Inside the Black Box of Medicaid: Provider-Level Spending Data
and a New Frontier for Health Economics Research

APEP Autonomous Research*   @SocialCatalystLab

February 15, 2026

Abstract

Medicaid spends over \$800 billion annually—one in five U.S. health care dollars—yet researchers have never observed *who delivers these services* at the provider level. We introduce the T-MSIS Medicaid Provider Spending dataset, released by HHS in February 2026: 227 million claim-month observations linking 617,503 billing providers to procedure codes and \$1.09 trillion in payments from January 2018 through December 2024. We document three facts. First, over half of total spending flows through Medicaid-specific procedure codes (T, H, S prefixes) with no Medicare equivalent—primarily personal care, behavioral health, and habilitation. Second, the provider panel is extraordinarily dynamic: only 6% of providers bill continuously across all 84 months. Third, joining to NPPES at a 99.5% match rate transforms anonymous identifiers into a geolocated, specialty-classified provider universe. We map the linkage architecture connecting T-MSIS to 30+ external datasets and outline analysis panels for studying provider supply, market structure, and workforce dynamics.

JEL Codes: I13, I18, H51, J44
Keywords: Medicaid, T-MSIS, provider spending, HCBS, health economics, data infrastructure

---

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

Medicaid is the largest health insurance program in the United States. It covers 92 million Americans—more than Medicare, more than any private insurer—and accounts for one in every five dollars spent on health care nationally (Medicaid and CHIP Payment and Access Commission, 2024). It is the dominant payer for long-term services and supports (Grabowski, 2006), behavioral health care, and births. And yet, for decades, the supply side of Medicaid has been essentially invisible to researchers.

The reason is data. Medicare's provider-level public use files have enabled a generation of health economics research on physician behavior, hospital markets, payment incentives, and quality (Clemens and Gottlieb, 2014; Dafny, 2005; Eliason et al., 2020). No comparable data existed for Medicaid. State-level T-MSIS claims were collected by CMS but never released publicly at the provider level. Researchers studying Medicaid providers had to rely on surveys (Decker, 2012; Polsky et al., 2015), state-specific administrative files obtained through individual data use agreements, or Medicare data as an imperfect proxy. The consequence was a vast blind spot: the economics profession knew remarkably little about who actually delivers Medicaid services, how the provider workforce is organized, or how spending varies across providers, services, and states.

On February 9, 2026, the Department of Health and Human Services released the Medicaid Provider Spending dataset through the HHS Open Data platform (Centers for Medicare & Medicaid Services, 2026). Derived from the Transformed Medicaid Statistical Information System (T-MSIS), the file contains 227 million rows mapping every billing provider to every procedure code to every month from January 2018 through December 2024—with total claims, unique beneficiaries, and Medicaid payments. It covers fee-for-service, managed care, and CHIP simultaneously. It is, in short, the provider-level Medicaid data that health economists have been waiting for.

We use these data to map the unobserved landscape of Medicaid providers, document how the file serves as a "universal joint" to 30+ external datasets through the National Provider Identifier (NPI), and construct the specific analysis panels needed to answer the program's most pressing questions.

What emerges from this exercise is a portrait of a health care system quite unlike the one visible through Medicare data. The Medicaid provider workforce is dominated not by physicians and hospitals, but by home health aides, personal care attendants, behavioral health workers, and community support organizations. Over half of total spending—and eight of the top ten procedure codes—flow through T-codes, H-codes, and S-codes that have no Medicare equivalent whatsoever. The standard tools of health economics, developed for

analyzing physician office visits and hospital admissions, must be adapted to a world of 15-minute personal care increments and per-diem habilitation services delivered in people's homes. The provider panel is remarkably dynamic: while only 6% of providers bill continuously across all 84 months, over 38% appear for fewer than 12 months, suggesting high workforce turnover or episodic program participation.

The research opportunities are correspondingly vast. With provider-level Medicaid claims linked to NPPES geography and Medicare billing data, researchers can study cross-payer substitution when Medicaid rates change (Clemens and Gottlieb, 2017), provider entry and exit in response to workforce policies, the organizational structure of home- and community-based services markets, and the effect of Medicaid managed care penetration on provider behavior—questions that were literally unanswerable before this data release. The 2018–2024 window captures three major natural experiments: the COVID-19 pandemic, the ARPA HCBS spending increase (April 2021), and the Medicaid unwinding (beginning April 2023), each with staggered state-level implementation suitable for difference-in-differences designs (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024).

2. The Dataset

2.1 Structure and Schema

The T-MSIS Medicaid Provider Spending dataset has a clean, minimal schema: seven columns defining a provider–procedure–month panel. Each row represents a unique combination of billing provider NPI, servicing provider NPI, HCPCS procedure code, and claim month, with three outcome measures: total unique beneficiaries served, total claims submitted, and total Medicaid amount paid. Table 1 presents the dataset dimensions.

Table 1: T-MSIS Medicaid Provider Spending Dataset Overview

Characteristic	Value
Total rows (billing × servicing × HCPCS × month)	227,083,361
Date range	January 2018 to December 2024
Months covered	84
Unique billing NPIs	617,503
Unique HCPCS codes	10,881
Total claims	18,825,564,012
Total Medicaid paid	$1,093,562,833,512
Median monthly payment per provider	$2,362
Mean monthly payment per provider	$27,417

Notes: T-MSIS = Transformed Medicaid Statistical Information System. Released by HHS February 9, 2026. Covers fee-for-service, managed care, and CHIP claims. Cell suppression applies when beneficiary count < 12. Monthly payment per provider = provider's total payments divided by active months. The 617,503 unique billing NPIs is the cumulative count over the full 84-month panel; at any given monthly peak, roughly 270,000 are active (see Table 4).

The unit of observation—billing NPI × servicing NPI × HCPCS code × month—is granular enough for provider-level analysis yet aggregated enough to protect beneficiary privacy. A minimum cell size of 12 beneficiaries is enforced through suppression, meaning that low-volume provider–procedure combinations are excluded from the data entirely. This suppression is not random: it disproportionately affects rural providers, rare procedures, and small organizations, a limitation we discuss further in the appendix. Because each suppressed cell contains fewer than 12 beneficiaries, the spending share removed is negligible even where the row count is substantial.

The presence of *both* billing and servicing NPI is a structural feature worth emphasizing. In 65% of rows, the billing and servicing NPI differ, encoding an organizational relationship: a Type 2 organization (the billing NPI) submits claims for services performed by an individual practitioner (the servicing NPI). In 31% of rows, they are identical—the provider bills for services they personally delivered. The remaining 4% have no servicing NPI recorded. But spending shares are reversed: self-billing rows account for 61% of total payments, while the numerous organizational rows account for 29%—organizations submit many small encounter records, while individual providers submit fewer but larger claims. This billing–servicing

structure mirrors the Medicare PECOS reassignment framework and enables the construction of provider–organization linkages *within Medicaid* without requiring external enrollment data.

2.2 Coverage and Scope

The data span 84 months: January 2018 through December 2024, though December 2024 is incomplete due to standard claims processing lags. Coverage is national, encompassing all 50 states, the District of Columbia, and U.S. territories. Crucially, the file commingles fee-for-service and managed care claims without a payer indicator—a design choice that captures total Medicaid spending regardless of delivery system but prevents within-data FFS/MCO comparisons.

A clarification on temporal coverage is warranted because different analyses in this paper use different subsets of the 84-month panel. The raw dataset contains all 84 months (January 2018 through December 2024). Provider tenure statistics—including the panel balance distribution in Figure 4 and the tenure groups in Table 5—count active months across all 84 months, since a provider submitting a claim in December 2024 is genuinely active regardless of whether that month's spending totals are complete. Time-series analyses of monthly spending and claims (Figure 2, Table 4, Figure 3), by contrast, exclude December 2024 because claims processing lags make that month's aggregate totals unreliable, yielding an 83-month analysis window for those exhibits.

The CHIP (Children's Health Insurance Program) is also included, consistent with T-MSIS reporting requirements. Since 84.8% of Medicaid enrollees nationally are in managed care plans (Medicaid and CHIP Payment and Access Commission, 2024), the data overwhelmingly reflect managed care encounters rather than traditional fee-for-service claims, a point that matters for interpreting payment amounts.

The institutional backstory matters for understanding what this release represents. The Affordable Care Act of 2010 mandated the creation of T-MSIS to replace the legacy Medicaid Statistical Information System (MSIS), which had operated since 1999 with inconsistent state-level reporting standards and limited analytic utility. CMS spent over a decade developing data quality standards, field definitions, and validation rules. States were required to transition from MSIS to T-MSIS by October 2017, but compliance was uneven: some states submitted production-quality files immediately, others required years of remediation. CMS publishes the Data Quality Atlas, which tracks state-by-state data quality metrics across dozens of dimensions—claim completeness, eligibility file consistency, provider identifier validity—and documents the dramatic improvement in usability from 2017 through 2024.

Before the February 2026 release, T-MSIS data were available to researchers only in two forms: state-level aggregates published by MACPAC and Medicaid.gov (useful for

national trend analysis but useless for provider-level research), and the restricted-use Research Identifiable Files available through the CMS Virtual Research Data Center under a data use agreement (accessible to a small number of institutional researchers at substantial cost and delay). The public release of provider-level T-MSIS data is unprecedented. For the first time, any researcher with an internet connection can download provider–procedure–month claims data for the entire Medicaid program. The analytic implications are comparable to the initial release of the Medicare Physician/Supplier PUF in 2014, which catalyzed an entire subfield of Medicare provider research.

The aggregate spending totals in the T-MSIS provider file are broadly consistent with independent expenditure benchmarks. MACPAC reports total Medicaid benefit spending of approximately $734 billion in federal fiscal year 2023 (Medicaid and CHIP Payment and Access Commission, 2024), while CMS-64 state-level expenditure data provide a complementary accounting. Exact reconciliation between T-MSIS provider payments and these benchmarks is complicated by several factors: T-MSIS captures provider-level payments, while MACPAC and CMS-64 totals include administrative costs, disproportionate share hospital (DSH) payments, and supplemental payments that are not attributed to specific providers or procedure codes. Additionally, the cell suppression protocol removes low-volume cells from the public file, and managed care encounter valuations may not sum to capitation payments at the plan level. Nevertheless, the order-of-magnitude correspondence between T-MSIS provider spending and established expenditure benchmarks provides confidence in the data's coverage and internal consistency.

2.3 What the Data Lack

Understanding the omissions is as important as understanding the contents. The dataset contains *no state identifier*, no provider name, no specialty classification, no diagnosis codes, no beneficiary demographics, and no managed care plan indicator. The NPI is the sole link to the outside world—every contextual variable must be obtained through external joins. This sparsity is by design: it enables public release while preserving confidentiality. But it means the raw file, without enrichment, supports only temporal and procedure-level analysis. The research value emerges from the linkage architecture described in Section 4.

A further limitation concerns the interpretation of payment amounts in managed care contexts. For managed care encounters, the "Medicaid Amount Paid" field may represent an imputed or allowed amount rather than actual provider revenue, since MCO encounter records are often valued at fee-schedule equivalents rather than reflecting negotiated capitation rates. Because managed care encounter valuation conventions vary across states, cross-state comparisons of dollar-denominated outcomes should account for this variation, and researchers

should exercise particular caution when interpreting payment levels—as opposed to within-state changes over time—as measures of provider compensation.

3. A Descriptive Portrait

What does the Medicaid provider landscape look like when observed for the first time at scale? Three features dominate: the overwhelming concentration of spending in Medicaid-specific service categories invisible to Medicare researchers, the extraordinary dynamism of the provider panel, and dramatic geographic variation revealed through NPI-based linkage.

3.1 A System Built on Home- and Community-Based Services

Table 2 decomposes total spending by HCPCS code category. The result is striking: over half of total spending—52%—flows through T-codes (home- and community-based services, 29%), H-codes (behavioral health, 14%), and S-codes (temporary state-specific services, 9%). These are Medicaid-specific procedure codes with no Medicare equivalent—they do not appear in the Medicare Physician/Supplier PUF or any other Medicare public use file.

Table 2: Medicaid Spending by Service Category

Category	Spending ($B)	% Total	Claims (B)	Codes
CPT Professional Services	367.6	33.6	10.3	5,731
HCBS/State (T-codes)	321.1	29.4	2.4	154
Behavioral Health (H-codes)	148.6	13.6	1.3	145
Temporary/State (S-codes)	95.7	8.8	1.2	299
Other	70.4	6.4	1.8	1,775
Ambulance/DME (A-codes)	40.3	3.7	0.9	584
CMS Procedures (G-codes)	27.7	2.5	0.5	832
Drugs (J-codes)	10.8	1.0	0.4	567
DME (E-codes)	9.5	0.9	0.1	410
Orthotics/Prosthetics (L-codes)	1.9	0.2	0.0	384
Total	1093.6	100.0	18.8	10,881

Notes: Categories defined by HCPCS code prefix. T-codes = HCBS/state-specific services; H-codes = behavioral health; S-codes = temporary/state codes; numeric = CPT professional services. Most T, H, and S codes have no Medicare equivalent.

The single largest procedure code, T1019 (personal care services, billed per 15-minute

increment), accounts for $123 billion—more than one-fifth of all Medicaid-specific code spending and 11.2% of total program payments (Table 3). The top six codes by spending are all Medicaid-specific: personal care, FQHC clinic visits, residential habilitation, attendant care, and community support services. Traditional physician office visits (CPT 99213, 99214) rank fourth and sixth overall, underscoring that even the familiar CPT universe looks different through Medicaid data.

Table 3: Top 15 HCPCS Codes by Total Medicaid Spending, 2018–2024

Rank	Code	Description	Total ($B)	Claims (M)	Providers	%
1	T1019	Personal care per 15 min	122.7	1100.6	9,780	11.2
2	T1015	FQHC clinic visit	49.2	322.2	13,829	4.5
3	T2016	Habilitation residential per diem	34.9	69.0	1,761	3.2
4	99213	Office visit, established	33.0	764.3	164,075	3.0
5	S5125	Attendant care per 15 min	31.3	398.1	4,555	2.9
6	99214	Office visit, established (complex)	29.9	502.5	150,306	2.7
7	99284	ED visit, level 4	20.2	170.7	21,452	1.8
8	H2016	Community support per diem	19.7	60.7	2,115	1.8
9	99283	ED visit, level 3	16.9	157.9	20,157	1.5
10	H2015	Community support per 15 min	16.5	109.9	3,908	1.5
11	99285	ED visit, level 5	15.1	111.0	17,705	1.4
12	90837	Psychotherapy, 60 min	12.1	127.2	44,034	1.1
13	S5102	Day care, adult per diem	9.3	111.1	2,146	0.9
14	90834	Psychotherapy, 45 min	8.8	108.8	27,908	0.8
15	T2021	Day habilitation per diem	8.7	55.6	2,363	0.8
		Top 15 total	428.3	4169.6		39.2

Notes: HCPCS = Healthcare Common Procedure Coding System. T-codes are state/HCBS specific, H-codes are behavioral health, S-codes are temporary/state codes. Most T, H, and S codes have no Medicare equivalent. Cumulative totals 2018–2024. % = share of total cumulative Medicaid spending ($1.09 trillion).

This composition has profound implications for research design. The standard approach in health economics—analyzing physician office visits, surgical procedures, and hospital encounters—captures a small fraction of Medicaid spending. The bulk of Medicaid resources flow to home health aides assisting with activities of daily living, residential programs for individuals with intellectual and developmental disabilities, and community-based behavioral health workers providing psychosocial rehabilitation. Understanding this workforce, its organization, and its response to policy is the core research opportunity.

Figure 1 displays the composition of spending over time. The category shares are remarkably stable, suggesting that the dominance of HCBS is structural rather than driven by pandemic-era policy changes.
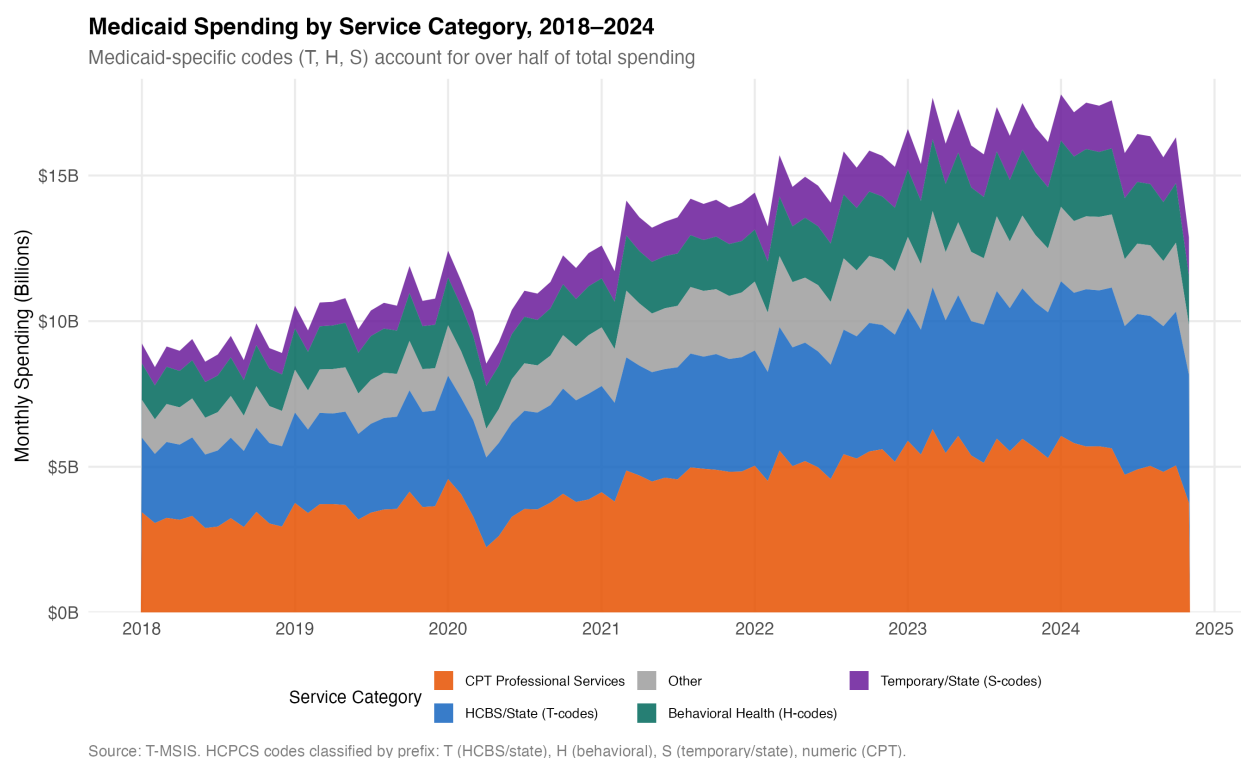
Medicaid Spending by Service Category, 2018–2024

Medicaid-specific codes (T, H, S) account for over half of total spending

Source: T-MSIS. HCPCS codes classified by prefix: T (HCBS/state), H (behavioral), S (temporary/state), numeric (CPT).

Figure 1: Medicaid Spending by Service Category, 2018–2024

Notes: Monthly spending decomposed by HCPCS code prefix. T-codes = HCBS/state services; H-codes = behavioral health; S-codes = temporary/state codes; numeric = CPT professional services. Data from T-MSIS.

3.2 Growth Trajectory

Figure 2 presents total monthly Medicaid provider spending from 2018 through 2024, with key policy events annotated. Total spending roughly doubled over this period. The COVID-19 pandemic produced a sharp but transient dip in claims during spring 2020 (Figure 3),

consistent with reduced service utilization during lockdowns: claims fell approximately 45% before recovering by summer 2020 and continuing their upward trajectory.
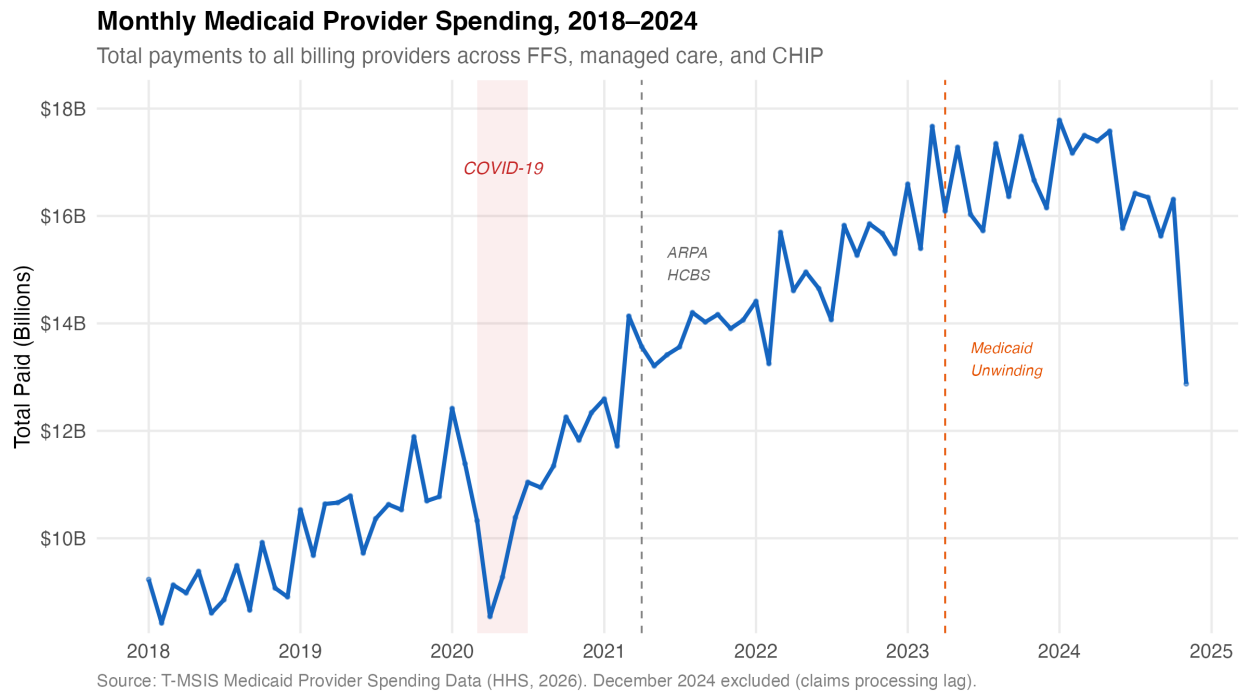
Monthly Medicaid Provider Spending, 2018–2024

Total payments to all billing providers across FFS, managed care, and CHIP

Source: T-MSIS Medicaid Provider Spending Data (HHS, 2026). December 2024 excluded (claims processing lag).

Figure 2: Monthly Medicaid Provider Spending, 2018–2024

Notes: Total payments to all billing providers. Vertical dashed lines indicate the ARPA HCBS spending increase (April 2021) and the beginning of Medicaid unwinding (April 2023). Shaded region = COVID-19 initial disruption. December 2024 excluded due to claims processing lag.

Table 4: Annual Growth in Medicaid Provider Spending

Year	Spending ($B)	% Δ	Claims (M)	% Δ	Providers	Codes
2018	108.7	N/A	2134.4	N/A	230,343	6,253
2019	126.9	16.8	2399.5	12.4	243,432	6,467
2020	132.1	4.1	2304.2	-4.0	242,759	6,507
2021	162.6	23.1	2904.6	26.1	263,709	6,815
2022	179.6	10.5	3087.7	6.3	266,470	6,753
2023	198.8	10.7	3218.6	4.2	271,651	6,889
2024	180.8	N/A	2713.2	N/A	256,264	6,678

Notes: Annual totals from T-MSIS. Providers = peak monthly unique billing NPIs (the cumulative total across all 84 months is 617,503; see Table 1). 2024 covers January–November only (December excluded due to claims processing lag); as a result, the 2024 row understates true annual totals and growth rates are suppressed. The sum of annual spending in this table ($1,089.5B) is approximately $4B less than the grand total in Table 1 because December 2024 is excluded here. % Δ = year-over-year growth rate.

The ARPA HCBS spending increase (April 2021) and the beginning of Medicaid unwinding (April 2023) are both visible in the spending trajectory, though disentangling these from secular trends requires the state-level variation exploitable through NPPES linkage.
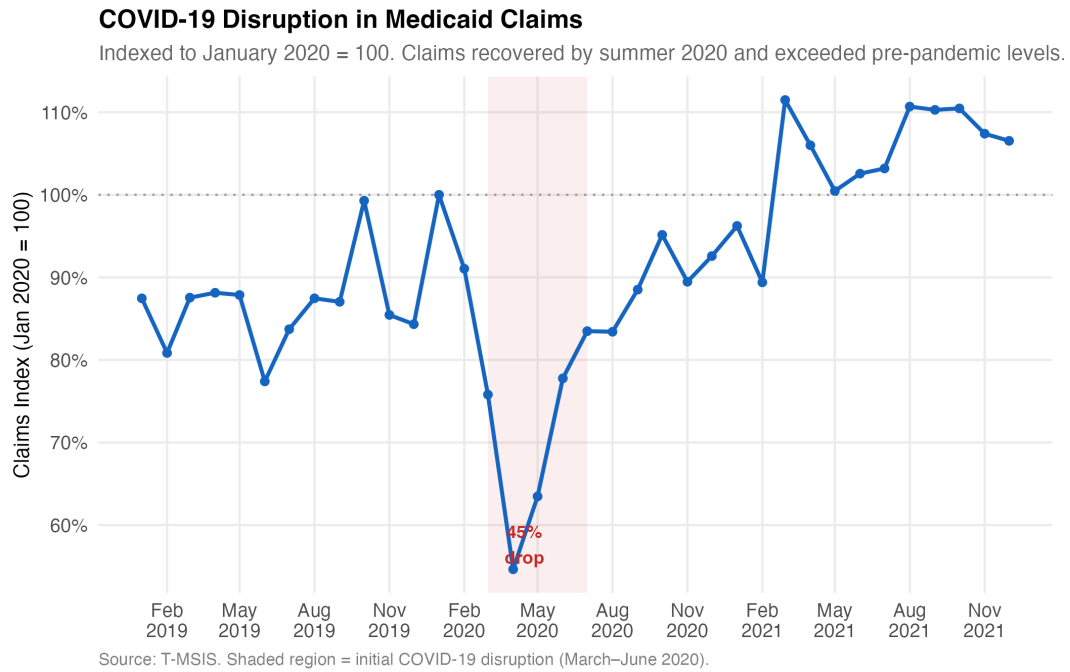
COVID-19 Disruption in Medicaid Claims

Indexed to January 2020 = 100. Claims recovered by summer 2020 and exceeded pre-pandemic levels.

Source: T-MSIS. Shaded region = initial COVID-19 disruption (March–June 2020).

Figure 3: COVID-19 Claims Disruption and Recovery

Notes: Claims indexed to January 2020 = 100. The initial disruption reduced claims by approximately 45%, with full recovery by July 2020.

3.3 A Dynamic Provider Workforce

The provider panel is strikingly unbalanced. Figure 4 shows the distribution of provider tenure—the number of months in which a billing NPI submits at least one claim. Only 6% of providers are active in all 84 months. Over 38% appear for fewer than 12 months. The median provider is active for just 22 months.

Distribution of Provider Tenure in T-MSIS Panel

Highly unbalanced: most providers appear for fewer than 12 months
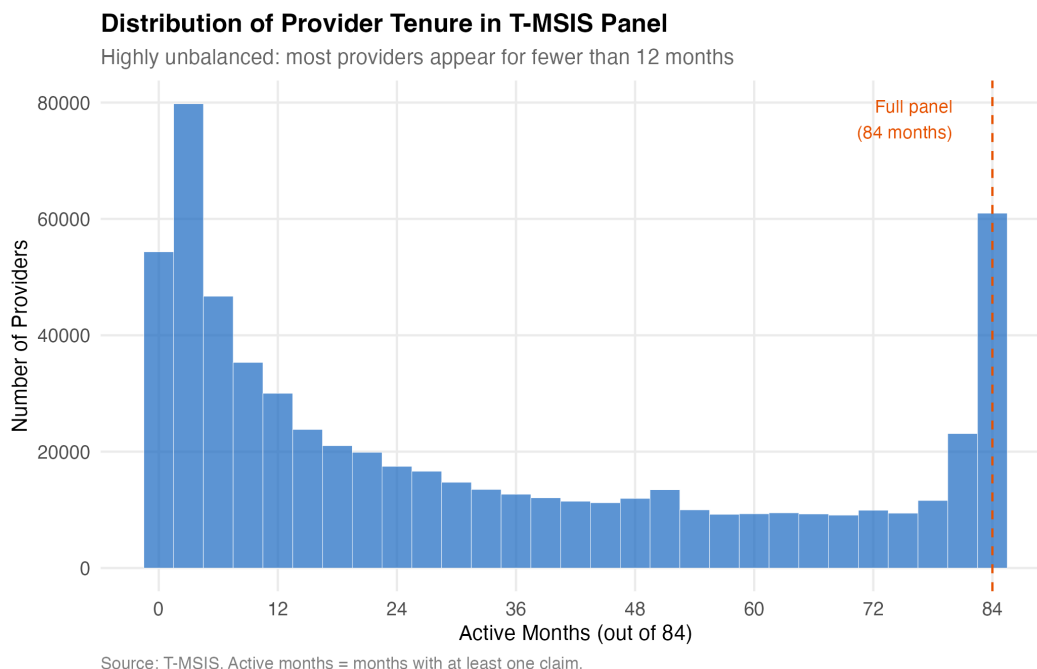
Figure 4: Distribution of Provider Tenure in the T-MSIS Panel

Notes: Active months = number of months with at least one billing claim. Panel covers January 2018 through December 2024 (84 months). Provider tenure counts include December 2024 billing activity; monthly spending totals exclude December 2024 due to claims processing lag. The heavy left tail reflects high workforce turnover in home- and community-based services.

Table 5: Provider Panel Properties: Tenure and Spending Concentration

Tenure	Providers	% of Providers	% of Spending	Median Total ($)
1 months	54,334	8.8	0.0	605
2–6 months	112,926	18.3	0.2	3,038
7–12 months	70,018	11.3	0.5	13,144
13–24 months	85,606	13.9	2.0	35,509
25–36 months	59,097	9.6	2.5	79,004
37–48 months	46,567	7.5	3.6	136,730
49–60 months	43,123	7.0	5.6	252,452
61–72 months	37,561	6.1	7.5	344,917
73–84 months	108,271	17.5	78.1	1,151,105

Notes: Panel covers January 2018 through December 2024 (84 months). Provider tenure counts include December 2024 billing activity; monthly spending totals exclude December 2024 due to claims processing lag. Tenure = months with at least one billing claim. Median total = median of cumulative payments across all months active.

This churn is consistent with what is known about the HCBS workforce: direct care workers experience turnover rates exceeding 40% annually in many states, driven by low wages, physically demanding work, and limited career advancement. But the T-MSIS data reveal for the first time the *scale* of this dynamism from the billing side. Figure 5 decomposes monthly provider counts into entry and exit flows, showing persistent net entry over the period.
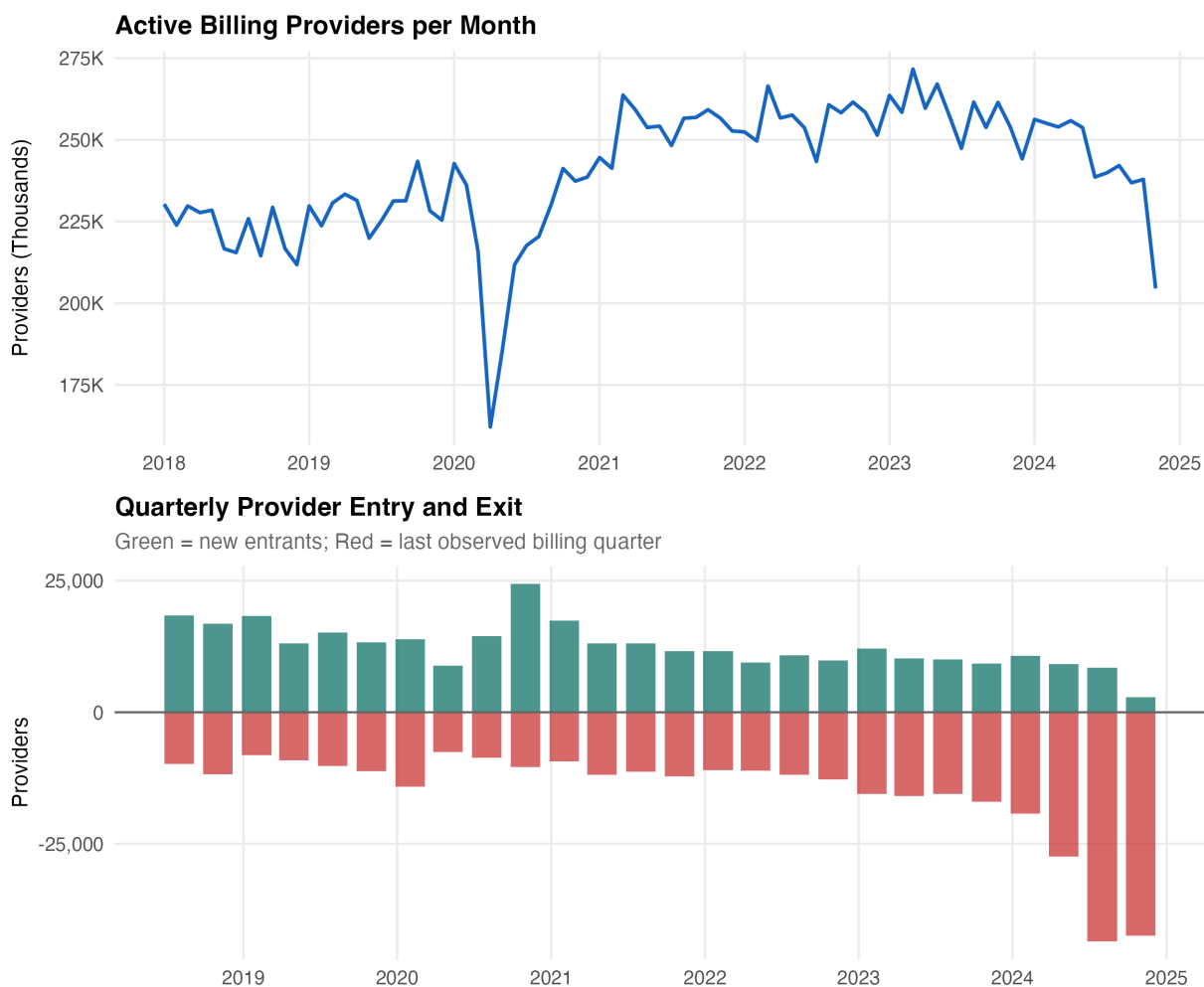
The Medicaid Provider Workforce, 2018–2024

Active Billing Providers per Month

Quarterly Provider Entry and Exit

Green = new entrants; Red = last observed billing quarter

Source: T-MSIS. Provider = unique billing NPI. Entry = first billing month. Exit = last billing month before data end.

Figure 5: Provider Entry, Exit, and Active Count, 2018–2024

Notes: Top panel: unique billing NPIs with at least one claim per month. Bottom panel (quarterly): green bars = providers billing for the first time; red bars = providers billing for the last time.

A few large organizations claim the lion's share of Medicaid dollars (Figure 6). Large HCBS agencies, behavioral health networks, and FQHCs account for a disproportionate share of total spending. The top 1% of providers by cumulative payments account for over half of all spending. This concentration has implications for market structure analysis: Medicaid provider markets may be substantially more concentrated than they appear from simple provider counts.

Distribution of Total Medicaid Payments by Provider, 2018–2024

Extreme right skew: a small number of providers account for most spending

Source: T-MSIS. Each bar represents a bin of providers grouped by total cumulative payments.
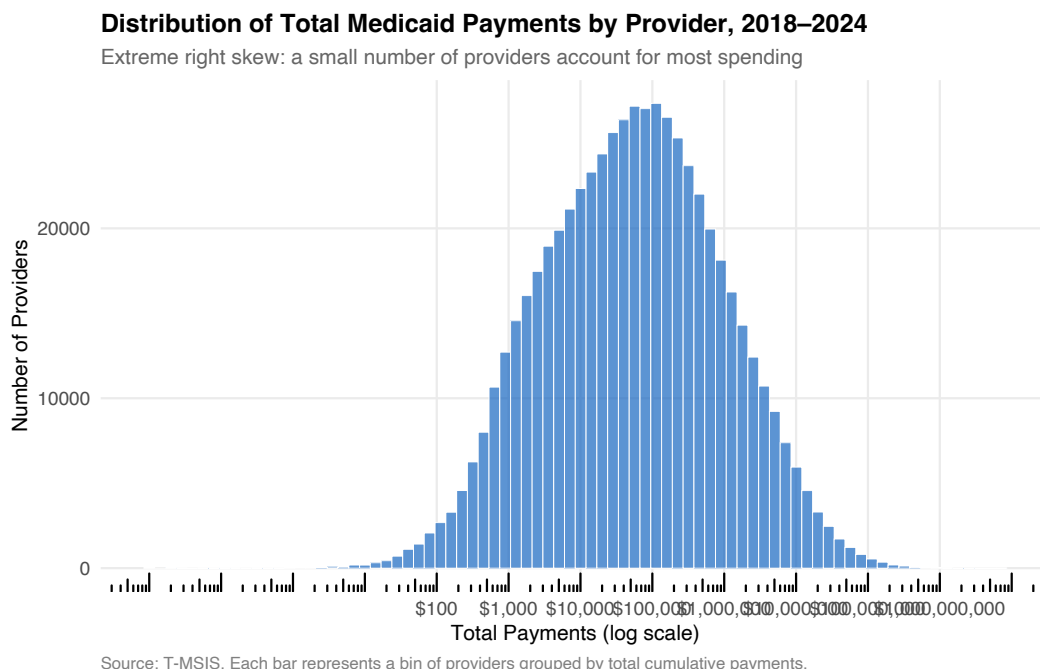
Figure 6: Distribution of Cumulative Provider Payments (Log Scale)

Notes: Each provider's total Medicaid payments summed across all months and HCPCS codes, 2018–2024. Displayed on log scale. The right tail represents large organizational billing providers.

3.4 Organizational Structure

The billing–servicing NPI relationship reveals the organizational backbone of Medicaid service delivery. Table 6 decomposes claims by billing structure. The majority of spending flows through self-billing providers (billing NPI = servicing NPI), suggesting individual practitioners or small organizations that bill directly. But a significant share flows through organizational structures where a Type 2 organization bills on behalf of individual practitioners, and a further share has no servicing NPI recorded at all.

Table 6: Billing Structure: Organizational Relationships in T-MSIS

Structure	Rows	% Rows	Spending ($B)	% Spending
Self-billing (billing = servicing)	69,731,487	30.7	671.1	61.4
Organization billing (billing != servicing)	147,861,529	65.1	320.1	29.3
Solo (no servicing NPI)	9,490,345	4.2	102.4	9.3

Notes: Self-billing = billing NPI equals servicing NPI (provider bills for own services). Solo = no servicing NPI recorded (typically Type 2 organizations). Organization billing = billing NPI differs from servicing NPI (org bills for individual practitioners).

These patterns can be enriched substantially through NPPES, which provides entity type (individual vs. organization), parent organization name and TIN, and sole proprietor status. The combination of billing structure from T-MSIS and organizational identifiers from NPPES enables the construction of firm-level panels: grouping NPIs by parent organization TIN, measuring firm size (NPI count), and tracking firm-level entry, exit, and growth over time.

3.5 Geographic Variation

Joining T-MSIS to NPPES practice location state and normalizing by ACS population reveals dramatic geographic variation in Medicaid provider spending. Figure 7 shows cumulative per-capita spending by state from 2018 through 2024. The variation reflects differences in Medicaid eligibility rules (Currie and Gruber, 1996; Buchmueller et al., 2016), HCBS waiver generosity, managed care penetration rates, and state-specific billing patterns.

Cumulative Medicaid Provider Spending per Capita, 2018–2024
Geographic variation reflects differences in Medicaid generosity, HCBS use, and managed care
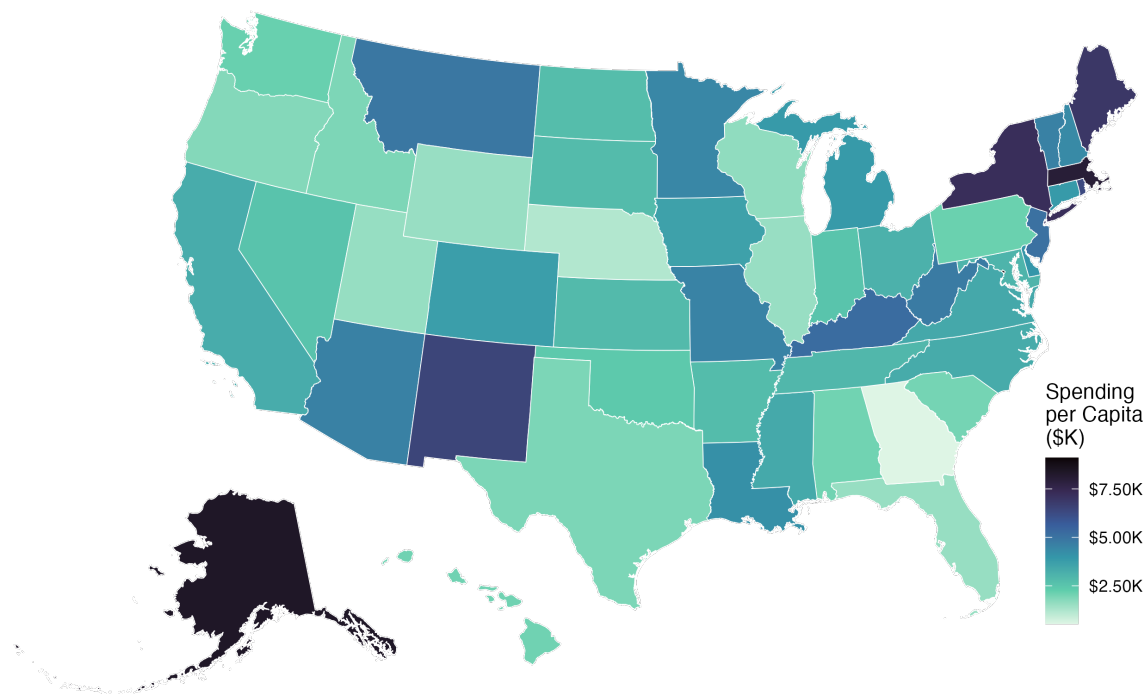
Source: T-MSIS joined to NPPES (state). Per capita using ACS 2022 total population.

Figure 7: Cumulative Medicaid Provider Spending per Capita by State, 2018–2024
Notes: Provider state assigned via NPPES practice location. Per-capita denominator from ACS 2022 total population. Includes all HCPCS categories.

3.6 A Comparison with Medicare

The most natural benchmark for the T-MSIS data is the Medicare Physician/Supplier PUF, which has the same NPI × HCPCS structure and has been publicly available since 2014. Placing the two datasets side by side reveals not minor differences in emphasis but two fundamentally different health care systems.

Start with procedure codes. The top ten Medicare codes by spending are almost exclusively CPT: evaluation and management visits (99213, 99214, 99215), imaging (radiology reads, echocardiograms), and procedures (cataract surgery, colonoscopy, joint replacement). These are physician-delivered clinical services performed in offices, hospitals, and ambulatory surgery centers. In T-MSIS, by contrast, eight of the top ten codes are Medicaid-specific T, H, and S prefixes: personal care per 15-minute increment, residential habilitation per diem, attendant care, community psychiatric support (Table 3). The overlap is minimal. A researcher who studied only Medicare procedure codes would miss over half of Medicaid spending entirely.

Provider composition diverges just as sharply. The Medicare PUF is dominated by physician specialties—internal medicine, cardiology, ophthalmology, orthopedic surgery, radiology—with hospitals and ambulatory surgery centers as the primary organizational providers. The Medicaid provider workforce visible in T-MSIS looks nothing like this. The modal billing provider is a home health agency, a personal care attendant organization, a behavioral health clinic, or a Federally Qualified Health Center. Individual providers are more likely to be licensed clinical social workers, certified nursing assistants, or applied behavior analysis technicians than board-certified physicians. The NUCC taxonomy codes that dominate Medicaid billing—home health aide (374U00000X), community health worker (171M00000X), residential treatment facility (324500000X)—barely appear in Medicare data.

The billing structure also differs markedly. Medicare claims flow overwhelmingly through individual provider billing: a physician renders a service and submits a claim under their own NPI. In the T-MSIS data, 65% of claim rows involve organizational billing—a Type 2 entity submitting claims on behalf of individual servicing providers. This reflects the organizational backbone of HCBS: agencies employ or contract with direct care workers and bill Medicaid on their behalf. The agency is the economic actor; the individual worker is the service delivery unit.

These differences carry a simple implication. Medicaid is not "Medicare for poor people," and analysis tools developed for Medicare will not transfer without substantial adaptation. The relevant unit of service is a 15-minute personal care increment, not an office visit. The relevant market participants are home health agencies and behavioral health organizations, not physician practices and hospitals. The relevant workforce questions concern direct care worker turnover and wage adequacy, not physician specialty choice and referral patterns. The

T-MSIS data make this alternative health care system visible for the first time at national scale.

4. The Linkage Universe

The raw T-MSIS file contains seven columns. Through the NPI, it connects to hundreds. Figure 8 illustrates the complete linkage architecture.

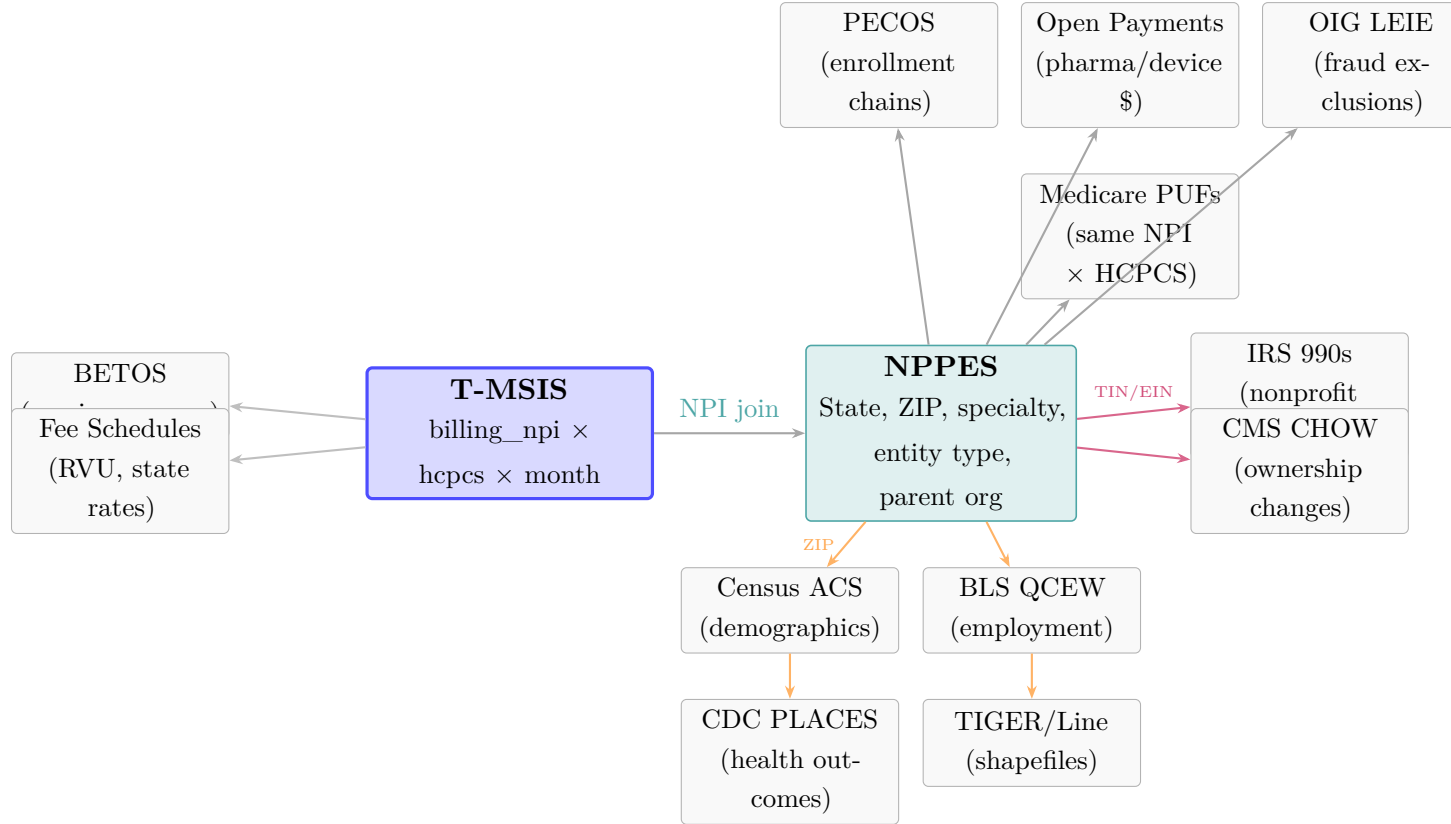**Figure 8:** The T-MSIS Linkage Universe

Notes: The NPI serves as the universal join key, connecting T-MSIS to NPPES (teal arrow), which in turn connects to external datasets via three channels: direct NPI match (gray), geographic match via ZIP code (orange), and organizational match via TIN/EIN (purple). Representative datasets shown; the full universe includes 30+ linkable sources.

4.1 The Essential First Link: NPPES

The National Plan and Provider Enumeration System (NPPES) transforms anonymous NPI numbers into identified, geolocated, specialty-classified providers. One join changes everything. The bulk NPPES download contains 329 fields per provider, of which we extract the most analytically valuable:

- **Geography:** Practice location state and 9-digit ZIP code—the foundation for all area-level analysis and mapping.

- **Provider classification:** Entity type (1 = individual, 2 = organization) and NUCC healthcare provider taxonomy code (800+ specialties, from internal medicine to home health aide).

- **Organizational structure:** Parent organization legal business name and TIN, sole proprietor flag, and organization subpart indicator.

- **Provider lifecycle:** Enumeration date (market entry), deactivation date and reason (market exit), and reactivation date.

- **Demographics:** Provider gender and credential text (for individual NPIs).

The NPPES match rate on billing NPI is 99.5%: we successfully match virtually all billing NPIs in T-MSIS to NPPES records, enabling state-level analysis, provider-type stratification, and all downstream geographic joins. The 0.5% of billing NPIs without NPPES matches are excluded from geographic and organizational analyses but remain in unaggregated spending totals.

4.2 Three Linkage Channels

From the NPPES-enriched dataset, three channels open:

Channel 1: NPI direct match. Any dataset containing NPI can be joined directly. The most important is the Medicare Physician/Supplier PUF, which has the *identical* NPI × HCPCS × year structure as T-MSIS—enabling provider-level cross-payer spending comparisons. Open Payments (the Sunshine Act database) adds pharmaceutical and device industry payments by NPI. PECOS provides Medicare enrollment chains and group practice affiliations. The OIG List of Excluded Individuals/Entities (LEIE) identifies providers excluded for fraud. Care Compare adds quality scores.

Channel 2: Geographic match via ZIP. NPPES provides 9-digit ZIP codes, which standard crosswalks map to counties, CBSAs, census tracts, and ZIP Code Tabulation Areas (ZCTAs). This unlocks the entire universe of area-level data: Census ACS (demographics, poverty, insurance), BLS QCEW (healthcare employment by county), CDC PLACES (chronic disease prevalence by tract), HRSA shortage area designations, USDA rural-urban classifications, and Census TIGER/Line shapefiles for mapping.

Channel 3: Organizational match via TIN/EIN. NPPES provides parent organization Tax Identification Numbers, enabling linkage to IRS Form 990 data (nonprofit financials

via ProPublica), CMS Cost Reports (hospital financials), CMS Hospital Change of Ownership data (M&A events), and SEC filings for publicly traded health care firms.

Each channel multiplies the research possibilities. A single NPI-based join to Medicare data enables cross-payer substitution analysis for every provider in the file. A single geographic join to QCEW enables labor market analysis for every county. The channels are independent and compose: a study of how hospital acquisitions (Channel 3) affect Medicaid billing in the acquiring hospital's county (Channel 2) for providers who also bill Medicare (Channel 1) requires all three channels simultaneously, and all three are feasible with public data.

4.3 Noteworthy Linkages for Health Economics

Several specific linkages deserve emphasis for their research potential:

Birth certificates. NCHS vital statistics birth certificates record the *attendant NPI*— the provider who delivered the baby. Since Medicaid pays for 42% of U.S. births, linking T-MSIS obstetric billing to birth outcomes (birth weight, gestational age, C-section rates) via attendant NPI creates one of the rare settings where provider spending and patient outcomes can be directly connected at the provider level.

Medicare Part D prescribing. The Medicare Part D Prescriber PUF records every drug prescribed by NPI. Linking to T-MSIS reveals which Medicaid providers also prescribe controlled substances under Medicare, enabling analysis of prescribing patterns by Medicaid billing volume, specialty, and HCBS versus clinical orientation.

CMS Hospital CHOW data. The Hospital Change of Ownership file records every Medicare-certified hospital acquisition since January 2016 with precise effective dates. Combining CHOW dates with T-MSIS spending creates a natural experiment: staggered hospital acquisitions as treatment events, with provider-level Medicaid billing as the outcome. This design can answer whether ownership changes alter Medicaid billing patterns—a question motivated by recent work on private equity in health care (Eliason et al., 2020; Gupta et al., 2024).

5. Constructed Analysis Panels

The linkage architecture tells researchers what they *can* connect to. This section describes what we *built*. We construct four analysis panels, each targeting a distinct class of research questions.

5.1 State–Provider-Type–Month Panel

The most natural unit of analysis for evaluating state-level Medicaid policies is a panel that aggregates T-MSIS billing data to the state–provider-type–month level. We assign each billing NPI a state using the NPPES practice location, classify providers by entity type and NUCC taxonomy, and collapse spending, claims, beneficiaries, and active provider counts within state × provider-type × HCPCS-category × month cells. This panel is purpose-built for difference-in-differences evaluation of staggered state policies. The primary application is HCBS wage increases: North Carolina raised its personal care reimbursement rate from $3.93 to $5.50 per 15-minute unit in March 2022; Colorado increased HCBS rates by 6–10% in 2022–23; other states followed. The panel enables estimation of how spending, claims volume, and provider counts in T-code services respond to wage floor changes, using the remaining states as controls. The 2018–2024 window contains at least three classes of staggered state-level policies suitable for this design: HCBS wage increases affecting six or more states, Medicaid postpartum coverage extensions to 12 months adopted by 47 states on different timelines, and the Medicaid unwinding beginning April 2023 with state-level start dates varying by several months.

5.2 Provider-Level Cross-Payer Panel

A second panel exploits the identical data structure shared by T-MSIS and the Medicare Physician/Supplier PUF. Because both files are organized at the NPI × HCPCS × time-period level, we merge them to create a provider-level cross-payer dataset with Medicaid payments from T-MSIS and Medicare payments and RVU-adjusted volume from the Medicare PUF for the same provider billing the same procedure code. This panel directly addresses longstanding questions about cross-payer substitution: when Medicaid reimbursement rates increase, do providers shift volume *toward* Medicaid or *away from* Medicare? When Medicaid enrollment drops during unwinding, do providers increase Medicare billing to compensate? Maryland's July 2022 Medicaid reimbursement increase to 100% of Medicare rates provides a particularly clean natural experiment—a single state with a precisely timed, large payment shock, with 49 control states—but the panel supports cross-payer analysis for any state-level Medicaid payment change during the sample period.

5.3 Firm-Level Market Structure Panel

The third panel shifts the unit of analysis from individual providers to firms. We group all billing NPIs that share the same parent organization TIN in NPPES, then aggregate spending, NPI counts, and market shares within firm × county × HCPCS-category × year cells. From

these aggregates, we compute Herfindahl-Hirschman Indices for specific service categories within county-year markets—something previously impossible for Medicaid services. The resulting panel enables studying how concentrated HCBS markets are, whether large multi-state agencies such as BrightSpring, Addus, and Maxim command different reimbursement or serve different populations than small local providers, and whether market concentration affects Medicaid payment adequacy. Combined with CMS Hospital Change of Ownership data, the panel also supports event-study designs around acquisitions, measuring how ownership changes alter Medicaid billing patterns at the firm level.

5.4 Provider Lifecycle and Workforce Dynamics Panel

The fourth panel tracks individual providers over time at the NPI $\times$ quarter level, recording whether each provider is actively billing, what HCPCS codes they bill, their total volume, and their tenure since initial NPPES enumeration. This panel supports survival analysis of Medicaid providers: what determines how long a provider bills, whether states with higher Medicaid reimbursement exhibit longer provider tenure, and whether wage increases reduce exit rates among HCBS providers. The NPPES enumeration and deactivation dates provide independent confirmation of the entry and exit patterns observed in T-MSIS billing data. BLS QCEW healthcare employment by county offers a complementary labor-market validation: if T-MSIS shows increasing HCBS provider counts in a state, QCEW should show rising employment in NAICS 6216 (home health care services) and 6241 (individual and family services). Discrepancies between the two sources would flag potential measurement issues in either dataset.

6. Research Agenda

The combination of T-MSIS data, NPI-based linkages, and the four constructed panels described above opens a research program spanning several subfields of health economics. We outline the most promising directions below, emphasizing the identification strategies that the data's structure makes possible.

The most immediate research opportunity concerns payment adequacy and provider supply. How do Medicaid reimbursement rates affect the supply of providers willing to serve Medicaid patients? Prior work has documented both the low level of Medicaid fees relative to Medicare (Zuckerman et al., 2014) and their effect on appointment availability (Decker, 2012; Polsky et al., 2015). The T-MSIS data provide the first national provider-level outcome variable for this question. The policy variation is unusually clean: North Carolina raised its HCBS personal care reimbursement rate from $3.93 to $5.50 per 15-minute unit in March

2022—a 40% increase. Colorado increased HCBS provider rates by 6–10% effective January 2022, with a further increase to a $15 minimum in 2023. Virginia raised its personal care rate by 12% in July 2021 as part of its ARPA HCBS spending plan. New Mexico implemented across-the-board HCBS rate increases of 8.5% in April 2022. Because these increases occurred at different times across states, the state–provider-type–month panel enables a standard staggered difference-in-differences design (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024): the treated states at the time of their rate increase versus the not-yet-treated and never-treated states as controls. The key outcome—T-code billing volume and active provider counts in the treated states—is directly observable in T-MSIS for the first time.

A closely related set of questions concerns cross-payer dynamics. Do providers treat Medicaid and Medicare as substitutes? The cross-payer panel enables direct estimation of how Medicaid payment changes affect the same provider's Medicare billing, speaking to longstanding questions about cost-shifting and crowd-out in health care markets (Clemens and Gottlieb, 2014; Curto et al., 2019). Maryland's July 2022 rate increase to 100% of Medicare rates offers the cleanest natural experiment, but the design generalizes to any state-level payment change during the sample period. The interaction between payment adequacy and cross-payer dynamics is itself of interest: if Medicaid rate increases draw providers toward Medicaid billing and away from Medicare, the welfare implications depend on the relative health returns to services across the two populations (Currie and Gruber, 1996).

The firm-level panel opens a third line of inquiry into market structure and consolidation. How concentrated are Medicaid HCBS markets, and does concentration affect prices, access, or quality? This question connects directly to a rapidly growing literature on private equity and corporate consolidation in health care. Eliason et al. (2020) showed that acquisitions in the dialysis industry led to changes in treatment intensity and patient outcomes; Gupta et al. (2024) documented quality declines in nursing homes acquired by private equity firms. Both studies relied on Medicare data. The T-MSIS firm-level panel extends this research program to Medicaid by enabling researchers to track multi-state HCBS firms—BrightSpring Health Services, Addus HomeCare, Maxim Healthcare—as they acquire smaller agencies and expand across markets, and to measure how their billing footprint, provider counts, and market share evolve over time. The HCBS market is structurally different from hospitals or dialysis centers: personal care and residential habilitation services are labor-intensive, low-margin, and geographically fragmented—precisely the conditions under which consolidation can produce large changes in market power with minimal regulatory scrutiny.

These market-level questions are inseparable from workforce dynamics. The extraordinary turnover documented in Figure 4—over 38% of providers billing for fewer than 12 months—

raises the question of what determines provider tenure in Medicaid. The provider lifecycle panel, combined with QCEW employment data and OEWS wage data, enables triangulating between billing-side and employment-side measures of workforce dynamics. State minimum wage changes, which disproportionately affect low-wage HCBS workers, provide identification. Market concentration itself may shape workforce outcomes: in concentrated HCBS markets, dominant agencies may suppress wages or reduce working conditions, accelerating turnover in ways that are measurable through the provider panel.

Finally, the 2018–2024 window captures three major shocks, each with state-level variation suitable for quasi-experimental analysis: the COVID-19 pandemic, the ARPA HCBS spending increase beginning April 2021, and the Medicaid unwinding beginning April 2023. The unwinding deserves particular emphasis. During the public health emergency, the Families First Coronavirus Response Act prohibited states from disenrolling Medicaid beneficiaries, causing enrollment to swell from 71 million in February 2020 to over 94 million by March 2023. When the continuous enrollment requirement expired on April 1, 2023, states began redetermining eligibility for their entire populations. By late 2024, over 25 million people had been disenrolled—many for procedural reasons rather than actual ineligibility. States varied dramatically in their unwinding timelines: Arkansas and Idaho began processing renewals in April 2023, while California and New York delayed until mid-2024. Some states adopted aggressive procedural disenrollment; others implemented ex parte renewals using existing data to confirm eligibility without requiring beneficiary action. This state-level variation in timing and aggressiveness is directly exploitable in T-MSIS, and the provider lifecycle panel can measure whether the unwinding triggers provider exit, particularly among small HCBS agencies whose entire caseload may consist of Medicaid beneficiaries.

These research directions are not independent. A study of how wage increases affect provider tenure in concentrated markets during the post-ARPA period would draw on all four panels simultaneously, and the richest empirical designs will combine multiple themes.

7. Conclusion

For the first time, researchers can see who delivers Medicaid services in the United States. The T-MSIS Medicaid Provider Spending dataset transforms the supply side of the nation's largest health insurance program from a black box into an observable, analyzable, linkable provider panel.

What we find inside is surprising. Over half of Medicaid provider spending flows through procedure codes—for personal care, behavioral health, habilitation—that have no Medicare equivalent whatsoever. The provider workforce churns rapidly: over 38% of billing NPIs

appear for fewer than 12 months, and only 6% persist across the full 84-month panel. The organizational structure—revealed through billing–servicing NPI relationships and NPPES parent organization identifiers—shows that two-thirds of claim rows flow through organizational billing arrangements, suggesting markets that may be substantially more concentrated than they appear from provider counts alone.

The NPI serves as a universal joint, connecting this Medicaid claims universe to Medicare billing, pharmaceutical industry payments, fraud exclusion lists, geographic characteristics, labor market data, health outcomes, and corporate ownership records. The resulting linkage architecture supports research designs that were literally impossible before this data release.

We have described the dataset, documented what it reveals about the Medicaid provider landscape, mapped the linkage universe, and constructed four specific analysis panels for future research. The research agenda is vast: payment adequacy, cross-payer dynamics, market structure, workforce retention, and the natural experiments of COVID-19, ARPA, and the Medicaid unwinding all become answerable with provider-level data.

The gap between what we know about Medicare providers and what we know about Medicaid providers has been, for decades, one of the most consequential blind spots in health economics. That gap was a choice; now, it is a relic.

Acknowledgements

References

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Buchmueller, Thomas C, Zachary M Levinson, Helen G Levy, and Barbara L Wolfe, "Effect of the Affordable Care Act on Racial and Ethnic Disparities in Health Insurance Coverage," *American Journal of Public Health*, 2016, *106* (8), 1416–1421.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Centers for Medicare & Medicaid Services, "Medicaid Provider Spending Dataset," Technical Report, U.S. Department of Health and Human Services 2026. Released February 9, 2026. Available at: opendata.hhs.gov.

Clemens, Jeffrey and Joshua D Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

_ **and** _ , "In the Shadow of a Giant: Medicare's Influence on Private Physician Payments," *Journal of Political Economy*, 2017, *125* (1), 1–39.

Currie, Janet and Jonathan Gruber, "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *Quarterly Journal of Economics*, 1996, *111* (2), 431–466.

Curto, Vilsa, Liran Einav, Amy Finkelstein, Jonathan Levin, and Jay Bhattacharya, "Health Care Spending and Utilization in Public and Private Medicare," *American Economic Journal: Applied Economics*, 2019, *11* (2), 302–332.

Dafny, Leemore S, "How Do Hospitals Respond to Price Changes?," *American Economic Review*, 2005, *95* (5), 1525–1547.

Decker, Sandra L, "In 2011 Nearly One-Third of Physicians Said They Would Not Accept New Medicaid Patients, But Rising Fees May Help," *Health Affairs*, 2012, *31* (8), 1673–1679.

Eliason, Paul J, Benjamin Heebsh, Ryan C McDevitt, and James W Roberts, "How Acquisitions Affect Firm Behavior and Performance: Evidence from the Dialysis Industry," *Quarterly Journal of Economics*, 2020, *135* (1), 221–267.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Grabowski, David C, "The Cost-Effectiveness of Noninstitutional Long-Term Care Services: Review and Synthesis of the Most Recent Evidence," *Medical Care Research and Review*, 2006, *63* (1), 3–28.

Gupta, Atul, Sabrina T Howell, Constantine Yannelis, and Abhinav Gupta, "Private Equity and the Quality of Nursing Home Care," *Review of Financial Studies*, 2024.

Medicaid and CHIP Payment and Access Commission, "MACStats: Medicaid and CHIP Data Book," Technical Report, MACPAC 2024.

Polsky, Daniel, Michael Richards, Simon Basseyn, Donna Bocour, David Blumenthal, Allison Hamblin, Genevieve M Kenney, Lisa Schwartz, Craig Evan Pollack, and Janet Weiner, "Appointment Availability after Increases in Medicaid Payments for Primary Care," *New England Journal of Medicine*, 2015, *372* (6), 537–545.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Zuckerman, Stephen, Dana Goin, and Urban Institute, "Trends In Medicaid Physician Fees, 2003–2012," *Health Affairs*, 2014, *33* (10), 1846–1853.

Appendix Table of Contents

A. Data Dictionary

A.1 T-MSIS Raw Schema

Table 7: T-MSIS Medicaid Provider Spending: Complete Field Definitions

Field	Type	Description
BILLING_PROVIDER_ NPI_NUM	string(10)	National Provider Identifier of the entity that submitted the claim. May be an individual (Type 1) or organization (Type 2).
SERVICING_PROVIDER_ NPI_NUM	string(10)	NPI of the provider who performed the service. Null in ∼4% of rows (typically organizational billing). Equals billing NPI in ∼31% of rows; differs in ∼65%.
HCPCS_CODE	string(5)	Healthcare Common Procedure Coding System code. Identifies the specific service performed. Includes CPT (numeric), T-codes (HCBS), H-codes (behavioral), S-codes (temporary), J-codes (drugs), and others.
CLAIM_FROM_MONTH	date (YYYY-MM)	Month of service. Not the claim submission date.
TOTAL_UNIQUE_ BENEFICIARIES	integer	Count of distinct Medicaid beneficiaries served in this NPI–code–month cell. Suppressed (row excluded) when < 12.
TOTAL_CLAIMS	integer	Total claim lines submitted.
TOTAL_PAID	float	Total Medicaid amount paid in USD. Includes FFS payments, MCO encounter values, and CHIP.

Notes: Primary key = (billing_npi, servicing_npi, hcpcs_code, month). No duplicates. Coverage: all 50 states + DC + territories, Jan 2018–Dec 2024. Source: HHS Open Data, derived from T-MSIS.

A.2 NPPES Fields Used

Table 8: NPPES Fields Extracted for T-MSIS Enrichment

Field	Type	Research Use
Entity Type Code	1/2	Individual (1) vs. Organization (2). Critical for provider-type stratification.
Practice State	string(2)	State assignment for DiD. From practice location, not mailing address.
Practice ZIP (9-digit)	string(9)	Foundation for all geographic joins: county, CBSA, tract, ZCTA.
Taxonomy Code (primary)	string(10)	NUCC specialty classification. 800+ codes, 3-level hierarchy.
Credential Text	string	MD, DO, NP, PA, LCSW, RN, etc.
Gender Code	M/F	Individual providers only.
Parent Org LBN	string	Legal name of parent organization. Groups subsidiaries.
Parent Org TIN	string(9)	Tax ID of parent. Most reliable firm identifier.
Is Sole Proprietor	Y/N	Distinguishes independent from employed providers.
Enumeration Date	date	When NPI was issued—proxy for market entry.
Deactivation Date	date	When deactivated—proxy for market exit.
Deactivation Reason	code	Death, disbandment, fraud, other.

Notes: NPPES bulk download contains 329 fields per NPI. We extract the 12 most analytically useful. Updated monthly by CMS.

B. HCPCS Code Reference

B.1 Code Categories by Prefix

Table 9: HCPCS Code Categories in T-MSIS Data

Prefix	Category	Examples and Description
T	HCBS / State	T1019 (personal care/15 min), T2016 (habilitation residential/diem), T1015 (FQHC visit). State-specific Medicaid codes for long-term services.
H	Behavioral Health	H2016 (community support/diem), H2015 (community support/15 min), H0036 (community psychiatric). Alcohol, drug, and mental health services.
S	Temporary / State	S5125 (attendant care/15 min), S5150 (unskilled respite care), S9470 (home infusion). Temporary codes for services not classified elsewhere.
0–9	CPT Professional	99213 (office visit, established), 99214 (office visit, complex), 97153 (ABA therapy). Standard medical procedures.
J	Drugs	J-codes for injectable drugs administered by providers.
A	Ambulance / DME	Ambulance transport, medical supplies.
G	CMS Procedures	CMS-specific procedure codes (e.g., telehealth modifiers).
E	DME	Durable medical equipment (wheelchairs, beds).
L	Orthotics / Prosthetics	Orthotic and prosthetic devices.

Notes: T, H, and S codes are Medicaid-specific—they have no Medicare equivalent. Numeric codes (CPT) are shared with Medicare. J, A, G, E, L codes appear in both programs.

B.2 Top 6 Codes: Service Descriptions

T1019 — Personal Care Services (per 15 minutes). Assistance with activities of daily living (bathing, dressing, eating, mobility) provided by a personal care aide in the beneficiary's home. This single code represents the backbone of Medicaid HCBS. Billing in 15-minute increments means a typical 4-hour shift generates 16 claims.

T2016 — Habilitation, Residential, Waiver (per diem). Per-diem payments for residential services under Medicaid HCBS waivers, primarily serving individuals with intellectual and developmental disabilities (I/DD) in group home settings. High per-unit cost reflects 24-hour residential care.

S5125 — Attendant Care Services (per 15 minutes). Similar to T1019 but classified as a temporary state code. Covers in-home attendant care for beneficiaries needing assistance with daily activities.

H2016 — Comprehensive Community Support (per diem). Per-diem community-based support for individuals with serious mental illness. Includes psychosocial rehabilitation, skills training, and supervised community living.

T1015 — Clinic Visit, FQHC (all-inclusive). Prospective payment for visits to Federally Qualified Health Centers, which serve as the primary care safety net for Medicaid populations, particularly in underserved areas.

H2015 — Comprehensive Community Support (per 15 minutes). Same service category as H2016 but billed in 15-minute rather than per-diem increments. Common for partial-day behavioral health programs.

C. Geographic Enrichment Methodology

C.1 ZIP Code to County Crosswalk

NPPES provides 9-digit ZIP codes. We map these to counties using the Census Bureau's ZCTA-to-County relationship file, which provides population-weighted correspondence between ZIP Code Tabulation Areas and FIPS county codes. For ZIPs spanning multiple counties, we assign to the county with the largest population share. This approach matches approximately 97% of provider ZIP codes to a unique county.

C.2 Mapping and Visualization

All choropleth maps use Census TIGER/Line shapefiles, downloaded via the `tigris` R package. State boundaries use the 2022 cartographic boundary file (cb = TRUE) for cleaner

coastlines. Alaska and Hawaii are shifted using `tigris::shift_geometry()` for continental U.S. display. Color scales use the `viridis` perceptually uniform palette family for accessibility.

C.3 ZCTA-Level Analysis

For sub-state analysis, we aggregate T-MSIS spending to the 5-digit ZIP level (from NPPES) and map to ZCTA polygons. The resulting ZCTA-level spending measures enable fine-grained geographic analysis—provider density, spending per capita, and market concentration at a resolution finer than counties but coarser than census tracts.

D. Data Quality Assessment

D.1 Cell Suppression

The minimum cell size of 12 beneficiaries means that low-volume provider–procedure–month combinations are excluded. This creates systematic censoring: rural providers serving small populations, providers offering rare services, and new entrants with few initial patients are disproportionately suppressed. The resulting dataset overrepresents high-volume urban providers and common procedures. Researchers should be cautious when drawing conclusions about rural access or rare services from this data.

D.2 Panel Completeness

The December 2024 data are incomplete due to standard claims processing lags. We recommend excluding December 2024 from longitudinal analysis. January 2018 data may also reflect incomplete T-MSIS reporting from some states; CMS's Data Quality Atlas provides state-by-state assessments of T-MSIS completeness.

D.3 FFS / Managed Care Commingling

The absence of a payer type indicator (FFS vs. MCO) means that state-level totals reflect the sum of FFS claims and MCO encounter records. Because MCO encounter "payment" amounts may reflect capitated rate imputations rather than actual fee-for-service payments, cross-state spending comparisons should be interpreted cautiously. States with high managed care penetration (e.g., Tennessee at $\sim$100%) may report encounter-based amounts that are not directly comparable to FFS payments in lower-penetration states.

D.4 Beneficiary Counting

The TOTAL_UNIQUE_BENEFICIARIES field counts distinct beneficiaries *within each cell*. Because the same beneficiary may appear in multiple provider–procedure–month cells, summing beneficiary counts across cells will overcount total individuals served. There is no beneficiary identifier in the data that would enable deduplication. Beneficiary counts are useful *within* cells (e.g., comparing beneficiaries per provider across states) but not for estimating total program enrollment.

D.5 NPPES Data Currency

The NPPES bulk download reflects current provider information, not historical. A provider who practiced in New York in 2018 but moved to Florida in 2025 will show a Florida practice address. For longitudinal analysis, researchers should consider using the NPPES historical deactivation/reactivation dates to identify providers whose addresses may have changed during the study period. The NPPES monthly update files (rather than the full replacement) can provide historical snapshots if archived.

E. Linkage Datasets: Detailed Access Information

E.1 Confirmed Accessible (Tested February 2026)

Dataset	Access Method	Notes
T-MSIS Provider Spending	Direct download (HHS)	2.9 GB Parquet, 11 GB CSV. No auth required.
NPPES Bulk Extract	Direct download (CMS)	∼1 GB ZIP, ∼7 GB CSV. Monthly updates.
Medicare Physician PUF	data.cms.gov Socrata	No auth. Same NPI × HCPCS structure.
Medicare Part D PUF	data.cms.gov Socrata	No auth. Drug prescribing by NPI.
Census ACS 5-Year	Census API	Requires API key (free). County/tract/ZCTA.
BLS QCEW	Direct CSV (bls.gov)	∼83 MB/year. NAICS 6216, 6241 for HCBS.

Dataset	Access Method	Notes
CDC PLACES	data.cdc.gov Socrata	No auth. Tract-level health estimates.
OIG LEIE	Direct CSV (oig.hhs.gov)	Exclusion list with NPI. ~15 MB.
ProPublica 990 API	REST API (no auth)	Nonprofit financials by EIN.
Census TIGER/Line	Direct FTP (census.gov)	ZCTA, county, tract shapefiles.
FRED	REST API	Requires key (free). Macro indicators.

Table 10: Confirmed Accessible Datasets for T-MSIS Linkage

E.2 Likely Accessible (Untested, Similar Patterns)

BLS OEWS (wage distributions by SOC), Census QWI (quarterly workforce indicators), BEA economic accounts, USDA rural-urban codes, County Health Rankings, CDC SVI, NUCC taxonomy CSV, BRFSS and NHIS public use files, NCHS vital statistics PUFs, CMS cost reports, CMS Provider of Services file, PECOS enrollment files, CMS Hospital CHOW data.

E.3 Restricted Access (Requires Application or Subscription)

State All-Payer Claims Databases (per-state DUA), HCUP databases (application + fee), AHA Annual Survey (institutional subscription), Ideon/RWJF network data (academic application), SEER-Medicare linked files (NCI application), NPDB full file, SafeGraph/Dewey foot traffic data, Lightcast job postings, Definitive Healthcare, NAIC financial data.
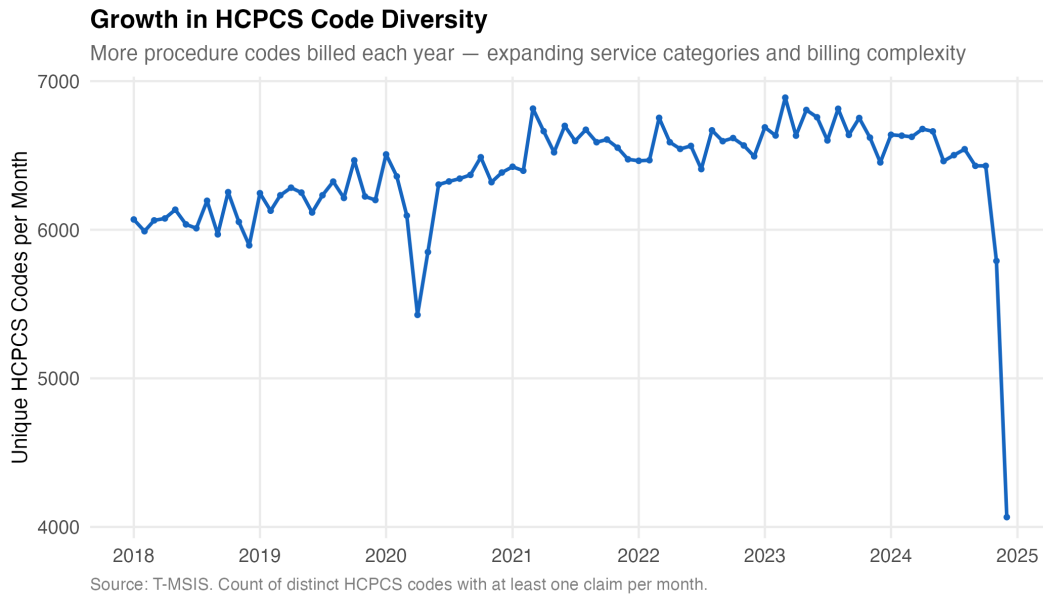
F. Additional Figures

Figure 9: Growth in HCPCS Code Diversity, 2018–2024

Notes: Number of distinct HCPCS codes with at least one claim per month. The expanding code universe reflects new service categories and billing complexity over time.