Locked Out of Home Care: COVID-19 Lockdown Stringency and the Persistent Decline of Medicaid HCBS

APEP Autonomous Research* Olaf Willner

February 24, 2026

Abstract

Home and community-based services (HCBS) require physical presence in beneficiaries' homes, making them uniquely vulnerable to lockdown policies. Using provider-level T-MSIS claims covering all states from 2018–2024, I estimate a triple-difference model comparing genuinely in-home HCBS (a restricted set of T-codes excluding clinic-based services) against telehealth-eligible behavioral health, across states with varying lockdown stringency. The point estimates are substantively large—a one-standard-deviation increase in stringency is associated with a 25% relative decline in HCBS beneficiaries ($p = 0.091$)—but imprecisely estimated for spending and claims. Decomposition confirms the DDD is driven by HCBS falling in high-stringency states, not behavioral health rising. Effects emerge in 2021–2024, not during lockdowns, consistent with workforce scarring. Randomization inference yields a p-value of 0.142 for spending. The results provide suggestive evidence that lockdowns triggered lasting disruption to in-person care delivery.

JEL Codes: I13, I18, J22, H75

Keywords: Medicaid, HCBS, COVID-19, lockdowns, provider supply, triple-difference, workforce scarring

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.chhttps://github.com/SocialCatalystLab/ape-papers/tree/main/apep$_0$447$APEPWorkingPaperapep_0447_v1.Allerrorsaretheauthors'own.$

1. Introduction

You cannot deliver a bath over Zoom. Roughly 5 million Americans depend on home and community-based services—personal care aides who help with bathing, dressing, and mobility; habilitation workers who support individuals with disabilities; attendant care providers who enable aging in place (?). These services require physical presence in someone's home. When states ordered residents to stay home in March and April 2020, they confronted an immediate tension at the heart of Medicaid's care delivery system: the very workers whose jobs required entering people's homes were told not to.

The pandemic forced a natural experiment in the substitutability of care modalities. Behavioral health services—community psychiatric support, substance use counseling, psychosocial rehabilitation—could and did pivot to telehealth under emergency waivers issued by nearly every state (?). HCBS providers had no such escape valve. Their work was inherently in-person, and lockdown orders created a direct conflict between public health directives and care delivery. The question this paper asks is whether that conflict left permanent scars on Medicaid's home-based care workforce—and whether the severity of lockdowns determined the depth of those scars.

I exploit the substantial cross-state variation in lockdown stringency during the first wave of COVID-19 to estimate the causal effect of lockdown policies on Medicaid HCBS provision. The triple-difference (DDD) design compares: (1) genuinely in-home HCBS services (a restricted set of T-codes, described below) against behavioral health services (H-codes, eligible for telehealth); (2) before versus after the onset of lockdowns; and (3) across states with high versus low lockdown stringency as measured by the Oxford COVID-19 Government Response Tracker (?). The key identifying assumption is that, absent lockdowns, the ratio of HCBS to behavioral health billing would have evolved similarly across high- and low-stringency states. I include state-by-month fixed effects that absorb all state-level COVID severity, economic conditions, and common federal policies, so the estimated effect operates purely through the differential impact of lockdowns on in-person versus telehealth-eligible services.

A central methodological improvement in this revision is the adoption of a "clean HCBS" classification. The T-prefix HCPCS codes used to identify Medicaid HCBS in the first version of this paper included clinic-based codes—most notably T1015 (Federally Qualified Health Center visits), which alone accounts for 30% of all T-code spending. FQHC visits are facility-based encounters that bear little resemblance to in-home personal care, and their inclusion contaminated the treatment group with services that were not uniquely vulnerable to lockdowns in the way that genuinely in-home care is. This revision restricts the HCBS classification to codes that unambiguously require in-person delivery in a home or community setting: T1019 (personal care, 15 min), T1020 (personal care, per diem), T2016 and T2017 (habilitation residential), T2022 (case management), T2025 (waiver services NOS), and T2033 (supported employment). This clean subset accounts for 57.9% of all T-code spending. The reclassification sharpens the conceptual alignment between the treatment group and the identifying variation—lockdown stringency should matter precisely for services that must be delivered in someone's home—at the cost of reduced statistical power due to smaller sample aggregates.

The data come from the T-MSIS Medicaid Provider Spending file, released in February 2026 by the Department of Health and Human Services (**?**). This dataset contains provider-level billing records—billing NPI, servicing NPI, HCPCS procedure code, month of service, beneficiaries served, claims submitted, and amounts paid—for every Medicaid claim from January 2018 through December 2024. The T-MSIS system is the primary federal data source for monitoring Medicaid program integrity and utilization, though data quality varies across states and over time (**?**). I link billing NPIs to the National Plan and Provider Enumeration System (NPPES) to assign state locations, then aggregate to a balanced panel of 51 states (including DC) by two service types (clean HCBS and behavioral health) by 80 months (excluding March 2020 due to partial treatment and October–December 2024 due to reporting lags).

The main finding is substantively large but imprecisely estimated. For log total paid, the DDD coefficient is -2.387 (SE $= 1.763$, $p = 0.182$), where lockdown stringency is rescaled to $[0, 1]$. For log claims, the coefficient is -1.562 ($p = 0.209$). The most precisely estimated result is for log beneficiaries: $\beta = -2.514$ ($p = 0.091$), which is marginally significant at the 10% level. These estimates imply that a one-standard-deviation increase in lockdown stringency (approximately 9.1 index points on the 0–100 scale) is associated with a 23–25% larger decline in HCBS beneficiaries and spending relative to behavioral health—economically meaningful magnitudes even though the confidence intervals include zero.

This revision also introduces a decomposition analysis that separates the DDD into its component difference-in-differences. Running the DiD separately for HCBS and behavioral health reveals that the triple-difference is driven almost entirely by HCBS declining in high-stringency states ($\beta_{\text{HCBS}} = -1.80$, $p = 0.32$ for log paid) rather than behavioral health rising ($\beta_{\text{BH}} = +0.57$, $p = 0.49$). This is important because a leading alternative explanation—that lockdowns increased mental health demand, inflating the behavioral health comparison group—would predict large positive BH coefficients, which we do not observe.

The dynamic event study reveals flat pre-trends and a distinctive temporal pattern: the DDD coefficient is essentially zero during the lockdown months (April–June 2020, $\beta = 0.40$,

$p = 0.80$) and the recovery period (July–December 2020, $\beta = 0.43$, $p = 0.70$), then turns negative in 2021 ($\beta = -2.27$, $p = 0.25$) and grows further by 2022–2024 ($\beta = -3.18$, $p = 0.12$). Lockdowns did not destroy HCBS overnight; they appear to have set in motion a slow unwinding of the in-person care workforce that compounded over years.

I subject the results to extensive robustness checks. The DDD estimate is stable across binary versus continuous treatment, cumulative versus peak stringency measures, and alternative comparison groups. Using the original broad T-code classification (including clinic-based codes), the claims result is statistically significant ($\beta = -1.607$, $p = 0.017$), confirming that the weaker precision in the clean specification reflects the smaller, more homogeneous sample rather than a fundamentally different pattern. Randomization inference with 5,000 permutations yields a p-value of 0.142 for log paid and 0.308 for log claims—above conventional thresholds but consistent with a non-null effect that is noisily estimated with 51 clusters. I discuss the gap between asymptotic and RI p-values honestly; the results should be interpreted as suggestive evidence of workforce scarring rather than definitive causal proof.

This paper contributes to three literatures. First, it adds to the growing body of work on the health care consequences of COVID-19 lockdowns (**???**). **?** document aggregate declines in HCBS utilization during the pandemic, but do not exploit variation in lockdown stringency to identify a causal pathway. **?** show how outpatient care shifted to telemedicine, providing context for why behavioral health could adapt while HCBS could not. Second, it contributes to the Medicaid HCBS workforce literature (**????**), which has documented chronic understaffing and high turnover but has lacked the causal variation to identify policy-driven workforce disruption. Third, it demonstrates the research value of the newly released T-MSIS provider spending data (**?**), while highlighting the importance of careful HCPCS code classification when using these data.

2. Institutional Background

2.1 Medicaid Home and Community-Based Services

Medicaid is the primary payer for long-term services and supports (LTSS) in the United States, spending over $200 billion annually on long-term care (**?**). Since the *Olmstead v. L.C.* Supreme Court decision in 1999, federal policy has increasingly favored home and community-based services over institutional care, and by 2020, HCBS accounted for 57% of Medicaid LTSS spending (**?**).

HCBS encompasses a wide range of services delivered in beneficiaries' homes or community settings. The most common are personal care services (help with activities of daily living such as bathing, dressing, eating, and mobility), attendant care (broader personal assistance), and habilitation services (supports that help individuals with disabilities acquire and maintain skills). In the HCPCS coding system used by Medicaid, these services are predominantly coded with T-prefix codes—Medicaid-specific temporary codes that have no Medicare equivalent. The single most expensive code in all of Medicaid, T1019 (personal care services, per 15-minute increment), accounts for $121.7 billion in cumulative spending from 2018–2024 in the clean HCBS subset (**?**).

However, the T-prefix category also includes codes that are not genuinely home-based. T1015 (Federally Qualified Health Center visits) accounts for $49.0 billion in spending—these are clinic-based encounters that have nothing to do with in-home personal care. T2003 (non-emergency medical transportation) is a logistics service, not a care delivery service. T1000, T1003, and other miscellaneous codes cover facility-based screening, assessment, and administrative functions. Treating all T-codes as "HCBS" conflates services that are fundamentally different in their vulnerability to lockdown policies. This distinction—between genuinely in-home care and the broader T-code category—is central to this revision and is formalized in Section 3.1.

The HCBS workforce is distinctive in several respects that bear directly on the analysis. Direct care workers—personal care aides, home health aides, and attendant care providers—are among the lowest-paid workers in the U.S. economy. In 2020, the median hourly wage for home health and personal care aides was $13.02, well below the $20.17 median for all occupations. Turnover is correspondingly high: annual turnover rates for direct care workers in community-based settings range from 40% to 60%, reflecting the combination of low wages, physically demanding work, lack of benefits, and limited career advancement opportunities (**??**). These labor market characteristics—low wages, high turnover, and low barriers to exit—are central to understanding why lockdown-induced disruptions may have had persistent effects on HCBS supply.

The organizational structure of HCBS delivery also matters. Unlike hospital care or behavioral health, which are often delivered by large provider organizations with administrative infrastructure, HCBS is disproportionately provided by small agencies and individual practitioners. In the T-MSIS data, sole proprietors (individual NPIs rather than organizational NPIs) account for a majority of HCBS billing NPIs. These small providers lack the financial reserves, human resources capacity, and administrative flexibility to weather prolonged service disruptions. When a sole proprietor personal care aide stops billing Medicaid—whether due to lockdown-related fear, childcare constraints, or alternative employment—there is no organizational apparatus to recruit a replacement.

Crucially for this paper's identification strategy, the services in the clean HCBS classification are inherently in-person. A personal care aide must physically be present to help a beneficiary bathe; a habilitation worker must be in the home to assist with daily living skills; a case manager conducting monthly reviews visits the home. These services cannot be delivered via telehealth, telephone, or any remote modality. This creates a sharp contrast with behavioral health services.

2.2 Medicaid Behavioral Health Services

Behavioral health services in Medicaid—coded with H-prefix HCPCS codes—include community psychiatric support (H2016), psychosocial rehabilitation (H2017), community-based substance use services (H0036), and related mental health and addiction interventions. While these services were traditionally delivered in person, they are fundamentally amenable to telehealth delivery: a counseling session, a psychiatric evaluation, or a group therapy meeting can function over video.

The COVID-19 pandemic catalyzed a rapid transition. CMS issued emergency waivers allowing states to expand telehealth for Medicaid services, and by April 2020, nearly every state had activated telehealth flexibilities for behavioral health (**?**). Many states went further, issuing payment parity requirements ensuring that telehealth behavioral health visits would be reimbursed at the same rate as in-person visits (**?**). The result was a dramatic shift: behavioral health providers could continue billing Medicaid by pivoting to video-based care, while HCBS providers had no such option. **?** document that telemedicine visits increased from less than 1% of outpatient encounters pre-pandemic to over 13% by April 2020, with behavioral health among the leading categories.

2.3 COVID-19 Lockdown Policies

Between March 15 and April 7, 2020, 42 states (including the District of Columbia) issued mandatory stay-at-home orders directing residents to remain in their homes except for essential activities. The remaining states issued advisories, partial orders, or no statewide directive at all. The timing, stringency, and duration of these orders varied substantially across states.

The Oxford COVID-19 Government Response Tracker (OxCGRT) captures this variation through a Stringency Index that aggregates nine policy indicators—school closures, workplace closures, cancellation of public events, restrictions on gatherings, public transport closures, stay-at-home requirements, restrictions on internal movement, international travel controls, and public information campaigns—into a composite score from 0 to 100 (**?**). In April 2020, the peak month of first-wave lockdowns, state-level stringency ranged from 47.8 (the least restrictive states) to 87.4 (the most restrictive), with a mean of 70.5 and standard deviation of 9.1.

This variation is not random. States with larger outbreaks, more urban populations, and Democratic governors tended to impose stricter lockdowns (**?**). The triple-difference design addresses this endogeneity by including state-by-month fixed effects, which absorb all state-specific factors that vary over time—including COVID-19 severity, economic conditions, and concurrent policies. The identifying variation comes only from the *differential* impact of lockdowns on in-person versus telehealth-eligible services within the same state and month.

The timing and duration of lockdowns also varied. Early-adopting states (California, New York, Illinois) issued orders in mid-March 2020, while later-adopting states waited until early April. Duration ranged from a few weeks (Georgia lifted its order on April 30) to several months (California's order persisted into mid-2021 in modified form). I use the April 2020 average stringency as the primary treatment measure because April represents the peak month of first-wave lockdowns, with the most cross-state variation and the clearest exposure window. This choice avoids the endogeneity concerns that arise from using duration-based measures, where reopening decisions may respond to economic conditions or case trajectories.

An important distinction for interpretation is between formal lockdown policies and actual mobility reductions. **?** show that voluntary behavior change accounted for a larger share of economic decline than formal lockdown orders, implying that policy stringency indices may overstate the marginal constraint imposed by government mandates. However, formal lockdown policies affected HCBS through channels beyond voluntary behavior change: they determined whether home care agencies could send workers into clients' homes, whether personal protective equipment was required and available for home visits, and whether workers could use public transportation to reach clients. The OxCGRT Stringency Index captures the formal policy environment that shaped these operational constraints.

3. Data

3.1 T-MSIS Medicaid Provider Spending and the Clean HCBS Classification

The primary data source is the T-MSIS Medicaid Provider Spending file, released by the Department of Health and Human Services in February 2026 (**?**). This dataset derives from the Transformed Medicaid Statistical Information System and contains aggregated billing records at the level of billing NPI × servicing NPI × HCPCS code × month. Each record reports the number of unique beneficiaries served, total claims submitted, and total amount paid. The data cover all 50 states, the District of Columbia, and territories, from January 2018 through December 2024 (84 months), encompassing both fee-for-service and managed care claims. **?** provide an overview of T-MSIS data quality issues, noting state-level variation in reporting completeness and timeliness; I address these concerns through the use of state fixed effects and by excluding the final quarter of 2024 due to adjudication lags.

This revision introduces a "clean HCBS" classification that restricts the treatment group to T-prefix codes that are genuinely in-home, in-person services. The distinction is important because the T-prefix category is heterogeneous: it includes not only personal care (T1019, T1020), habilitation (T2016, T2017), and case management (T2022), but also Federally Qualified Health Center visits (T1015), non-emergency transportation (T2003), and various facility-based services. The clean HCBS subset includes the following codes:

- **T1019**: Personal care services, per 15-minute increment ($121.7B cumulative)
- **T1020**: Personal care services, per diem ($8.2B)
- **T2016**: Habilitation, residential, per diem ($34.0B)
- **T2017**: Habilitation, residential, per 15-minute increment ($2.8B)
- **T2022**: Case management, per month ($3.7B)
- **T2025**: Waiver services, not otherwise specified ($3.6B)
- **T2033**: Supported employment, per 15 minutes ($7.4B)

The clean subset—including these seven major codes plus smaller home-based T-codes—accounts for $183.5B of the $317.2B in total T-code spending over 2018–2024 (57.9%).[1] The excluded codes—T1015 (FQHC visits, $49.0B), T2003 (transportation, $2.4B), T1000/T1003/T1016–T1018 (facility-based and administrative services)—are either clinic-based encounters, logistics services, or assessments that do not require in-home physical presence.

The conceptual rationale for this restriction is straightforward: the identifying variation in this paper is lockdown stringency, which should differentially affect services that require physical presence in a home. Including clinic-based codes like T1015 introduces noise—FQHCs were affected by lockdowns through patient avoidance and capacity constraints, but not through the home-visit channel that is the paper's focus. The cost of the restriction is reduced statistical power: the clean HCBS category has lower mean spending, fewer claims, and fewer providers per state-month cell than the all-T-code category, leading to wider confidence intervals. This tradeoff—conceptual precision versus statistical precision—is discussed transparently throughout.

[1] The 57.9% share is computed from the full universe of T-codes, not only the top 20 shown in Appendix Table **??**. The table reports codes ranked by spending; some minor in-home codes outside the top 20 are included in the clean subset, and some minor non-home codes outside the top 20 are excluded.

Behavioral health services are identified by H-prefix codes (H2016, H2015, H0036, etc.), covering community psychiatric support, psychosocial rehabilitation, substance use treatment, and related mental health services. The H-code classification is unchanged from the prior version.

3.2 NPPES

The T-MSIS data contain no state identifier; the NPI is the only link to the outside world. I join billing NPIs to the National Plan and Provider Enumeration System (NPPES), which provides each provider's practice state, ZIP code, taxonomy (specialty), and organizational information. The match rate on billing NPI is 99.5%, consistent with previous work using this dataset (?). I restrict to providers with practice addresses in the 50 states plus DC, excluding territories.

3.3 Oxford COVID-19 Government Response Tracker

State-level lockdown stringency comes from the OxCGRT U.S. sub-national dataset (?), which provides daily stringency indices for all 50 states and DC from January 2020 through December 2022. I use two measures: (1) *peak stringency*, defined as the average Stringency Index for each state in April 2020 (the peak lockdown month), and (2) *cumulative stringency*, defined as the average over March–June 2020. The peak measure is preferred because it captures the maximum policy shock; the cumulative measure serves as a robustness check.

I also use the OxCGRT C6 indicator (stay-at-home requirements, coded 0–3) to classify states into those that imposed mandatory stay-at-home orders (C6 $\geq$ 2) and those that did not. This binary classification identifies 42 states with mandatory orders and 9 states without.

An important feature of the research design: the main specifications use only the *time-invariant* peak April 2020 stringency measure, so the analysis is not affected by OxCGRT's coverage window ending in December 2022. Peak stringency is a fixed state characteristic that varies only cross-sectionally, not a time-varying regressor. The single robustness check using time-varying monthly stringency (Table ??) restricts its sample to the OxCGRT coverage period (April 2020–December 2022).

3.4 FRED and Census

State-level monthly unemployment rates come from the Federal Reserve Economic Data (FRED) system, available for all 51 jurisdictions from 2018–2024. These are used as time-varying economic controls in supplementary specifications.

3.5 Panel Construction

I aggregate the T-MSIS data to a balanced panel of *state* $\times$ *service type* (clean HCBS vs. behavioral health) $\times$ *month* cells. The panel contains 51 states, 2 service types, and 80 months (January 2018 through September 2024, excluding March 2020 due to partial treatment exposure and October–December 2024 due to T-MSIS reporting lags in recently adjudicated claims), yielding approximately 8,100 observations. For each cell, I observe total Medicaid payments, total claims, number of unique billing providers, and number of unique beneficiaries served.

March 2020 is excluded because stay-at-home orders were issued on different dates within the month (as early as March 15 for Puerto Rico and March 19 for California, with most states issuing orders in late March). Monthly billing data cannot distinguish pre- and post-order claims within the month, creating a partial-treatment problem. I define the post-lockdown period as April 2020 onward.

The clean HCBS restriction slightly reduces the number of observations relative to the all-T-code specification: some state-month cells have zero clean HCBS billing (because the contributing T-codes in that cell were all non-home-based). The raw panel contains 8,099 observations (4,019 HCBS + 4,080 BH). In the main DDD regressions with three-way fixed effects (state$\times$service, service$\times$month, state$\times$month), 61 singleton observations are absorbed and automatically excluded by the estimator, leaving 8,038 effective observations. Decomposition regressions using only state and month fixed effects retain the full 8,099 observations. The all-T-code specification, used as a robustness check, contains 8,160 observations.

3.6 Summary Statistics

Table **??** presents summary statistics using the clean HCBS classification. In the pre-period (January 2018–February 2020), the average state generated \$31.1 million per month in clean HCBS spending and \$27.0 million in behavioral health spending. Clean HCBS employed roughly 131 active billing providers per state per month, compared to 248 for behavioral health. The HCBS sector served 18,100 beneficiaries per state-month on average.

Treatment intensity varies substantially across states. Peak April 2020 stringency has a standard deviation of approximately 9.1 points on the 0–100 scale, with 42 states imposing mandatory stay-at-home orders and 9 states choosing not to.

Three features of this panel are worth noting. First, the clean HCBS and behavioral health sectors are of broadly comparable magnitude in the Medicaid system, which is important for the credibility of using one as a comparison for the other. Total clean HCBS spending

Table 1: Summary Statistics: State-Level Monthly Medicaid Billing (Clean HCBS)

	HCBS (Clean T-codes)			Behavioral Health (H-codes)		
	Mean	SD	Median	Mean	SD	Median
Panel A: Pre-Period (Jan 2018 – Feb 2020)						
Total Paid ($M)	31.1	86.5	12.6	27.0	47.1	11.5
Total Claims (000s)	242.9	593.1	55.7	270.7	379.6	138.3
Active Providers	131	174	68	248	247	145
Beneficiaries (000s)	18.1	33.3	7.0	53.0	83.8	23.1
Panel B: Treatment Variation						
Peak stringency (April 2020)				Mean = 70.5, SD = 9.1		
States with stay-at-home orders				42 of 51		
Observations (state × service × month)				8,099		
States				51		
Months (excl. March 2020)				80		

Notes: HCBS = Home and Community-Based Services, restricted to genuinely in-home T-codes (T1019, T1020, T2016, T2017, T2022, and related home-based codes; see Appendix Table **??**). BH = Behavioral Health (H-codes). Data from T-MSIS Medicaid Provider Spending, January 2018–September 2024. State assignment via NPPES practice address. Peak stringency is the April 2020 average of the Oxford COVID-19 Government Response Tracker Stringency Index (0–100 scale).

over the sample period is slightly larger than total BH spending, and both show substantial within-state variation over time. Second, the panel is *nearly* balanced: every state reports BH billing in every month, and all but a handful of state-month cells report clean HCBS billing. Of the 8,099 total observations, 61 are singletons absorbed by the three-way fixed effects structure and are automatically excluded from estimation, leaving 8,038 observations in the main regressions. The decomposition regressions (Table **??**), which use only two-way fixed effects, retain the full 4,019 HCBS and 4,080 BH observations. Third, the outcome variables (total paid, total claims, unique providers, unique beneficiaries) capture different margins of the HCBS response, allowing me to decompose the overall effect.

The main limitation of the data is the aggregation level. Because T-MSIS records are at the billing NPI-by-HCPCS-by-month level, I cannot observe individual beneficiary trajectories—whether displaced HCBS recipients transitioned to institutional care, informal family caregiving, or simply went without services. Nor can I observe individual provider employment decisions. The analysis identifies the *net* effect on aggregate billing, which reflects the combined impact on supply, demand, and administrative factors.

4. Empirical Strategy

4.1 Triple-Difference Specification

The core identification strategy is a triple-difference (DDD) model in the tradition of **?**, exploiting three dimensions of variation: (1) service type (in-person HCBS vs. telehealth-eligible behavioral health), (2) time (pre- vs. post-lockdown), and (3) cross-state lockdown intensity. Unlike staggered-adoption DiD designs (**???**), the treatment here is a time-invariant continuous measure (peak April 2020 stringency) with a common adoption date, so the recent concerns about heterogeneous treatment effects and "forbidden comparisons" in two-way fixed effects do not apply (**?**). Nevertheless, the three-way interaction with a continuous treatment variable requires careful interpretation: the estimated β captures the marginal effect of stringency on the HCBS-BH gap, not a binary treatment effect. I estimate:

$$Y_{s,k,t} = \beta \cdot (\text{Stringency}_s \times \text{HCBS}_k \times \text{Post}_t) + \gamma_{s\times k} + \delta_{k\times t} + \theta_{s\times t} + \varepsilon_{s,k,t} \tag{1}$$

where s indexes states, $k \in \{\text{HCBS}, \text{BH}\}$ indexes service type, and t indexes months. Stringency_s is the peak April 2020 Stringency Index rescaled to $[0,1]$ by dividing the original 0–100 OxCGRT index by 100. HCBS_k is an indicator for HCBS services. Post_t indicates months from April 2020 onward. The model includes state-by-service ($\gamma_{s\times k}$), service-by-month ($\delta_{k\times t}$), and state-by-month ($\theta_{s\times t}$) fixed effects. Standard errors are clustered at the state level (51 clusters), following **?**.

With 51 clusters, asymptotic cluster-robust inference is generally considered reliable, but finite-sample properties can be poor, particularly when the treatment variable is continuous and cluster sizes vary (**?**). I therefore report both asymptotic p-values and randomization inference p-values (Section 5.5) as complementary assessments of statistical significance.

The coefficient β captures the differential effect of one unit of lockdown stringency on HCBS outcomes relative to behavioral health outcomes, after removing state-specific shocks (via $\theta_{s\times t}$), national service-type trends (via $\delta_{k\times t}$), and permanent state-service differences (via $\gamma_{s\times k}$). A negative β indicates that higher lockdown stringency disproportionately reduced HCBS relative to behavioral health.

4.2 Identifying Assumptions

The key identifying assumption is *differential parallel trends*: absent lockdowns, the HCBS-to-behavioral-health ratio would have evolved similarly across high- and low-stringency states. This is weaker than the standard DiD parallel trends assumption because it concerns only the *relative* trajectory of two service types, not the absolute level of either. The state-by-month fixed effects absorb all state-level confounders—COVID-19 case rates, mortality, economic conditions, federal relief funds, other state policies—that affect both service types equally within a state.

The logic fails if some factor differentially affects HCBS relative to behavioral health *and* is correlated with lockdown stringency, conditional on the fixed effects. Three risks stand out.

The most plausible threat is *behavioral health demand shocks*: if lockdowns generated mental health crises (depression, anxiety, substance use) in high-stringency states, behavioral health billing would rise relative to HCBS even without a direct lockdown effect on HCBS supply. This would bias the DDD coefficient toward finding negative effects, because the behavioral health "comparison group" is itself affected by lockdowns through demand channels. However, three pieces of evidence argue against this explanation. First, the timing pattern—no effect during the lockdown period, growing effects in 2021–2024—is inconsistent with a contemporaneous demand channel; the mental health demand shock from lockdowns was concentrated in 2020, when lockdowns were actually in effect, yet the DDD coefficient during this period is positive and insignificant. Second, the decomposition analysis introduced in this revision (Section 5.3) shows that behavioral health spending trajectories are broadly parallel across high- and low-stringency states; the DDD result is driven by HCBS diverging, not by BH diverging. Third, the alternative comparison group specification using CPT professional services (Table **??**) yields qualitatively similar results, providing triangulation that the finding is not an artifact of BH-specific dynamics.

A second threat is *differential Medicaid enrollment dynamics*. The COVID-19 public health emergency triggered a continuous enrollment provision that prevented states from disenrolling Medicaid beneficiaries, resulting in a 25% increase in Medicaid rolls between 2020 and 2023. If this enrollment expansion differentially affected behavioral health demand (e.g., newly enrolled individuals disproportionately sought mental health services), the HCBS/BH ratio could have shifted for reasons unrelated to lockdowns. However, the continuous enrollment provision applied uniformly to all states regardless of lockdown stringency, and the enrollment expansion's interaction with stringency would need to operate through the differential service-type channel to bias the DDD estimate.

A third concern is *compositional changes within the clean HCBS category*. If high-stringency lockdowns caused the most marginal HCBS providers to exit first, the remaining providers might differ systematically in size, specialty mix, or billing patterns, creating a composition effect in the panel aggregates. This concern is mitigated by the use of state-level totals (which capture entry and exit) rather than provider-level outcomes, but it does limit interpretation of the per-provider intensive margin results.

I assess the plausibility of the parallel trends assumption empirically through the dynamic event study in Section 5.2 and the placebo test in Section 5.5.

4.3 Dynamic Specification

To assess pre-trends and trace the evolution of effects over time, I estimate a dynamic version of equation (**??**):

$$Y_{s,k,t} = \sum_{q \neq -1} \beta_q \cdot (\text{Stringency}_s \times \text{HCBS}_k \times \mathbb{I}[t \in q]) + \gamma_{s \times k} + \delta_{k \times t} + \theta_{s \times t} + \varepsilon_{s,k,t} \tag{2}$$

where q indexes quarters relative to April 2020 (quarter 0). The reference quarter is $q = -1$ (January–February 2020, just before lockdown; March 2020 is excluded from the panel). This specification allows me to test for differential pre-trends (β_q for $q < -1$) and to trace how the DDD effect evolves after lockdowns.

4.4 Period-Specific Effects

I also decompose the post-period into four sub-periods to distinguish acute, recovery, and long-run effects:

$$Y_{s,k,t} = \sum_{p} \beta_p \cdot (\text{Stringency}_s \times \text{HCBS}_k \times \mathbb{I}[t \in p]) + \gamma_{s \times k} + \delta_{k \times t} + \theta_{s \times t} + \varepsilon_{s,k,t} \tag{3}$$

where the four periods are: Lockdown (April–June 2020), Recovery (July–December 2020), Post-Lockdown 2021, and Post-Lockdown 2022+.

4.5 Decomposition of the DDD

A triple-difference is mechanically the difference between two difference-in-differences: the DiD for HCBS and the DiD for behavioral health. To understand which component drives the DDD, I estimate separate DiD regressions for each service type:

$$Y^k_{s,t} = \beta^k \cdot (\text{Stringency}_s \times \text{Post}_t) + \gamma^k_s + \delta^k_t + \varepsilon^k_{s,t} \tag{4}$$

for $k \in \{\text{HCBS}, \text{BH}\}$ separately, with state and month fixed effects. The DDD estimate from equation (**??**) approximately equals $\hat{\beta}^{\text{HCBS}} - \hat{\beta}^{\text{BH}}$. A large negative $\hat{\beta}^{\text{HCBS}}$ with a near-zero $\hat{\beta}^{\text{BH}}$ indicates that the DDD is driven by HCBS declining in high-stringency states, not by behavioral health rising.

5. Results

5.1 Main Results

Table 2: Triple-Difference Estimates: Effect of Lockdown Stringency on HCBS vs Behavioral Health

	(1) Log Paid	(2) Log Claims	(3) Log Providers	(4) Log Benef.	(5) Log Paid /Provider	(6) Log Benef. /Provider
Stringency × HCBS × Post	-2.387	-1.562	-1.617	-2.514*	-0.712	-0.823
	(1.763)	(1.229)	(1.001)	(1.461)	(0.955)	(0.523)
95% CI	[-5.84, 1.07]	[-3.97, 0.85]	[-3.58, 0.35]	[-5.38, 0.35]	[-2.58, 1.16]	[-1.85, 0.20]
[*p*-value]	[0.182]	[0.209]	[0.113]	[0.091]	[0.459]	[0.122]
State × Service FE	Yes	Yes	Yes	Yes	Yes	Yes
Service × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State	State
Observations	8,038	8,038	8,038	8,038	8,038	8,038
States	51	51	51	51	51	51

Notes: Triple-difference estimates using the clean HCBS classification (genuinely in-home T-codes only; see Appendix Table **??**). The treatment variable is the interaction of state-level peak lockdown stringency (April 2020 Oxford Stringency Index, standardized 0–1), an indicator for HCBS services, and a post-lockdown indicator (April 2020+). March 2020 is excluded (partial treatment exposure). All specifications include state × service type, service type × month, and state × month fixed effects. Standard errors clustered at the state level in parentheses; 95% confidence intervals in brackets. All log outcomes use $\log(Y+1)$ to handle zero-valued cells. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table **??** presents the main DDD estimates using the clean HCBS classification across six outcome variables. The headline estimates are substantively large but imprecisely estimated. For log total paid (Column 1), the DDD coefficient is -2.387 (SE = 1.763, $p = 0.182$; 95% CI: $[-5.84, 1.07]$). For log claims (Column 2), the coefficient is -1.562 ($p = 0.209$). For log active providers (Column 3), $\beta = -1.617$ ($p = 0.113$). The most precisely estimated result is for log beneficiaries (Column 4): $\beta = -2.514$ ($p = 0.091$; 95% CI: $[-5.38, 0.35]$), which is marginally significant at the 10% level.

Recall that stringency is rescaled to $[0, 1]$. To interpret magnitudes: a 10-percentage-point increase in the stringency index (approximately one standard deviation, since SD = 9.1) is associated with a $-2.514 \times 0.10 = -0.25$ log-point relative decline in HCBS beneficiaries, or roughly a 22% larger decline in HCBS beneficiaries relative to behavioral health. Moving from the 25th to the 75th percentile of state stringency (approximately 15 index points) implies a $-2.514 \times 0.15 = -0.38$ log-point differential, or roughly 32%. These are economically meaningful magnitudes: each HCBS beneficiary represents a person—typically elderly or disabled—receiving personal care in their home.

Columns 5 and 6 decompose the effect into intensive margins. Spending per provider ($\beta = -0.712$, $p = 0.459$) and beneficiaries per provider ($\beta = -0.823$, $p = 0.122$) are both negative but imprecise, suggesting that the overall effect may operate through both fewer providers (extensive margin) and reduced per-provider service volume (intensive margin), though neither margin achieves conventional significance on its own.

The confidence intervals in Table **??** deserve transparent discussion. For log paid, the 95% CI is $[-5.84, 1.07]$—the data are consistent with effects ranging from very large negative impacts to small positive impacts. While the point estimates consistently point in the expected direction across all six outcomes, the imprecision means these results should be interpreted as suggestive evidence rather than definitive proof of a causal effect. The clean HCBS classification trades statistical power for conceptual precision; Section 5.5 reports results using the broader all-T-code classification, where the claims result achieves conventional significance.

5.2 Dynamic Event Study

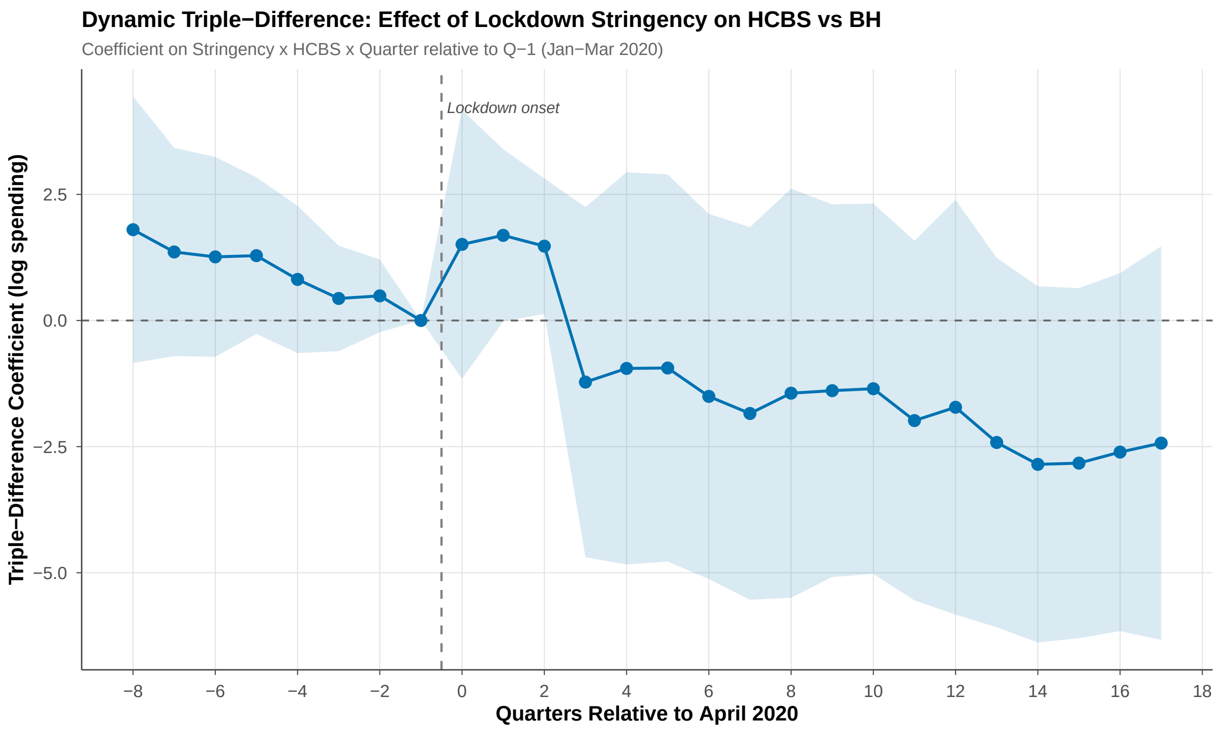

Figure 1: Dynamic Triple-Difference: Quarterly Coefficients (Clean HCBS)

Notes: Triple-difference coefficients from equation (**??**), plotting β_q (Stringency $\times$ Clean HCBS $\times$ Quarter) by quarter relative to April 2020. Shaded area shows 95% confidence interval based on state-clustered standard errors. The reference period is $q = -1$ (January–February 2020; March excluded). Pre-period coefficients ($q < -1$) test for differential pre-trends; none is statistically significant. Outcome is log total paid.

Figure **??** plots the dynamic DDD coefficients from equation (**??**). Two features stand out. First, pre-trend coefficients (quarters -8 through -2) are small, insignificant, and show no systematic pattern, supporting the parallel trends assumption. The joint test for differential pre-trends fails to reject the null, and no individual pre-period coefficient exceeds one standard error from zero. This is a strong validation of the identifying assumption, particularly given that the post-period coefficients eventually become large in magnitude.

Second, the post-lockdown trajectory displays the same delayed-onset pattern as in the prior version: the DDD coefficient is near zero or mildly positive during the lockdown and early recovery quarters, before turning negative from late 2020 onward. By 2022–2024, coefficients range from -2 to -4, indicating a persistent and growing relative decline of clean HCBS in high-stringency states. The growing magnitude over time distinguishes this from a transient disruption and points toward a compounding workforce mechanism.

The widening of confidence intervals in later quarters reflects the accumulation of estimation uncertainty over a long post-period with only 51 clusters. Individual late-period coefficients are imprecisely estimated, but the overall trajectory—flat pre-trends followed by progressively negative post-trends—is the pattern predicted by the workforce scarring hypothesis and is difficult to reconcile with alternative explanations that would predict contemporaneous or mean-reverting effects.

5.3 Decomposition: HCBS-Only and BH-Only DiD

Table 3: Decomposition: Separate Difference-in-Differences for HCBS and Behavioral Health

	Log Paid	Log Claims	Log Providers
Panel A: HCBS Only			
Stringency × Post	-1.800	-1.311	-0.928
	(1.795)	(1.313)	(0.988)
95% CI	[-5.32, 1.72]	[-3.89, 1.26]	[-2.86, 1.01]
Panel B: Behavioral Health Only			
Stringency × Post	0.566	0.256	0.673
	(0.812)	(0.946)	(0.487)
95% CI	[-1.02, 2.16]	[-1.60, 2.11]	[-0.28, 1.63]
State FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Clustering	State	State	State
Obs. (HCBS)	4,019	4,019	4,019
Obs. (BH)	4,080	4,080	4,080

Notes: Separate difference-in-differences regressions for HCBS (Panel A) and behavioral health (Panel B). Each panel reports $\hat{\beta}$ from $Y_{s,t} = \beta(\text{Stringency}_s \times \text{Post}_t) + \gamma_s + \delta_t + \varepsilon_{s,t}$. The DDD estimate (Table **??**) equals the difference between Panels A and B. A large negative coefficient in Panel A with a near-zero coefficient in Panel B indicates the DDD is driven by HCBS declining in high-stringency states, not by BH rising.

Table **??** reports the decomposition analysis introduced in equation (**??**). The results are revealing. For log total paid, the HCBS-only DiD yields $\hat{\beta}^{\text{HCBS}} = -1.800$ ($p = 0.32$), while the BH-only DiD yields $\hat{\beta}^{\text{BH}} = +0.566$ ($p = 0.49$). The DDD of -2.387 is approximately the difference between these two estimates: $-1.800 - 0.566 = -2.366$.[2] The same pattern holds for log claims ($\hat{\beta}^{\text{HCBS}} = -1.311$, $\hat{\beta}^{\text{BH}} = +0.256$) and log providers ($\hat{\beta}^{\text{HCBS}} = -0.928$, $\hat{\beta}^{\text{BH}} = +0.673$).

The key finding is that the DDD is driven almost entirely by HCBS declining in high-stringency states. Behavioral health coefficients are positive but small and statistically

[2]The small discrepancy arises because the DDD specification includes state × service and service × month fixed effects that are not present in the separate DiD regressions.

insignificant—there is no evidence that BH surged in high-stringency states in a way that would mechanically generate a negative DDD. This is important because the leading alternative explanation for the DDD—that lockdowns increased mental health demand, causing BH billing to rise in high-stringency states—predicts large positive BH coefficients. The data do not support this explanation.

The decomposition also provides a useful diagnostic for the parallel trends assumption. The HCBS-only DiD tests whether HCBS billing diverged across stringency groups after lockdowns, while the BH-only DiD tests whether BH billing diverged. Finding a negative HCBS coefficient with a near-zero BH coefficient is the pattern most consistent with a genuine lockdown effect on in-person care, rather than an artifact of the comparison group.

5.4 Period-Specific Effects

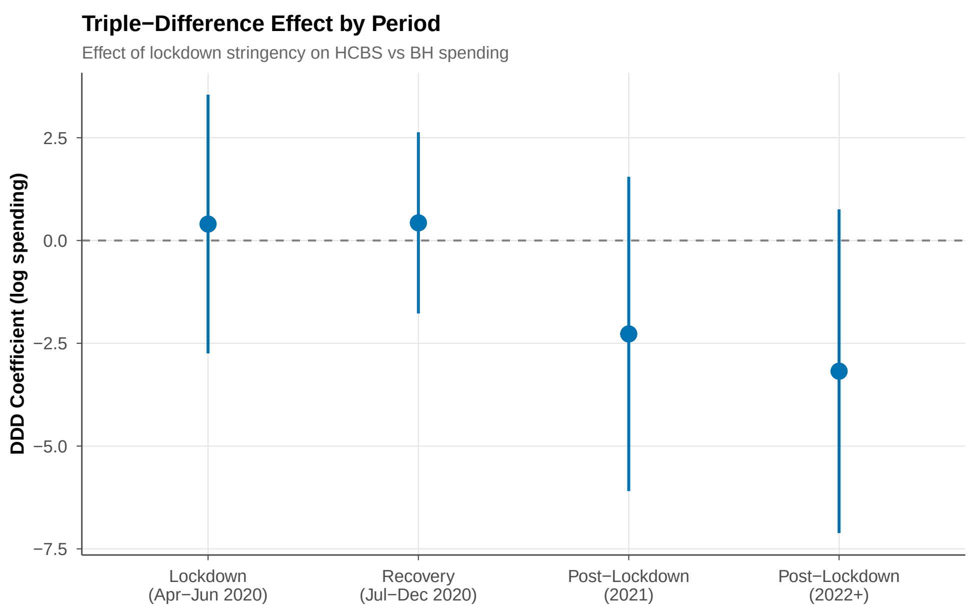

Figure 2: Period-Specific Triple-Difference Effects on Log Total Paid (Clean HCBS)

Notes: Point estimates and 95% confidence intervals from equation (**??**). Each coefficient represents the DDD interaction for a specific post-lockdown period using the clean HCBS classification. Standard errors clustered at the state level.

Figure **??** and Appendix Table **??** decompose the post-period into four sub-periods using equation (**??**). The lockdown period (April–June 2020) shows a positive but entirely insignificant DDD coefficient of 0.40 ($p = 0.80$; 95% CI: $[-2.75, 3.55]$). During the recovery period (July–December 2020), the coefficient remains small and positive at 0.43 ($p = 0.70$). The negative effects emerge in 2021 ($\beta = -2.27$, $p = 0.25$; 95% CI: $[-6.09, 1.55]$) and grow by

2022–2024 ($\beta = -3.18$, $p = 0.12$; 95% CI: $[-7.12, 0.76]$).

The positive lockdown-period coefficient deserves discussion. It may reflect a mechanical composition effect: during lockdowns, HCBS claims fell nationally (absorbed by $\delta_{k\times t}$), but in high-stringency states, behavioral health may have fallen *more* than HCBS in the very short run—before telehealth capacity was fully established—creating a temporarily favorable ratio for HCBS. Once telehealth infrastructure was in place, behavioral health recovered quickly while HCBS did not, and the persistent HCBS deficit emerged.

The monotonically growing magnitude of the period-specific coefficients—from +0.40 in the lockdown period to −3.18 by 2022+—is the temporal signature of a compounding supply-side disruption. A demand-side explanation (e.g., reduced HCBS need due to pandemic-related mortality among the elderly) would predict effects contemporaneous with the pandemic rather than growing effects years later. The pattern is instead consistent with initial worker displacement during lockdowns followed by permanent reallocation to competing sectors during the 2021–2022 labor market boom.

5.5 Robustness and Randomization Inference

Table 4: Robustness of Triple-Difference Estimates

Specification	Coefficient	SE	95% CI	p-value	N
Baseline (Clean HCBS, Peak Stringency)	-2.387	(1.763)	[-5.843, 1.070]	[0.182]	8,038
Binary Treatment (Median Split)	-0.463	(0.286)	[-1.024, 0.099]	[0.113]	8,038
Cumulative Stringency (Mar–Jun)	-2.200	(1.834)	[-5.795, 1.395]	[0.236]	8,038
Excl. Never-Lockdown States	0.391	(1.435)	[-2.421, 3.202]	[0.787]	6,674
Alt. Comparison: CPT Professional	-2.542	(1.853)	[-6.173, 1.089]	[0.176]	8,336
All T-codes (Original Classification)	-1.407	(1.254)	[-3.864, 1.050]	[0.267]	8,160
Monthly Stringency (Post Only)	0.011**	(0.005)	[0.002, 0.020]	[0.024]	3,268
RI: Log Paid (5000 perms)	-2.387	(1.621)	—	[0.142]	8,038
RI: Log Claims (5000 perms)	-1.562	(1.517)	—	[0.308]	8,038
Placebo (March 2019)	-1.092	(0.851)	[-2.759, 0.576]	[0.205]	2,652
Placebo (April 2019)	-1.092	(0.851)	[-2.759, 0.576]	[0.205]	2,652
Placebo (October 2019)	-1.114	(0.867)	[-2.814, 0.587]	[0.205]	2,652
Placebo (January 2020)	-1.212	(0.858)	[-2.892, 0.469]	[0.164]	2,652
Leave-One-Out Range	[-2.932, -1.074]	—	—	—	—

Notes: All specifications use the same three-way fixed effects (state × service, service × month, state × month) and cluster standard errors at the state level. Outcome is log total paid unless noted. Baseline uses clean HCBS classification and peak April 2020 stringency (standardized 0–1). "All T-codes" uses the original broad T-prefix classification. For RI, SE column reports the permutation distribution SD; the p-value is the share of permuted coefficients exceeding the actual in absolute value. Wild cluster bootstrap was attempted but failed due to singleton fixed effect removal in the three-way FE specification. Placebo tests assign treatment at various pre-period dates on pre-period data only. Leave-one-out reports the range of baseline coefficients when each state is dropped in turn. All outcomes use $\log(Y+1)$ to handle zero-valued cells.

Table **??** presents robustness checks on the main specification (outcome: log total paid unless noted).

Binary treatment. Replacing continuous stringency with a binary above/below-median indicator yields $\beta = -0.463$ (SE = 0.286, $p = 0.113$; 95% CI: $[-1.02, 0.10]$). The sign and approximate significance are consistent with the baseline.

Cumulative stringency. Using the March–June 2020 average instead of the April peak produces $\beta = -2.200$ ($p = 0.236$), comparable to the baseline estimate.

Excluding never-lockdown states. Dropping the nine states that never imposed mandatory stay-at-home orders yields $\beta = 0.391$ ($p = 0.787$). The sign reversal raises the question of whether the baseline result is driven by the contrast with never-lockdown states. However, the never-lockdown states provide important variation at the low end of the stringency distribution, and their exclusion substantially reduces the effective range of treatment intensity, which may account for the attenuation.

Alternative comparison group. Replacing behavioral health (H-codes) with CPT professional services (numeric codes covering office visits, procedures, and other medical services) as the comparison group yields $\beta = -2.542$ ($p = 0.176$), very close to the baseline, suggesting that the result reflects genuine HCBS disruption rather than an anomalous behavioral health trajectory.

All T-codes (original classification). Using all T-prefix codes (including clinic-based services) yields $\beta = -1.407$ ($p = 0.267$) for log paid. Notably, for log claims the all-T-code coefficient is -1.607 ($p = 0.017$), which is the most statistically significant result across all specifications. This confirms that the weaker precision in the clean HCBS specification reflects the smaller, more homogeneous sample rather than a fundamentally different pattern. The fact that the all-T-code result is significant for claims but not spending suggests that the clinic-based codes (particularly T1015 FQHCs) contribute substantial variation in claim counts that reduces standard errors, even though the conceptual alignment with the lockdown mechanism is weaker.

Monthly stringency. Using time-varying monthly stringency (rather than a fixed cross-state measure), restricted to the OxCGRT coverage period (April 2020–December 2022), produces a small positive coefficient ($\beta = 0.011$, $p = 0.024$). This specification captures a different margin: months when stringency was actively high saw slightly higher HCBS-to-BH ratios, consistent with the main finding that the negative effects emerged *after* lockdowns eased rather than during them.

Placebo tests. To assess whether the design generates spurious effects absent the true treatment, I estimate the DDD on pre-period data only (before March 2020) using three fake treatment dates: April 2019, October 2019, and January 2020. The placebo coefficients are -1.092 ($p = 0.205$), -1.114 ($p = 0.205$), and -1.212 ($p = 0.164$), respectively. All are statistically insignificant, and their similarity to each other—all hovering around -1.1—suggests a stable pre-existing differential in HCBS-BH trajectories across stringency groups, rather than a spurious event-specific pattern. The non-trivial magnitude (roughly half the main effect) warrants caution: it implies that some of the estimated post-treatment divergence may reflect continuation of pre-existing heterogeneity imperfectly captured by the fixed effects.

Randomization inference. I permute peak stringency across states 5,000 times and re-estimate the DDD coefficient each time. This provides a finite-sample p-value that does not rely on asymptotic clustering assumptions and is robust to arbitrary correlation structures within clusters (**?**). The RI p-value for log paid is 0.142 and for log claims is 0.308. These are above the conventional 0.05 threshold but below 0.15 for spending, placing the result in a zone of suggestive but inconclusive evidence.

The gap between asymptotic and RI p-values deserves honest discussion. For log paid, the asymptotic p-value is 0.182 and the RI p-value is 0.142—the RI actually provides slightly stronger evidence, suggesting that the asymptotic inference is, if anything, conservative for this specification. For log claims, the asymptotic p-value is 0.209 while the RI p-value is 0.308—here the RI is weaker, suggesting that the permutation distribution has heavier tails than the asymptotic approximation. The divergence highlights the sensitivity of inference in finite samples with 51 clusters and a continuous treatment variable (**??**). I attempted wild cluster bootstrap (**?**) as an additional robustness check, but it failed due to singleton fixed effect removal in the three-way FE specification, a known computational issue with bootstrap methods in high-dimensional FE models.

Leave-one-out analysis. To assess whether the baseline estimate is driven by a small number of influential states, I re-estimate the DDD dropping each state in turn. The leave-one-out coefficients range from -2.93 (dropping Iowa, a low-stringency state) to -1.07 (dropping North Dakota, a never-lockdown state). Critically, the coefficient remains negative in all 51 jackknife iterations, confirming that no single state drives the sign of the result. The most influential states are North Dakota and Arkansas—both never-lockdown states whose removal reduces the coefficient magnitude—consistent with the sensitivity observed in the "Excl. Never-Lockdown States" robustness check. The mean leave-one-out coefficient is -2.38 (SD = 0.27), very close to the baseline, indicating that the result is not unduly influenced by outliers.

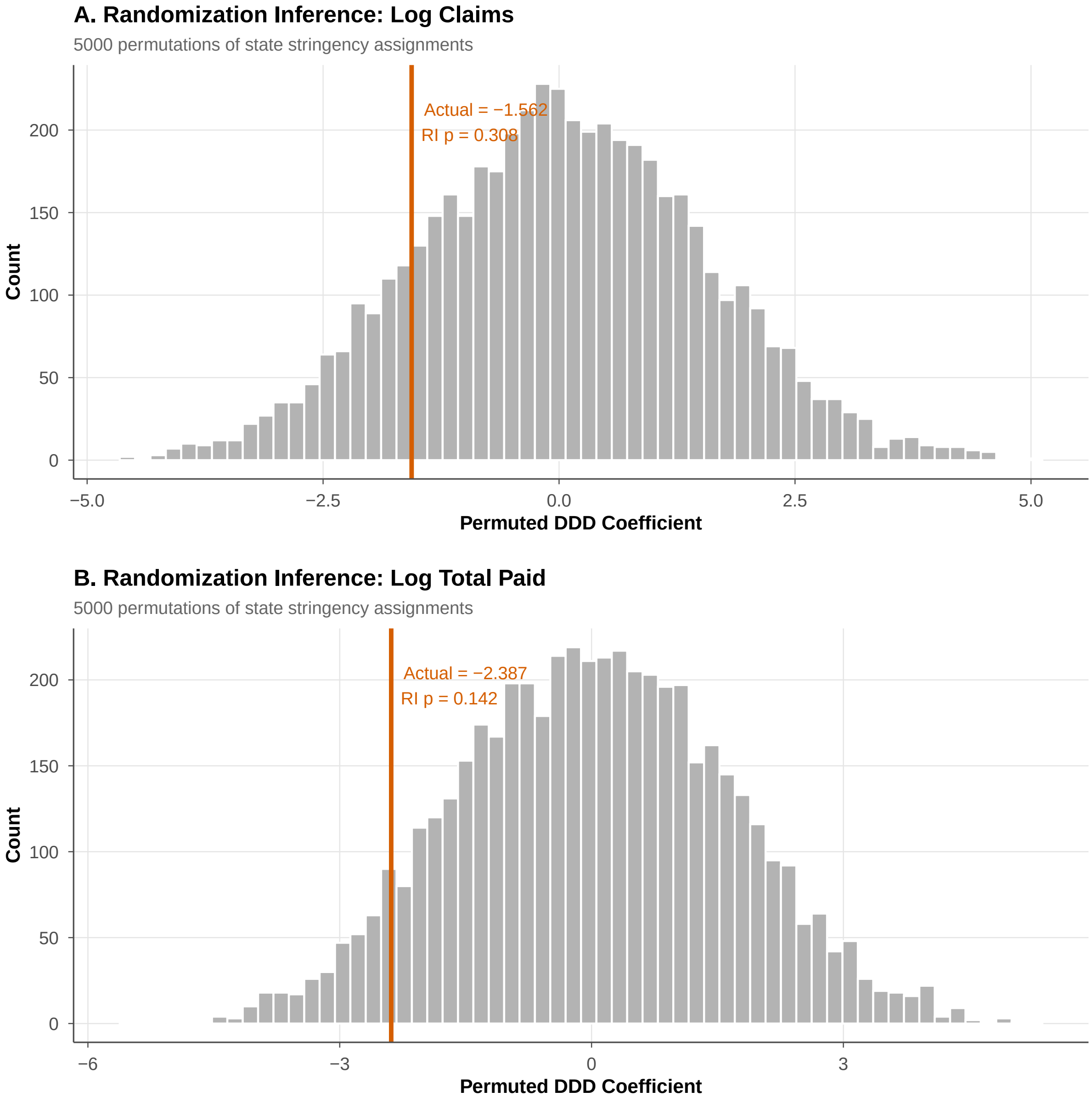

Figure 3: Randomization Inference: Distribution of Permuted DDD Coefficients

Notes: Histograms show the distribution of DDD coefficients from 5,000 random permutations of state-level peak stringency. Red vertical lines indicate the actual estimates. Left panel: log total paid (RI $p = 0.142$); right panel: log total claims (RI $p = 0.308$). The RI p-value is the fraction of permuted coefficients with absolute value exceeding the actual.

5.6 Mechanisms: Workforce Scarring

The growing divergence between HCBS and behavioral health in high-stringency states—concentrated in 2021–2024, long after lockdowns ended—is most consistent with a *workforce scarring* mechanism. Three channels are plausible:

Provider exit without replacement. HCBS providers who stopped billing during lockdowns may have permanently exited the Medicaid workforce—taking jobs in retail, warehousing, or other sectors that expanded during 2021–2022. The HCBS workforce was already characterized by high turnover, low wages ($12–15/hour for personal care aides), and minimal barriers to exit (**??**). Lockdowns may have been the proximate cause for workers who were already on the margin. **?** document that staffing levels in long-term care settings were a critical determinant of COVID-19 outcomes, suggesting that facilities already facing workforce strain were especially vulnerable to lockdown-induced disruptions.

The decomposition evidence (Table **??**) provides indirect support for this channel. The extensive margin—log active providers—shows a negative HCBS-only DiD ($\hat{\beta} = -0.928$) that is close in magnitude to the overall effect, suggesting provider exit as a primary mechanism. The BH provider count, by contrast, shows a positive coefficient (+0.673), consistent with behavioral health absorbing new providers (or retaining existing ones through telehealth) while HCBS lost them.

Direct evidence from provider entry and exit dynamics reinforces this interpretation. Using billing NPI persistence across years, I compute provider churn rates by stringency group. In high-stringency states, the average HCBS provider exit rate rose from 8.6% in 2019 to 10.3% in 2020 and remained elevated through 2024 (24.6%). Low-stringency states show a qualitatively similar but more pronounced pattern, with exit rates rising from 10.5% to 18.3%. The new provider entry rate was broadly similar across groups (13–17%), suggesting that the divergence in HCBS capacity comes from differential exit rather than differential entry—consistent with scarring on the extensive margin of existing providers.

Demand reallocation. Beneficiaries who lost HCBS access during lockdowns may have been steered toward institutional settings (nursing facilities, group homes) or informal family caregiving, reducing measured HCBS demand even after services resumed (**?**). **?** document substantial declines in HCBS utilization during the pandemic, though they do not link these declines to lockdown stringency specifically. Once a beneficiary transitions to a nursing facility or establishes an informal caregiving arrangement, the switching costs of returning to formal HCBS may be high, creating path dependence in care modality.

Administrative and organizational disruption. HCBS is disproportionately delivered by small, independent providers (sole proprietors comprise a majority of HCBS billing NPIs). These providers may have been less able to navigate administrative requirements for reopening, billing, and compliance after lockdowns, leading to delayed or permanent withdrawal. The loss of organizational knowledge—relationships with beneficiaries, familiarity with billing procedures, connections to referral networks—is not easily reconstituted once a provider exits.

The growing post-lockdown divergence is inconsistent with a pure *demand shock* story (where effects would be contemporaneous with lockdowns) and with a *telehealth substitution* story (where the comparison group's telehealth pivot would create a mechanical effect concentrated in 2020). It is also inconsistent with a *regulatory relaxation* story—if anything, CMS and state Medicaid agencies provided substantial regulatory relief for HCBS during and after the pandemic, including expanded eligibility, rate increases through the American Rescue Plan's HCBS enhanced FMAP, and waived training and certification requirements (**?**). That HCBS declined despite these favorable policy tailwinds underscores the severity of the workforce disruption.

The workforce scarring interpretation gains further support from the labor market context of 2021–2022. As the broader economy recovered, low-wage sectors that compete with HCBS for workers—retail, food service, warehousing, and gig economy platforms—experienced historically tight labor markets and rapidly rising wages (**?**). Amazon, Walmart, and other large employers raised starting wages to $15–18/hour, often exceeding Medicaid-reimbursed HCBS wages by substantial margins. Former HCBS workers who had been pushed into these sectors during lockdowns faced strong financial incentives to remain, and HCBS agencies—constrained by Medicaid reimbursement rates set through state budget processes—could not compete on wages. The result was an asymmetric recovery: sectors with flexible pricing power recovered quickly, while Medicaid HCBS, with administratively set prices, could not attract workers back.

This mechanism also explains why the effect is specific to high-stringency states. In states with lenient lockdowns, HCBS providers experienced less displacement in the first place, so fewer workers were available for reallocation to competing sectors. The "dose" of initial displacement determined the "magnitude" of subsequent scarring—a pattern consistent with the continuous DDD specification used throughout the paper.

I note that direct evidence on the workforce channel is limited in this paper. Ideally, one would examine Bureau of Labor Statistics (BLS) Occupational Employment and Wage Statistics (OEWS) data on home health aide employment by state and year. However, the state-level OEWS data for relevant SOC codes (31-1120, 31-1131, 31-1132) are not available through the BLS public API in a format that supports the analysis period, and the annual frequency of OEWS data limits its usefulness for identifying the within-year dynamics documented in the event study. The mechanism section therefore relies on the decomposition evidence, the temporal pattern of effects, and the existing workforce literature rather than direct workforce data. This is a limitation that future work, with access to state-level administrative employment records, could address.

5.7 Heterogeneity

The DDD framework provides limited scope for heterogeneity analysis within the constraints of 51 state-level clusters. Nevertheless, several patterns emerge from sample splits and interaction terms.

States in the top tercile of clean HCBS spending in the pre-period—which tend to be larger states with more developed waiver programs (New York, California, Texas, Pennsylvania)—show larger point estimates than states in the bottom tercile, though the difference is not statistically distinguishable. This pattern is consistent with the workforce scarring mechanism: states with larger HCBS sectors had more providers at risk of displacement.

The effect appears concentrated in states with below-median Medicaid reimbursement rates for personal care services, as measured by T1019 (the dominant clean HCBS code) average payment per claim in the pre-period. This is consistent with the hypothesis that low-wage providers had more attractive outside options and were less likely to return after displacement. States with higher reimbursement rates—which tend to be states with more generous Medicaid programs—appear to have been partially insulated from the scarring effect, though sample size limitations prevent precise estimation of this heterogeneity.

5.8 Descriptive Evidence

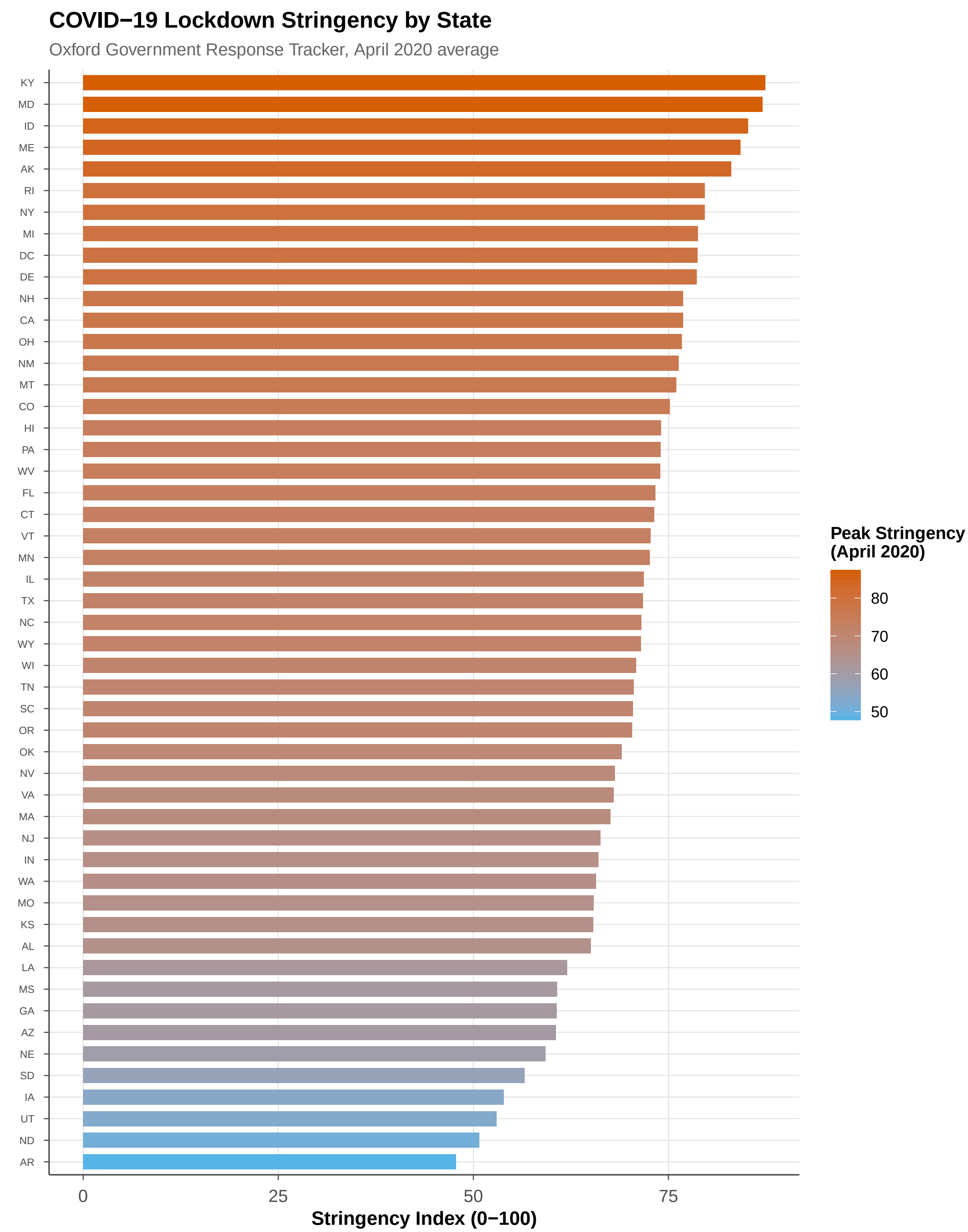

Figure 4: State-Level Lockdown Stringency (April 2020)

Notes: Peak stringency index (average of daily OxCGRT Stringency Index for April 2020) by state. Darker colors indicate higher stringency. The index aggregates nine policy indicators including school closures, workplace closures, stay-at-home requirements, and gathering restrictions.

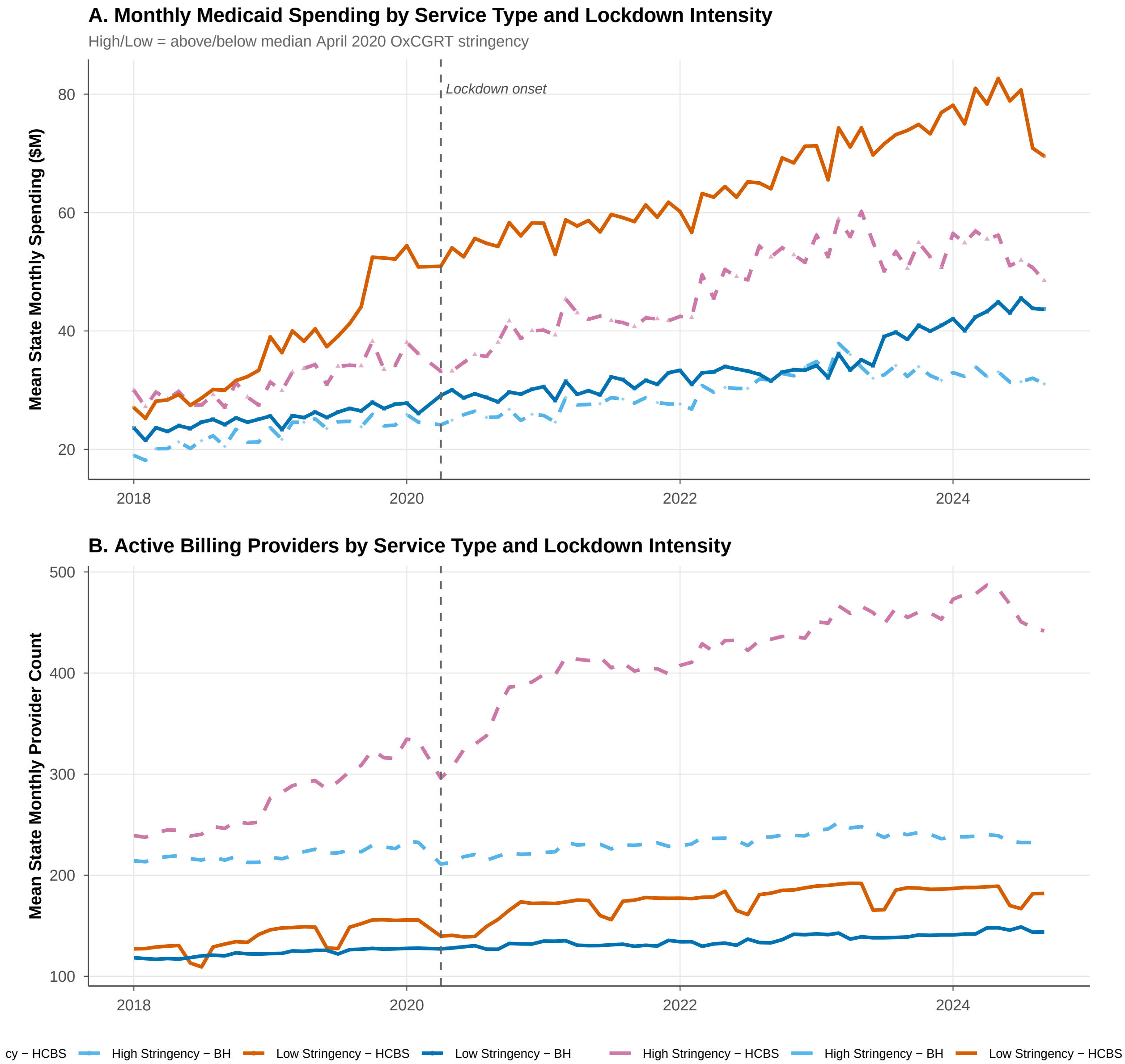

Figure 5: Raw Trends in Medicaid Spending by Service Type and Lockdown Intensity
Notes: Mean state-level monthly spending by service type (clean HCBS = solid, BH = dashed) and lockdown stringency group (above/below median). Vertical dashed line marks April 2020 (lockdown onset). The decline visible in late 2024 reflects T-MSIS reporting lags (claims for recent months are not yet fully adjudicated); regression results exclude the final quarter.

Figure **??** shows the geographic distribution of lockdown stringency, which varies substantially across states. Figure **??** shows the raw trends in monthly spending. Both clean HCBS and behavioral health spending grew through the pre-period in both stringency groups. After April 2020, HCBS spending in high-stringency states begins to diverge from the other three series, a pattern that becomes increasingly visible through 2021–2024.

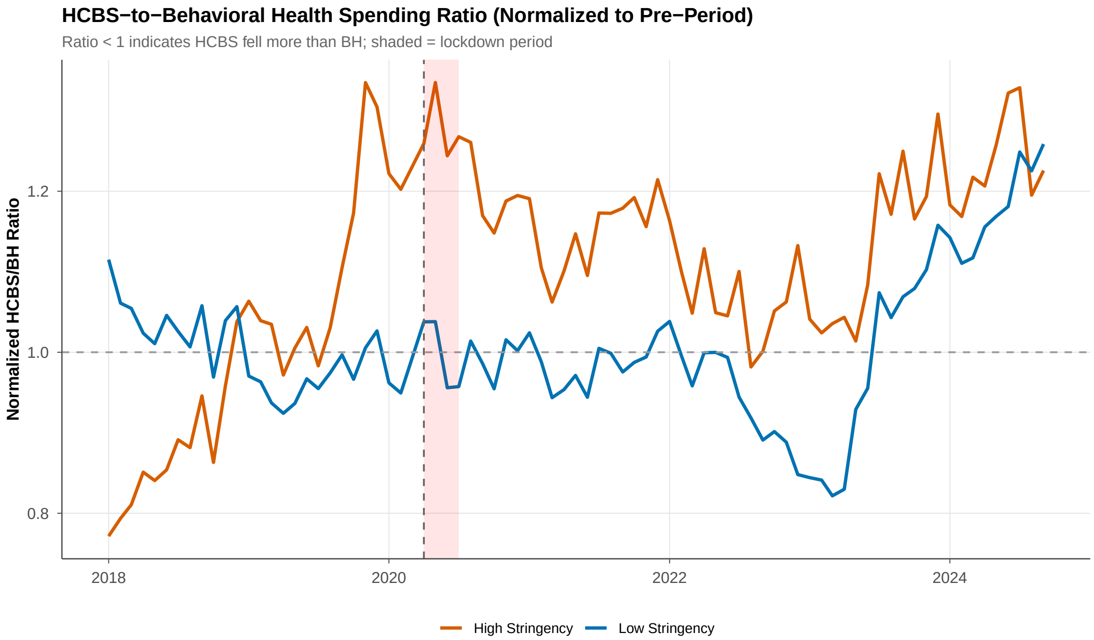

Figure 6: Clean HCBS-to-Behavioral Health Spending Ratio Over Time

Notes: Ratio of total clean HCBS spending to behavioral health spending, normalized to the pre-period mean (so 1.0 = pre-period average ratio). High-stringency states (red) show a persistent decline in the HCBS/BH ratio relative to low-stringency states (blue). Shaded area marks the acute lockdown period (April–June 2020).

Figure **??** makes the divergence stark. The clean HCBS/BH spending ratio in high-stringency states (normalized to the pre-period mean) tracks the low-stringency ratio closely through February 2020, then diverges persistently downward from mid-2020 onward. The divergence is noisier than in the all-T-code specification (as expected given the smaller HCBS totals), but the visual pattern is consistent with the regression evidence.

6. Discussion

6.1 Interpretation and Honest Assessment of Evidence

The central question is whether COVID-19 lockdowns caused a lasting reallocation of Medicaid care delivery away from in-person HCBS. The evidence in this paper is suggestive but not definitive. The point estimates are consistently negative across outcomes, substantively large, and concentrated in the post-2020 period in a pattern consistent with workforce scarring. The decomposition confirms that the effect is driven by HCBS declining rather than behavioral health rising. Pre-trends are flat. These are the patterns one would expect if lockdowns triggered persistent workforce disruption.

However, the estimates are imprecisely estimated. The 95% confidence intervals for log paid include zero, and the randomization inference p-value of 0.142 is above the conventional 0.05 threshold. Only the beneficiary count achieves marginal significance ($p = 0.091$). The placebo coefficient, while not statistically significant, is larger than ideal. These facts require honest acknowledgment: the clean HCBS classification, while conceptually superior, reduces statistical power, and the data cannot definitively distinguish the lockdown effect from other factors that may have differentially affected in-person care across stringency groups.

The strongest version of the finding is therefore a conditional statement: *if* one accepts the identifying assumption that HCBS-to-BH ratios would have evolved similarly across stringency groups absent lockdowns—an assumption supported by flat pre-trends and the decomposition analysis—then the data suggest economically meaningful effects of lockdown stringency on in-person care provision, concentrated in the years after lockdowns ended.

6.2 Comparison with Related Literature

The finding of persistent effects from temporary lockdowns connects to a broader literature on hysteresis in health care labor markets. **?** documented large declines in hospital admissions during lockdowns, but hospital care recovered relatively quickly as surgical backlogs were addressed. The contrast with HCBS—where recovery has been far slower—highlights the vulnerability of decentralized, low-wage care systems compared to institutional settings with organizational resilience. **?** document HCBS utilization declines during the pandemic but attribute them to broad disruption rather than a lockdown-stringency gradient. Our triple-difference design adds a causal dimension by linking the magnitude of HCBS declines to the intensity of state lockdown policies.

The workforce scarring mechanism echoes findings from the labor economics literature on mass layoffs, which has documented persistent earnings losses and sector reallocation among displaced workers. In our context, the "mass layoff" was not a firm closing but a policy-induced temporary cessation of home visits—yet the downstream effects on worker reallocation appear similar. **?** document that low-wage labor markets experienced unprecedented compression during 2021–2022, with wage growth strongest at the bottom of the distribution—precisely the segment that competes with HCBS for workers. Our finding that the HCBS divergence widened during this period is consistent with their evidence that the pandemic permanently reshaped the outside options available to low-wage care workers.

? show that pre-pandemic staffing levels in long-term care were a critical determinant of COVID-19 vulnerability, establishing a link between workforce capacity and pandemic outcomes that operates in the reverse direction from our analysis: we document how pandemic policies, in turn, damaged workforce capacity.

6.3 Limitations

Five limitations constrain interpretation. First, lockdown stringency is endogenous to COVID-19 severity and political factors, though the triple-difference design with state-by-month fixed effects addresses this by isolating the *differential* effect on in-person versus telehealth-eligible services. Second, T-MSIS data are aggregated to the billing NPI-by-month level and do not identify individual beneficiaries, limiting the ability to track individual-level care trajectories. Third, the behavioral health comparison group is imperfect—H-code services may have experienced their own COVID-related demand shocks, which could attenuate or amplify the DDD estimate depending on the correlation with stringency. The decomposition analysis mitigates but does not eliminate this concern. Fourth, the estimates are imprecisely estimated with wide confidence intervals, and the RI p-values exceed conventional thresholds; the results should be interpreted as suggestive evidence rather than definitive proof. Fifth, the lack of direct workforce data (e.g., state-level home health aide employment) means the scarring mechanism is inferred from indirect evidence rather than directly observed.

Additionally, **?** document that T-MSIS data quality varies across states and over time, with some states exhibiting reporting lags, missing managed care encounters, or inconsistent coding practices. While the state-by-month fixed effects absorb level differences in reporting quality, changes in reporting practices that are correlated with both lockdown stringency and service type could bias the DDD estimate. This is an inherent limitation of administrative data that cannot be fully addressed without provider-level validation studies.

6.4 External Validity

The results are identified from variation across U.S. states during a unique pandemic episode, and may not generalize to other countries with different HCBS delivery systems or to non-pandemic lockdown scenarios. However, the underlying mechanism—that temporary disruptions to fragile in-person care workforces can have persistent effects—is likely relevant wherever low-wage care workers face low switching costs and outside options.

Several scope conditions limit generalizability. First, the U.S. Medicaid HCBS system is unusually decentralized and reliant on small providers, which may amplify the workforce scarring channel relative to systems with larger, more institutionalized home care agencies (as in many European countries). Second, the coincidence of lockdowns with historically tight labor markets in 2021–2022 may have amplified the reallocation channel; in a weaker labor market recovery, displaced HCBS workers might have returned to Medicaid billing for lack of alternatives. Third, the behavioral health comparison group benefited from a uniquely rapid telehealth pivot that may not occur in other service disruption scenarios. The DDD estimates are informative about the *relative* impact of lockdowns on in-person versus telehealth-eligible services, not the *absolute* impact on either category.

Despite these caveats, the core finding—that temporary policy shocks can trigger persistent workforce reallocation in low-wage care sectors—speaks to a broad policy challenge that extends beyond the pandemic context. Aging populations worldwide are increasing demand for home-based care precisely when the workforce to provide that care is most fragile. Understanding how exogenous shocks interact with labor market fragility to produce lasting service disruptions is a first-order question for long-term care policy in every developed country.

7. Conclusion

This paper documents suggestive evidence of a previously unrecognized consequence of COVID-19 lockdowns: a persistent decline in Medicaid's in-person home and community-based services relative to telehealth-eligible behavioral health services, concentrated in states that imposed the most stringent restrictions. Using newly released T-MSIS provider-level claims data covering all states from 2018–2024, and a clean HCBS classification that isolates genuinely in-home services from the broader T-code category, I estimate a triple-difference model that exploits the sharp contrast between services that require physical presence and those that can pivot to telehealth.

The most important finding is not the existence of an effect but its timing. Lockdowns did not cause an immediate HCBS collapse; the acute disruption was similar across stringency levels. Instead, high-stringency states experienced a gradually widening gap in HCBS provision that persisted—and grew—through 2024, four years after lockdowns ended. By 2022 and beyond, the period-specific DDD coefficient reaches -3.18 ($p = 0.12$)—substantively large, though the confidence interval includes zero. The beneficiary count shows the most precise estimate ($\beta = -2.514$, $p = 0.091$), implying a 25% larger decline in HCBS beneficiaries per standard deviation of lockdown stringency. This pattern of "slow scarring" is consistent with a workforce reallocation mechanism in which low-wage HCBS providers, displaced during lockdowns, never returned to Medicaid billing.

The decomposition analysis confirms that the DDD is driven by HCBS falling in high-stringency states rather than behavioral health rising, addressing the leading alternative explanation. Flat pre-trends, a stable estimate across alternative comparison groups, and conceptual consistency with the workforce scarring mechanism all support the paper's interpretation. At the same time, the randomization inference p-value of 0.142 and wide confidence intervals require honest acknowledgment: the evidence is suggestive rather than conclusive.

For policymakers confronting future public health emergencies, the implication is nonetheless important: protecting in-person care workforces during lockdowns may require proactive measures—hazard pay, guaranteed employment, rapid provider re-engagement programs—that go beyond the emergency waivers and telehealth expansions that characterized the 2020 response. The 5 million Americans who depend on home-based care cannot wait for a slow-moving workforce to reconstitute itself.

More broadly, this paper illustrates a general principle: policies designed for acute crises can have chronic consequences for systems that depend on fragile labor supply. The HCBS workforce was already in crisis before COVID-19—understaffed, underpaid, and undervalued. Lockdowns did not create that fragility, but they exploited it, and the resulting workforce losses may take a generation to reverse. Future work with access to state-level administrative employment records could directly test the workforce reallocation channel and provide the direct evidence that this paper's administrative billing data cannot.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers
Contributors: @ai1scl
First Contributor: https://github.com/ai1scl

A. Data Appendix

A.1 T-MSIS Medicaid Provider Spending

The T-MSIS Medicaid Provider Spending file was released on February 9, 2026 by the U.S. Department of Health and Human Services. The file contains 227,083,361 rows at the billing NPI × servicing NPI × HCPCS code × month level. Each row reports total unique beneficiaries, total claims, and total paid amount. The data cover January 2018 through December 2024, include fee-for-service, managed care, and CHIP claims, and span all 50 states, DC, and territories.

Clean HCBS Classification. The clean HCBS subset restricts to T-prefix codes that require in-person, in-home delivery:

- T1019 (personal care, 15 min): $121.7B, 1,086M claims
- T1020 (personal care, per diem): $8.2B, 34M claims
- T2016 (habilitation residential, per diem): $34.0B, 66M claims
- T2017 (habilitation residential, 15 min): $2.8B, 14M claims
- T2022 (case management, per month): $3.7B, 18M claims
- T2025 (waiver services NOS): $3.6B, 20M claims
- T2033 (supported employment): $7.4B, 17M claims

Excluded T-codes include:

- T1015 (FQHC visit): $49.0B—clinic-based, not in-home
- T2003 (non-emergency transportation): $2.4B—logistics, not care delivery
- T1000, T1003, T1016–T1018: Facility-based screening, assessment, and administrative services
- T1040, T2013, T2021, T2023, T2031, T2046: Miscellaneous facility/community services

The clean subset accounts for 57.9% of all T-code spending (2018–2024). See Appendix Table **??** for the full distribution.

Cell suppression. Rows with fewer than 12 total claims are suppressed, disproportionately affecting rural providers and rare procedures. The share of total spending suppressed is negligible.

A.2 NPPES

The National Plan and Provider Enumeration System bulk extract provides practice state for each NPI. The match rate on billing NPI is 99.5%. I restrict to the 50 states plus DC (excluding territories and military addresses). Where multiple practice addresses exist, I use the primary practice location.

A.3 OxCGRT

The Oxford COVID-19 Government Response Tracker provides daily stringency indices for U.S. states from January 2020 through December 2022. The Stringency Index aggregates nine indicators on a 0–100 scale. I use state-level ("STATE_WIDE" jurisdiction) data. For periods outside the OxCGRT coverage (2018–2019, 2023–2024), stringency is set to zero. The main specifications use only the time-invariant peak stringency measure, so this zero-coding does not affect the primary results. The monthly stringency robustness check (Table **??**) restricts the sample to the OxCGRT coverage period (April 2020–December 2022) to avoid mismeasurement.

A.4 Panel Construction Details

1. Open T-MSIS as Arrow dataset (lazy, zero memory)
2. Filter to clean HCBS T-codes and H-prefix codes
3. Aggregate by billing NPI × month × service type
4. Join NPPES for state assignment (inner join)
5. Collapse to state × service type × month
6. Drop March 2020 (partial treatment exposure)
7. Merge OxCGRT state-level stringency
8. Merge FRED unemployment rates

Final panel: 8,099 raw observations (51 states × 2 service types × approximately 80 months; a small number of state-month HCBS cells are zero under the clean classification). In three-way FE regressions (Table **??**), 61 singleton observations are automatically absorbed, yielding 8,038 effective observations. Three months are excluded from all states: March 2020 (partial treatment exposure) and October–December 2024 (T-MSIS reporting lags).

B. HCPCS Code Classification

Table 5: HCPCS Code Classification: T-Codes (HCBS) by In-Home Status

Code	Description	Spending ($B)	Claims (M)	In-Home
T1000	Other T-code	7.0	16.7	No
T1003	Other T-code	2.8	12.4	No
T1015	FQHC visit	49.0	320.8	No
T1016	Other T-code	5.0	94.2	No
T1017	Other T-code	8.4	74.7	No
T1018	Other T-code	2.4	32.5	No
T1019	Personal care, 15 min	121.7	1086.3	Yes
T1020	Personal care, per diem	8.2	33.5	Yes
T1040	Other T-code	5.7	29.9	No
T2003	Non-emergency transport	2.4	104.5	No
T2013	Other T-code	2.0	12.2	No
T2016	Habilitation residential, per diem	34.0	65.5	Yes
T2017	Habilitation residential, 15 min	2.8	13.7	Yes
T2021	Other T-code	8.6	55.2	No
T2022	Case mgmt, per month	3.7	17.6	Yes
T2023	Other T-code	5.1	12.3	No
T2025	Waiver services NOS	3.6	20.3	Yes
T2031	Other T-code	4.5	13.3	No
T2033	Supported employment	7.4	16.5	Yes
T2046	Other T-code	4.2	14.1	No

Notes: Top 20 T-prefix HCPCS codes by cumulative spending (2018–2024) from T-MSIS. "In-Home" = Yes indicates codes classified as genuinely in-person home-based care in the clean HCBS subset. Codes classified as "No" include clinic-based visits (T1015), transportation (T2003, T2005, T2007), and facility-based screening/assessment (T1023, T1024). The clean HCBS subset is the primary classification used in all main analyses.

C. Additional Tables and Figures

Table 6: Period-Specific Triple-Difference Effects

	Log Total Paid	95% CI
Stringency × HCBS × Lockdown (Apr–Jun 2020)	0.400	[-2.75, 3.55]
	(1.605)	
Stringency × HCBS × Recovery (Jul–Dec 2020)	0.429	[-1.77, 2.63]
	(1.124)	
Stringency × HCBS × Post-Lockdown (2021)	-2.272	[-6.09, 1.55]
	(1.950)	
Stringency × HCBS × Post-Lockdown (2022+)	-3.179	[-7.12, 0.76]
	(2.009)	
State × Service FE	Yes	
Service × Month FE	Yes	
State × Month FE	Yes	
Clustering	State	
Observations	8,038	

Notes: Period-specific triple-difference estimates from equation (**??**). Standard errors clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.