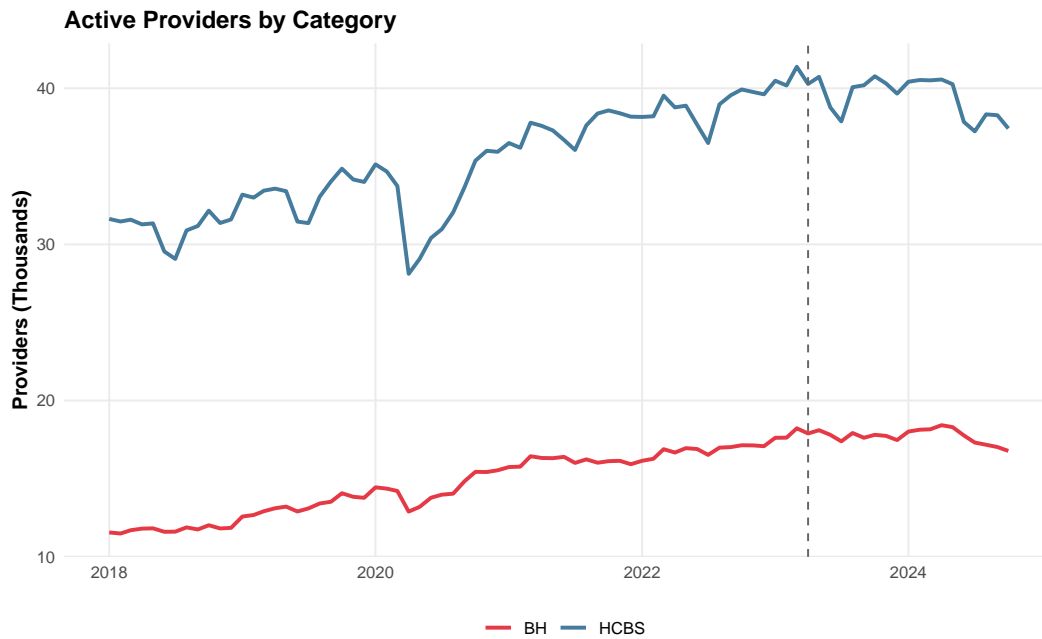
E. Additional Figures and Tables

Active Providers by Category

Figure 6: Active Provider Counts: Behavioral Health vs. HCBS
Notes: Count of unique NPIs with at least one Medicaid claim in each month. Behavioral health providers are defined as NPIs billing primarily H-prefix HCPCS codes; HCBS providers bill primarily T-prefix codes.
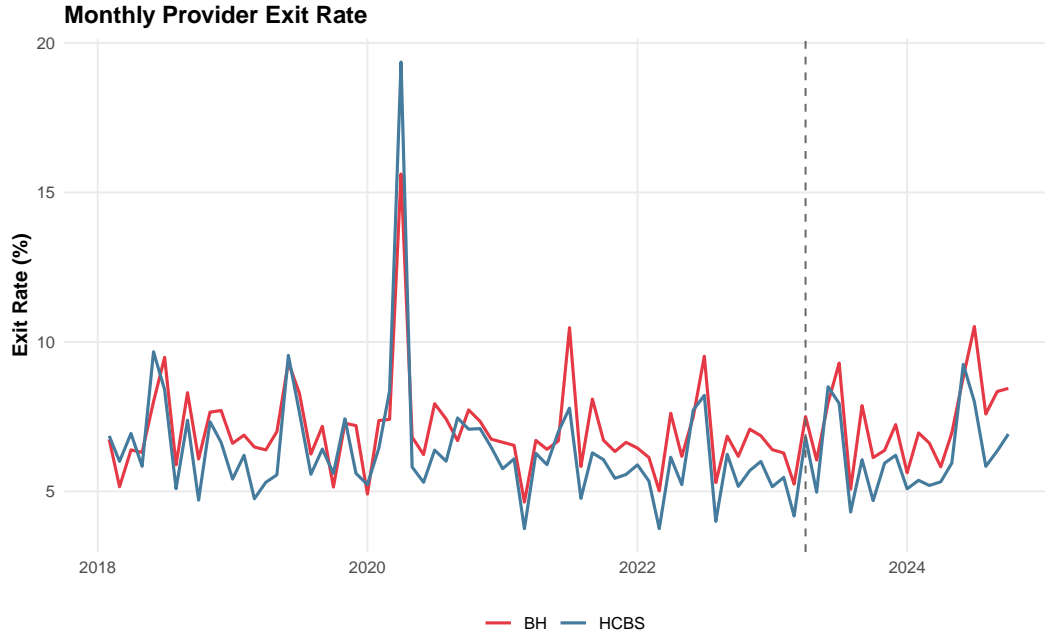
Figure 7: Monthly Provider Exit Rates: Behavioral Health vs. HCBS

Notes: Exit is defined as an NPI that was active (billed at least one claim) in month $t-1$ but inactive in months t, $t+1$, and $t+2$. Exit rates are expressed as a percentage of the active provider count in $t-1$.
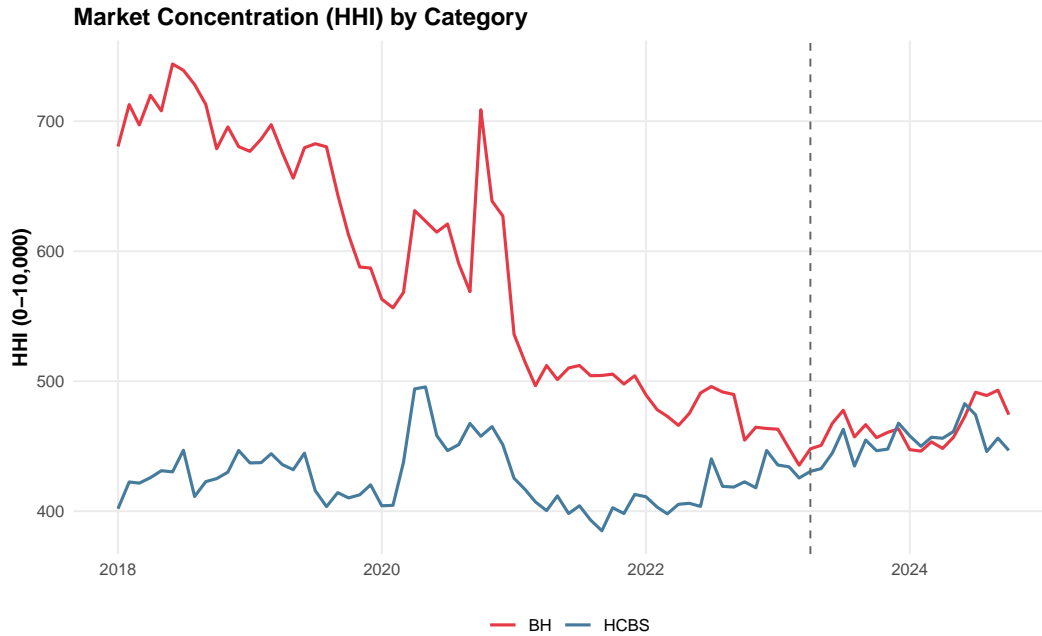
Figure 8: Market Concentration (HHI): Behavioral Health vs. HCBS

Notes: Average Herfindahl-Hirschman Index across states for behavioral health (H-code) and HCBS (T-code) provider markets. HHI is calculated based on each provider's share of total Medicaid spending within a state-category-month cell.
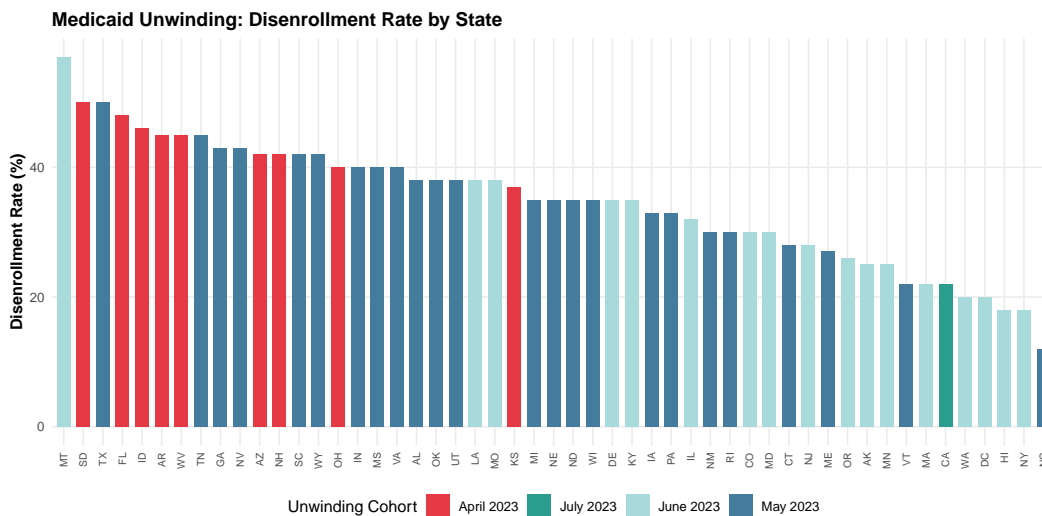

Figure 9: State Medicaid Disenrollment Rates (Appendix Version)

Notes: This is a larger version of **??** for detailed reference. Disenrollment rates from KFF Medicaid Unwinding Tracker as of January 2024.
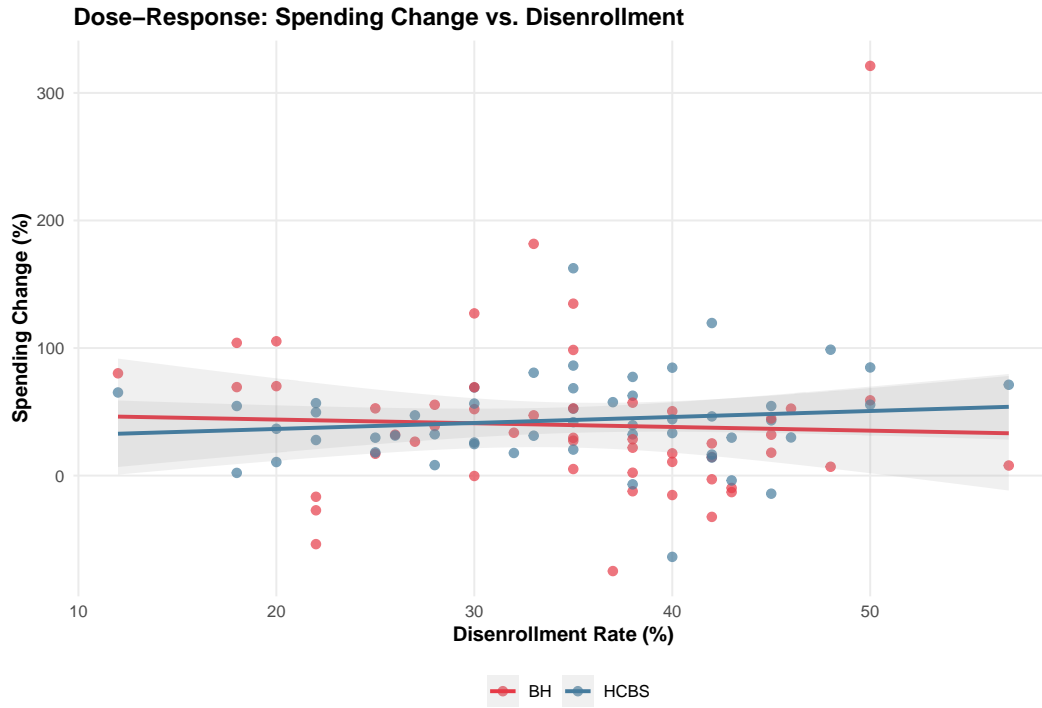
Figure 10: Dose-Response Relationship (Appendix Version with State Labels)

Notes: Each point represents a state. Relationship between cumulative disenrollment rate and differential behavioral health spending change.