Watching the Workers: Electronic Visit Verification and the Medicaid Home Care Market

Abstract

Background. The 21st Century Cures Act (2016) mandated Electronic Visit Verification (EVV) for all Medicaid personal care services (PCS) by January 1, 2020, and home health care services (HHCS) by January 1, 2023, with the Congressional Budget Office projecting $290 million in federal savings through fraud deterrence. No causal evaluation of EVV's effects on spending or service delivery has been published.

Methods. I exploit staggered EVV implementation timing across 51 jurisdictions to estimate effects on spending, utilization, provider participation, and service delivery patterns using difference-in-differences. My primary outcomes are drawn from T-MSIS Medicaid claims data (2018–2024) at quarterly frequency, with CMS Financial Management Report (FMR) expenditure data as secondary validation. I implement two-way fixed effects (TWFE) models, the Gardner (2022) DID2S estimator, a Goodman-Bacon decomposition, and a triple-difference design leveraging the PCS-HHCS deadline gap. I decompose spending effects into price (dollars per claim), utilization intensity (claims per provider), and extensive margin (provider count) channels. Estimates are interpreted as intent-to-treat effects given evidence that operational implementation lags formal compliance.

Results. EVV compliance had no statistically significant effect on quarterly PCS claims spending (TWFE: +$0.62M per state-quarter, SE=$18.61, p=0.974; N=1,379) or total HCBS claims spending (+$2.37M, SE=$17.54, p=0.893). Mechanism decomposition revealed no significant effects on any channel: price (PCS $/claim: -$36.71, p=0.145), utilization intensity (PCS claims/provider: +2,870, p=0.284), provider counts (PCS: -1.2, p=0.780), or beneficiary counts (PCS: +4,058, p=0.615). FMR-based estimates corroborate the null. Pre-trends are not rejected for PCS claims (Wald p=0.875) but are rejected for HHCS (p=0.0006). The minimum detectable effect ($52.14M per state-quarter, 55.6% of baseline) exceeds the CBO projection by a factor of 209. As a secondary mechanism check, annual HHCS spending shows a directionally positive but imprecise estimate under the observed-cells primary specification (+$24.26M per state-year, SE=$13.71, p=0.083); under a zero-fill convention that treats unobserved state-quarters as zero claims, the quarterly HHCS estimate shifts to +$4.55M (p=0.094). The HHCS direction is therefore positive across conventions but not causally interpretable, and the HHCS results should be read as a measurement-policy-sensitive signal rather than a causal estimate.

Conclusions. EVV has not produced large aggregate spending reductions. The most informative mechanism evidence is against large provider-exit effects, while

smaller price and utilization changes remain unresolved. The CBO-projected savings cannot be ruled out given power limitations. The bounded null is consistent with multiple scenarios: genuine ineffectiveness, offsetting documentation improvements, effects too small for aggregate data, or incomplete implementation. The HHCS pre-trend violation (Wald p=0.0006) and measurement-policy sensitivity together warrant strong caution in interpreting any HHCS-specific result as causal.

Introduction

The United States spends over $100 billion annually on Medicaid home and community-based services (HCBS), a figure that has grown rapidly as states have shifted long-term care from institutional to community settings (MACPAC, 2019; Mathematica, 2023). Personal care services (PCS)—in which attendants assist beneficiaries with activities of daily living in their homes—represent a particularly large and fast-growing component of this spending, exceeding $15 billion in fee-for-service expenditures alone by 2015 (GAO, 2017). The decentralized nature of home care delivery, occurring in private residences with limited direct supervision, has made these services a persistent target of fraud enforcement. Personal care attendant fraud has consistently represented the largest single category of Medicaid Fraud Control Unit convictions, accounting for 34 to 43 percent of all convictions from 2012 through 2022 (OIG MFCU Annual Reports, multiple years).

In response, Congress included an Electronic Visit Verification (EVV) mandate in the 21st Century Cures Act (Public Law 114-255), enacted in December 2016. Section 12006(a) requires all state Medicaid programs to implement EVV systems that electronically record six visit elements—the service recipient, the service provider, the service type, the date, the location, and the start and end times—for PCS by January 1, 2020 (extended one year for good-faith-effort compliance) and for HHCS by January 1, 2023. States that fail to comply face escalating reductions in their Federal Medical Assistance Percentage (FMAP) of 0.25 to 1.0 percentage points. The Congressional Budget Office scored this provision as generating approximately $290 million in federal savings over ten years, based on an assumed 1 percent reduction in home care payments through fraud deterrence (CBO, 2016).

The EVV mandate is now operationally active in all 50 states and the District of Columbia for PCS, and in 44 jurisdictions for HHCS. Yet despite the policy's scale—affecting every state Medicaid program's home care infrastructure—no published study has used quasi-experimental methods to evaluate its causal effects on spending, utilization, or access. The existing evidence base consists of qualitative studies documenting EVV's burden on workers and consumers (Steckler et al., 2020; Mateescu, 2021), state audit reports revealing incomplete implementation (Ohio Auditor of State, 2024; OIG, 2024), and descriptive fraud statistics suggesting possible reductions in PCS-related convictions (OIG MFCU Annual Report, FY 2024).

This paper provides the first quasi-experimental evaluation, to my knowledge, of EVV's effect on Medicaid home care spending, utilization, and provider participation. I exploit the substantial variation in EVV implementation timing across states—driven by vendor procurement delays and IT infrastructure challenges rather than strategic policy choices—to implement a staggered difference-in-differences design. My identification strategy combines two complementary approaches. First, I compare spending changes in states that implemented EVV earlier versus later, using state and time fixed effects to absorb time-invariant heterogeneity and common temporal shocks. Second, I leverage the three-year gap between the PCS and HHCS deadlines to construct a supplementary triple-difference design: during the period when PCS EVV was operational but HHCS EVV was not yet implemented, HHCS spending within the same state serves as a within-state control.

I draw on two complementary data sources. My primary outcomes come from the T-MSIS Medicaid Provider Spending dataset, which provides provider-procedure-month Medicaid claims aggregated to the state-quarter level. These claims data offer three key advantages over aggregate administrative expenditure data: genuine quarterly frequency (versus annual Financial Management Report data divided by four), HCPCS-level specificity that isolates exactly the services subject to EVV, and the ability to decompose spending effects into price (dollars per claim), utilization intensity (claims per provider), extensive margin (provider count), and beneficiary reach. I use CMS Financial Management Report (FMR) data as secondary validation.

I construct a novel state-level panel covering 51 jurisdictions over 2018–2024 by combining these outcome data with a hand-coded EVV implementation timing panel drawn from CMS compliance records and state administrative documents, and state-level control variables including the ARPA Section 9817 enhanced FMAP, the COVID-19 public health emergency, and Medicaid expansion status.

My central finding is that EVV compliance has not produced large aggregate spending reductions. The quarterly TWFE estimate for PCS claims spending is +\$0.62 million per state-quarter (SE=\$18.61, p=0.974). Mechanism decomposition reveals no statistically significant effects on any channel: PCS dollars per claim (-\$36.71, p=0.145), claims per provider (+2,870, p=0.284), provider counts (-1.2, p=0.780), or beneficiaries (+4,058, p=0.615). FMR-based results corroborate the bounded-null pattern. Pre-treatment trends are cleanly not rejected for PCS claims (Wald p=0.875), and randomization inference confirms the null (p=0.966). A formal power analysis reveals a minimum detectable effect of \$52.14 million per state-quarter—209 times the CBO's projected total computable savings of approximately \$0.25 million per state-quarter—so the CBO projection cannot be ruled out.

This bounded-null finding constitutes the paper's primary contribution. I show that EVV compliance has not generated large aggregate spending reductions and—crucially, using claims data—that the most informative mechanism evidence is against large provider-exit effects. However, I stress that drawing strong policy conclusions from the null is not warranted given power limitations. The results remain consistent with multiple interpretations: (A) EVV genuinely does not reduce spending; (B) savings are offset by documentation improvements that capture previously unreported legitimate services; (C) EVV's true effect is too small for aggregate data to detect; or (D) EVV was never effectively implemented at scale.

I report HHCS spending in parallel as a secondary mechanism check rather than as a co-equal primary outcome. Under the observed-cells primary specification, annual HHCS spending shows a directionally positive but imprecise estimate (+\$24.26M per state-year, SE=\$13.71, p=0.083); under a zero-fill convention that treats unobserved state-quarters as zero claims, the quarterly HHCS estimate shifts to +\$4.55M (p=0.094). The HHCS result is directionally positive across conventions but not causally interpretable, and pre-trends for HHCS are rejected (Wald p=0.0006). For both reasons the HHCS finding cannot be read as a causal estimate; I retain it in the main text because it is a substantive mechanism check that any reader will want to see, but I disclose its measurement-policy sensitivity explicitly so readers do not over-interpret a marginal p-value. Other important limitations include T-MSIS data quality issues for managed care states and the gap between formal compliance and operational use, which means my estimates are intent-to-treat effects.

The paper proceeds as follows. Section II provides background on the HCBS policy landscape, the fraud concerns motivating EVV, and the existing evidence base. Section III describes my data sources and panel construction. Section IV presents the empirical strategy. Section V reports results. Section VI discusses interpretation, policy implications, and limitations. Section VII concludes.

Background

The Growth of Home and Community-Based Services

Medicaid is the nation's primary payer for long-term services and supports (LTSS), financing care for elderly individuals, people with disabilities, and others who require ongoing assistance with activities of daily living. Over the past three decades, Medicaid LTSS spending has shifted dramatically from institutional to community-based settings. The share of Medicaid LTSS expenditures devoted to HCBS rose from 12 percent in FY 1989 to approximately 64 percent by FY 2022 (MACPAC, multiple years; Mathematica, 2023). In absolute terms, joint federal-state Medicaid HCBS spending reached \$116 billion in FY 2020, serving roughly 2.5 million beneficiaries through 1915(c) waiver programs and an additional 2.5 million through state plan benefits including personal care services and home health (KFF, 2022). This rebalancing was driven by the Supreme Court's *Olmstead v. L.C.* (1999) decision, the Money Follows the Person demonstration, and the ACA's Balancing Incentive Program (Konetzka, 2014).

Personal care services represent a particularly large and fast-growing category within HCBS, encompassing assistance with activities of daily living (ADLs) and instrumental activities of daily living (IADLs) delivered in the beneficiary's home. States deliver PCS through multiple Medicaid authorities—the state plan personal care option (Section 1905(a)(24)), the Community First Choice option (Section 1915(k)), 1915(c) waivers, and managed LTSS programs—creating substantial cross-state variation in financing and reporting. Fee-for-service PCS spending reached at least \$15 billion by 2015, up \$2.3 billion from 2012 (GAO, 2017). Self-directed services, in which beneficiaries hire and manage their own attendants—often family members—are available in all 50 states and served over 1.5 million individuals in 2023 (MACPAC, 2024). The combination of rapid spending growth, decentralized service delivery, and consumer-directed arrangements created conditions that federal policymakers viewed as particularly vulnerable to fraud and abuse.

Home health care services constitute a distinct category within HCBS, encompassing skilled nursing, physical therapy, occupational therapy, speech-language pathology, and home health aide services delivered by certified home health agencies. While smaller in aggregate spending than PCS, HHCS involve a different workforce (licensed professionals rather than personal care attendants), different regulatory structures, and different service delivery models (agency-based rather than consumer-directed). The EVV mandate's separate deadlines for PCS and HHCS—January 1, 2020 and January 1, 2023, respectively—reflect these structural differences and provide a key source of identification for my analysis.

The Fraud and Program Integrity Motivation

The EVV mandate arose from longstanding concerns about fraud, waste, and abuse in Medicaid PCS. The decentralized nature of home care delivery—services provided in private residences, often by individual attendants with limited supervision—creates inherent monitoring challenges. Unlike nursing home services, where institutional billing systems provide some check on claims accuracy, PCS are documented primarily through paper timesheets or self-reported logs that are difficult to verify independently.

The OIG documented that PCS-related cases represented 38 percent of MFCU indictments during FY 2012–2015 and 34 percent of convictions, with an average of over 400 PCS-related convictions annually from 2015 through 2022. Common fraud schemes include billing for services never rendered ("phantom visits"), attendants clocking in without being physically present, inflated service hours, and billing for services provided by unqualified individuals. The GAO (2017) found that \$4.9 billion in PCS claims—approximately one-third of total spending—did not even identify the person providing the service, making verification of service delivery impossible. CMS has consistently identified PCS among the highest improper payment rate categories in Medicaid (CMS, 2018).

The fraud concern has intensified recently. In February 2026, CMS announced a major crackdown on Medicaid home care fraud, with particular focus on Minnesota, where CMS deferred $259.5 million in federal Medicaid funding and identified "unusually high spending" (KFF, 2026; CMS, 2026). The Minnesota action involved organized fraud rings exploiting consumer-directed PCS programs, including billing for services to beneficiaries who were out of the country—schemes that EVV's location verification should in principle detect. This enforcement action underscores both the continuing policy salience of my research question and the possibility that EVV alone may be insufficient to address certain categories of fraud.

The 21st Century Cures Act and EVV Implementation

Section 12006(a) of the Cures Act established the first federal EVV mandate, requiring electronic verification of six data elements for each home care visit: the service type, recipient, date, location, provider, and start and end times. The statute establishes escalating FMAP penalties for noncompliance (0.25–1.0 percentage points) and includes a one-year good-faith-effort (GFE) exemption. The CBO scored the provision as generating approximately $290 million in federal savings over ten years—less than 0.5 percent of annual HCBS spending—based on an assumed 1 percent payment reduction through fraud deterrence (CBO, 2016).

EVV systems typically rely on GPS-enabled mobile applications, telephony-based check-in systems, fixed verification devices, or biometric verification. States have adopted several implementation models: under the *closed model* (state-mandated), a single vendor serves all providers; under the *open model* (provider choice), providers select their own systems and the state operates a data aggregator to collect and validate visit data; MCO-directed models let managed care organizations select vendors; and *hybrid models* combine state-provided and provider-selected systems. All states must operate an EVV data aggregator that cross-references visit records against billed claims for program integrity monitoring.

CMS compliance data reveal substantial variation in actual go-live dates. For PCS, the modal go-live quarter is 2021Q1 (33 states), reflecting the GFE deadline. An additional 6 states achieved compliance later in 2021, and the remaining states staggered into compliance over 2022–2024. The earliest PCS implementer is Maryland (2020Q1); the latest is Massachusetts (2024Q4). For HHCS, 44 states were compliant by January 2024; 7 remained noncompliant (AR, GA, MA, MI, MS, ND, SC). This produces a cohort distribution heavily concentrated at 2021Q1, with progressively smaller later cohorts—a feature that constrains my staggered design.

A critical finding from state audits is that formal compliance does not correspond to effective operational implementation. The Ohio Auditor of State (2024) found that 56 percent of all PCS and HHCS paid claims were not matched to an EVV visit record, representing approximately $1.1 billion in claims outside the EVV system—despite Ohio having spent $146 million on implementation. The OIG's Kansas audit (2024) similarly identified gaps in verification procedures and exception management. These findings suggest that the "treatment" in my analysis—formal compliance—may substantially overstate the degree to which EVV is functioning as designed, motivating my intent-to-treat framing.

Existing Evidence and Contribution

The existing research on EVV impacts is remarkably thin. My systematic review confirms that no published study has employed quasi-experimental methods to estimate EVV's causal effect on spending, utilization, or access. Steckler et al. (2020) conducted interviews with 21 PCS consumers and 20 workers, finding EVV was perceived as intrusive, reducing flexibility, and creating technical difficulties. Mateescu (2021) documented how EVV creates an "atmosphere of ambient criminalization" and reported that one-third of EVV users left home less often (National Council on Independent Living survey, 2020). PCS attendant fraud convictions declined from over 400 annually during 2015–2022 to 298 in FY 2024 (OIG MFCU Annual Report), consistent with EVV deterrence but not causally attributable to EVV.

The HCBS workforce context is critical for interpretation. Approximately 2.4 million workers provide HCBS, with annual turnover rates of 40–60 percent, median hourly wages of $13.81, and chronic shortages in all 50 states (MACPAC, 2022; Scales, 2020). Herd and Moynihan (2019) predict that technology mandates impose learning, compliance, and psychological costs that disproportionately burden vulnerable populations—precisely the profile of the HCBS workforce. As of 2024, 692,679 people were on HCBS waiver waiting lists across 38 states (KFF, 2024).

My identification strategy draws on the literature exploiting staggered policy adoption for causal inference and recent econometric advances addressing heterogeneous treatment effects in staggered designs (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; de Chaisemartin and D'Haultfoeuille, 2020). The HITECH Act (2009) offers the closest health IT parallel: Adler-Milstein and Jha (2017) documented that HITECH drove large gains in hospital electronic health record adoption, though their analysis focused on hospitals rather than the community-based providers subject to EVV. The triple-difference design builds on Gruber (1994) and Olden and Moen (2022).

This paper fills four gaps. First, no causal evaluation of EVV exists despite the mandate affecting all 50 state Medicaid programs. Second, the PCS-versus-HHCS timing contrast has not been exploited for identification. Third, no prior study has decomposed EVV's effects into price, utilization, and extensive margin channels using claims-level data. Fourth, the policy moment is critical: with the February 2026 CMS fraud crackdown and congressional scrutiny of HCBS spending under the current budget reconciliation process (H.R. 1), rigorous evidence on EVV's effects is urgently needed.

Data

Data Sources

I construct a state-level panel dataset covering 50 states and the District of Columbia over the period 2018–2024 by combining four data sources: (1) a novel EVV implementation timing panel; (2) T-MSIS Medicaid claims data (primary outcomes); (3) CMS Financial Management Report expenditure data (secondary validation); and (4) state-level control variables.

EVV Implementation Timing Panel I construct the treatment variable by synthesizing information from multiple sources: CMS EVV compliance status assessments published on Medicaid.gov, state Medicaid agency bulletins documenting operational go-live dates, OIG audit reports, and EVV vendor announcements. For each jurisdiction, I code the go-live date separately for PCS and HHCS, the EVV system model (open, closed, or hybrid), and the primary vendor or data aggregator.

The resulting panel identifies substantial variation. For PCS, 33 states share the 2021Q1 modal cohort, with progressively smaller cohorts at later dates. Maryland (2020Q1) is the earliest implementer; Massachusetts (2024Q4) is the latest. I distinguish between formal CMS compliance dates (self-reported by states) and operational go-live dates where these differ, coding the formal compliance date as my baseline treatment variable with sensitivity analyses using alternative timing definitions.

T-MSIS Medicaid Claims Data (Primary Outcomes) My primary outcomes are drawn from the HHS Medicaid Provider Spending dataset, which contains provider-procedure-month Medicaid claims from the Transformed Medicaid Statistical Information System (T-MSIS). The dataset covers January 2018 through December 2024, containing approximately 227 million rows.

I filter to 13 HCPCS codes corresponding to EVV-relevant services. For PCS: T1019 (personal care, per 15 min; 457,916 claims), S5125 (attendant care, per 15 min; 214,380), S5130 (homemaker, per 15 min; 125,420), T1020 (personal care, per diem; 39,648), S5131 (homemaker, per diem; 2,528), and S5126 (supportive home care, per diem; 1,905). For HHCS: G0299 (direct nursing by RN; 144,999), G0300 (direct nursing by LPN; 62,805), G0151 (physical therapy; 62,759), G0156 (home health aide; 52,701), G0152 (occupational therapy;

28,520), T1021 (home health aide/CNA; 14,627), and G0153 (speech-language pathology; 9,785). This reduces the dataset to 1,217,993 EVV-relevant claims.

I construct an NPI-to-state crosswalk to assign each claim to a state. The raw claims data identify providers by National Provider Identifier (NPI) but do not include a state field. I obtain state assignments from the NPPES companion tables, using the servicing provider NPI as the preferred identifier (reflecting where services are delivered) with the billing provider NPI as a fallback. For providers appearing in multiple states, I assign the modal state. The rerun crosswalk contains 1,762,665 NPIs after adding 170,225 billing-only fallback NPIs, and the claim-level state-assignment rate rises to 98.5 percent (1,199,784 of 1,217,993 EVV-relevant claims). Billing-only fallback assigns 317,983 claims (26.1 percent), while 18,209 claims remain unmatched.

I aggregate the matched claims from provider-procedure-month to the state-by-quarter level. Quarterly values are computed as sums of monthly values for additive variables (spending, claims, beneficiaries) and averages for rate variables (provider counts). The resulting outcome variables include: total Medicaid spending (separately for PCS, HHCS, and combined HCBS), claims volume, beneficiary counts (upper bounds due to potential double-counting across providers within a state-quarter), average monthly unique provider counts, dollars per claim (spending divided by claims), and claims per provider (claims divided by average monthly provider count).

The claims data have four key advantages over FMR data. First, they provide genuine quarterly variation—monthly data aggregated to quarters—rather than annual data divided by four, enabling credible within-year event studies. Second, HCPCS-level filtering isolates exactly the services subject to EVV. Third, the decomposition of spending into price, quantity, and extensive margin channels enables mechanistic analysis. Fourth, the higher statistical precision increases power.

The claims data also have known limitations. CMS T-MSIS managed care data quality varies substantially across states; KFF (2023) identifies 6 states with "unusable" and 16 with "high concern" managed care encounter data. I test sensitivity to excluding these states. CMS suppresses cells with fewer than 12 beneficiaries, creating non-random missingness primarily affecting small states and rare HHCS codes. Beneficiary counts are upper bounds due to potential double-counting across providers. The panel covers 1,379 of 1,632 possible PCS state-quarter observations (84.5 percent) and 1,262 of 1,632 HHCS observations (77.3 percent).

CMS Financial Management Report Data (Secondary Validation) I use FMR data derived from the CMS-64 quarterly statement of expenditures that each state submits to CMS. While the CMS-64 is submitted quarterly, the publicly available FMR data report state-level Medicaid expenditures by service category at annual frequency, aggregated by federal fiscal year (FY2018–

FY2024). I extract personal care services regular payments (Row 72), self-directed PCS under Section 1915(j) (Row 73), home health services (Row 37), and Community First Choice under Section 1915(k). I construct total PCS spending as the sum of regular and self-directed personal care, and total HCBS as PCS plus home health plus Community First Choice. For quarterly comparison, I divide annual values by four, noting this eliminates within-year variation—within any fiscal year, all four "quarterly" FMR values are identical. Annual FMR models (N=408 state-years) align with the native data frequency; quarterly FMR models (N=1,428) are supplementary. Negative values reflecting CMS-64 adjustments and states reporting zero state-plan PCS spending (due to waiver or managed care delivery) are retained.

Control Variables I construct the following state-level controls. The *ARPA Section 9817 enhanced FMAP* indicator equals one for quarters overlapping with the American Rescue Plan Act Section 9817 period (April 2021–March 2022), during which all states received a 10 percentage point increase in the federal match for HCBS expenditures. This confound is important because the ARPA period directly coincides with the peak EVV rollout window. Quarter fixed effects absorb the common ARPA effect, but cross-state variation in ARPA-induced spending changes correlated with EVV timing could bias estimates. I also include the *COVID-19 PHE* indicator (January 2020–May 2023), *Medicaid expansion* status, *state population* (2020 Census), and *total Medicaid enrollment* (MBES).

Analysis Panel Construction

The final analysis panel merges the EVV policy panel, claims outcomes, FMR outcomes, and control variables at the state-by-quarter level. The panel contains 51 jurisdictions observed over 32 quarters (2017Q1–2024Q4), yielding 1,632 possible state-quarter observations. Due to claims suppression and missing data, nonmissing observations vary by outcome: 1,379 state-quarters for PCS claims (84.5 percent coverage), 1,262 for HHCS claims (77.3 percent), and 1,428 for FMR outcomes (87.5 percent).

Descriptive Statistics

Table 1 reports summary statistics for the analysis panel. Several features merit discussion.

Spending distributions. Mean quarterly PCS claims spending is $123.7 million per state-quarter, but the median is only $28.3 million, reflecting substantial right skew driven by large states (New York, California, Texas). The standard deviation ($441.2M) exceeds the mean by a factor of nearly four. Mean quarterly HHCS claims spending is much smaller at $5.6 million (median $1.0M), reflecting both the smaller HHCS program size and the later stage of HHCS EVV implementation during my sample period.

Price and utilization. Mean PCS dollars per claim is \$150.39 (median \$101.22), while mean HHCS dollars per claim is \$71.18 (median \$66.29). The higher PCS price reflects the mix of per-15-minute and per-diem billing codes. States have an average of 115.4 unique PCS providers per month (median 61.7) and 39.6 HHCS providers (median 16.3), with considerable variation in market size.

Treatment distribution. PCS EVV is active in 43.5 percent of state-quarter observations (710/1,632), HHCS EVV in 13.4 percent (218/1,632), and 30.1 percent (492/1,632) fall in the PCS-only window for the triple-difference design.

Balance. Early PCS implementers (34 states by 2021Q1) and late implementers (17 states) differ in pre-treatment characteristics. Late implementers tend to have higher PCS spending (mean \$81.2M vs. \$29.0M quarterly, normalized difference = -0.38), consistent with larger PCS programs facing more complex implementation challenges. Population and HHCS spending are more balanced (normalized differences of -0.04 and -0.07). These imbalances are absorbed by state fixed effects but motivate caution in heterogeneity analyses comparing early versus late implementers.

Methods

Staggered Difference-in-Differences

My primary specification is a two-way fixed effects (TWFE) model:

$$Y_{st} = \alpha_s + \gamma_t + \beta \cdot \mathrm{EVV}_{st} + \varepsilon_{st}$$

where Y_{st} is the outcome for state s in quarter t, α_s are state fixed effects, γ_t are quarter fixed effects, and EVV_{st} indicates whether state s has implemented EVV for the relevant service type by quarter t. The coefficient β estimates the average treatment effect on the treated. Standard errors are clustered at the state level (51 clusters for PCS, 50 for HHCS).

I estimate at two temporal resolutions. The primary specification uses quarterly claims data (N=1,379 for PCS, N=1,261 for HHCS). Secondary models use annualized claims data (N=408) and FMR data. Because 33 of 51 states share the 2021Q1 cohort, limiting identifying variation, I report a Goodman-Bacon (2021) decomposition. Quarter fixed effects absorb national-level confounds including the ARPA enhanced FMAP and the COVID-19 PHE.

Heterogeneity-Robust Estimation Because standard TWFE can produce biased estimates when treatment effects are heterogeneous across cohorts and time (de Chaisemartin and D'Haultfoeuille, 2020; Goodman-Bacon, 2021), I implement the Gardner (2022) DID2S estimator using annualized data. In the first stage, state and time fixed effects are estimated using only untreated observations; always-treated states are excluded. In the second stage, residualized outcomes are regressed on the treatment indicator with cluster-robust standard errors.

Event Study I estimate event study models to assess pre-trends and characterize the dynamic treatment path:

$$Y_{st} = \alpha_s + \gamma_t + \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[\text{EventTime}_{st} = k] + \varepsilon_{st}$$

where EventTime_{st} is quarters since state s's go-live date, $k = -1$ is the reference period, and event times are binned at -8 and $+12$ quarters. The claims data provide genuine quarterly observations, so event study coefficients reflect actual within-year dynamics.

Triple-Difference Design

The PCS-HHCS timing gap enables a supplementary DDD specification (Gruber, 1994; Olden and Moen, 2022):

$$Y_{sjt} = \alpha_{sj} + \gamma_{jt} + \delta \cdot \text{DDD}_{sjt} + \varepsilon_{sjt}$$

where $j \in \{\text{PCS}, \text{HHCS}\}$ indexes service type, α_{sj} are state-by-service fixed effects, and γ_{jt} are time-by-service fixed effects. The DDD indicator takes the value 1 for PCS observations in the PCS-only treatment window. Standard errors are clustered at the state level. The DDD requires only that no contemporaneous shock differentially affects PCS relative to HHCS spending within the same state—a weaker assumption than standard DiD.

Mechanism Decomposition

A key contribution enabled by the claims data is the decomposition of aggregate spending:

$$\text{Total Spending} = \underbrace{(\$/\text{claim})}_{\text{Price}} \times \underbrace{(\text{claims/provider})}_{\text{Intensity}} \times \underbrace{(\text{providers})}_{\text{Extensive margin}}$$

I estimate the TWFE model separately for each component plus beneficiary counts. A reduction in dollars per claim would be consistent with EVV deterring inflated charges; fewer claims per provider would indicate deterrence of phantom visits; fewer providers would suggest fraudulent provider exit. This decomposition addresses a fundamental limitation of aggregate analyses: a null spending effect could mask offsetting changes across margins. The decomposition is descriptive rather than structural.

Outcomes

My primary outcomes are quarterly claims-based measures at the state-quarter level. Spending outcomes include: PCS total paid (in millions of dollars), HHCS total paid, and total HCBS paid (PCS + HHCS). Volume outcomes include PCS and HHCS claims counts. Mechanism outcomes include: dollars per claim (PCS and HHCS separately), claims per provider (PCS and HHCS), average monthly unique provider count (PCS and HHCS), and upper-bound beneficiary counts (PCS and HHCS).

Secondary outcomes are FMR-based: PCS spending, HHCS spending, total HCBS spending, PCS share of HCBS, and normalized variants (per capita and per enrollee). The FMR outcomes serve as validation of the claims-based results using an independent data source with different scope, measurement, and temporal resolution.

Estimand and Intent-to-Treat Interpretation

My treatment variable captures formal EVV compliance, not effective operational implementation. Given evidence of a substantial compliance-operation gap (Ohio: 56% of claims unmatched post-compliance), I interpret all estimates as intent-to-treat (ITT) effects. If one assumes a 44 percent effective compliance rate (based on Ohio), the implied TOT is approximately 2.3 times the ITT. For PCS claims spending, this yields approximately -$4.1M per state-quarter—still statistically insignificant. This Wald-type calculation requires the assumption that EVV has no effect on non-verified claims, which may be violated if the mandate's existence deters fraud among non-compliant providers.

Power Analysis

I calculate the minimum detectable effect (MDE) at 80 percent power and 5 percent significance: MDE = $(1.96 + 0.84) \times$ SE = $2.80 \times$ SE. I compare MDEs to CBO projections converted to total computable amounts at the average 57 percent FMAP (≈$0.25M per state-quarter, ≈$1.0M per state-year).

Robustness and Sensitivity

I conduct an extensive battery of robustness checks: (1) pre-trends joint Wald tests on pre-treatment event study coefficients; (2) randomization inference with 500 permutations; (3) ARPA Section 9817 sensitivity across four specifications; (4) T-MSIS data quality sensitivity, progressively excluding states with poor managed care data; (5) exclusion of 7 HHCS-noncompliant states; (6) go-live date shifts ($±$1 quarter); (7) early-versus-late and open-versus-closed heterogeneity (exploratory); (8) Oster (2019) bounds; (9) FMR validation at quarterly and annual frequency; and (10) early-implementer cutoff sensitivity.

Identifying Assumptions and Threats

The key assumption is that PCS spending trends would have been parallel across early- and late-implementing states absent EVV, supported by the clean pre-trends test (Wald p=0.875). Principal threats include: correlated state policy changes contemporaneous with EVV adoption; ARPA Section 9817 overlap with the modal 2021Q1 cohort (while quarter FE absorb the common effect, cross-state differential ARPA responses correlated with EVV timing could bias estimates); treatment timing endogeneity (mitigated by vendor-driven timing variation and the federal deadline structure); and limited staggered variation given the concentrated 2021Q1 cohort. All analyses are implemented in Python 3.13 using pyfixest for high-dimensional fixed-effects estimation.

Results

I begin with the raw PCS spending pattern, treatment-timing support, and event-study evidence so the main null estimate is read in the context of the underlying claims data rather than as a stand-alone coefficient table.

Main Estimates

Table 2 presents the primary difference-in-differences estimates. Across all outcomes and specifications, I find no statistically significant effect on PCS or total HCBS spending.

Quarterly claims TWFE (primary specification). For PCS claims spending, the TWFE estimate with state and quarter fixed effects is +$0.62 million per state-quarter (SE=$18.61, p=0.974, N=1,379). The 95 percent confidence interval ranges from approximately -$35.9M to +$37.1M. The point estimate represents 0.5 percent of the mean quarterly PCS claims spending of $123.7 million and is bounded near zero relative to the observed mean, but it is not precise enough to adjudicate the CBO-projected savings range: the standard error of $18.61 million implies that the design rules out only large effects. Adding explicit controls for ARPA Section 9817, the COVID-19 PHE, and Medicaid expansion yields a similarly null estimate (+$3.10M, SE=$20.25, p=0.879). The insensitivity of the estimate to the inclusion of explicit time-varying controls confirms that quarter fixed effects adequately absorb the common effects of the ARPA enhanced FMAP and the COVID-19 PHE.

For HHCS claims spending, the quarterly TWFE estimate is +$5.18 million (SE=$3.09, p=0.100, N=1,261) under the observed-cells primary specification. This is the largest point estimate in percentage terms—representing 93 percent of mean quarterly HHCS spending—and is marginally insignificant at conventional levels. However, HHCS pre-treatment trends are rejected (Wald p=0.0006), meaning the parallel trends assumption is violated for HHCS. I therefore do not interpret the HHCS coefficient as a causal effect of EVV on HHCS spending, and I treat HHCS throughout as a secondary mechanism check rather than a co-equal primary outcome alongside PCS.

HHCS sensitivity to the missing-data convention. A second reason for caution on HHCS is that the result is sensitive to how unobserved state-quarter cells are treated. My primary specification uses only observed claims cells—state-quarters in which T-MSIS records claims for the HHCS HCPCS codes—and drops cells suppressed for low claim counts or absent from the dataset. Under an alternative convention that zero-fills these unobserved cells during the 2018–2024 T-MSIS era (treating absence of records as zero claims), the quarterly HHCS TWFE estimate shifts to +\$4.55M (SE=\$2.67, p=0.094, N=1,428). The annual HHCS estimate under the primary observed-cells convention is +\$24.26M per state-year (SE=\$13.71, p=0.083, N=325). This is a measurement-policy issue, not a coding bug or post-hoc specification search: the analytic registry locks the observed-cells convention as primary and discloses the zero-fill alternative as a sensitivity. The PCS quarterly headline (the primary outcome) is *not* sensitive to this convention. Together with the rejected HHCS pre-trends, the missing-data sensitivity is the second independent reason to read HHCS as a secondary mechanism signal rather than evidence of a causal HHCS effect.

The remaining discussion therefore focuses on PCS outcomes, where the design is well-identified, well-powered for large effects, and not sensitive to missing-data conventions.

For total HCBS claims spending (PCS + HHCS), the TWFE estimate is +\$2.37 million (SE=\$17.54, p=0.893), effectively zero. Claims volume outcomes mirror the spending null. PCS claims counts show an insignificant positive estimate of +140,458 claims per state-quarter (SE=227,567, p=0.540), representing an 11.7 percent increase relative to the mean of 1,199,217 claims.

Annual claims models. Annual results yield qualitatively identical conclusions for PCS and total HCBS spending. PCS: +\$9.39M per state-year (SE=\$69.83, p=0.894, N=347). HHCS: +\$24.26M (SE=\$13.71, p=0.083, N=325). Total HCBS: +\$13.23M (SE=\$66.91, p=0.844, N=353). Annual HHCS claims counts are significant (+283,162, p=0.044), but I do not elevate HHCS because the spending estimate remains imprecise and HHCS pre-trends fail. Annual aggregations use sum(min_count=1) and drop all-missing state-years to avoid an earlier zero-fill artifact in the 2017 cell; the HHCS annual coefficient is the observed-cells primary referenced in the abstract.

Mechanism Decomposition

Table 3 presents the mechanism decomposition—a key contribution enabled by the claims data. No channel shows a statistically significant effect, but the informativeness of these nulls differs substantially across margins.

Price (dollars per claim). PCS: -\$36.71 (SE=\$24.77, p=0.145), a 24 percent reduction relative to the mean of \$150.42. While the largest point estimate in the decomposition, it is not statistically significant. HHCS: +\$0.01 (p=0.999).

Utilization intensity (claims per provider). PCS: +2,870 (SE=2,650, p=0.284), a 25 percent increase relative to the mean of 11,586. HHCS: -272 (p=0.831).

Extensive margin (provider count). PCS: -1.2 (SE=4.4, p=0.780), a 1 percent reduction relative to the mean of 115 providers. HHCS: +1.5 (p=0.810). No evidence of EVV-induced provider exit.

Beneficiaries. PCS: +4,058 (SE=8,011, p=0.615). HHCS: +3,928 (p=0.244). Neither significant.

The absence of statistically significant effects across all four margins offers no evidence of large offsetting changes across channels, but the less precise price and intensity estimates cannot rule out smaller countervailing movements. The provider-count result is the most informative margin, while the other channel tests are better interpreted as bounded nulls than definitive mechanism exclusions.

DID2S Estimates

Annual DID2S: PCS +\$33.28M (SE=\$111.33, p=0.766), HHCS +\$33.37M (SE=\$7.99, p<0.001), total HCBS +\$43.68M (p=0.688). The significant HHCS DID2S result should be interpreted with extreme caution given the rejected HHCS pre-trends and the HHCS missing-data sensitivity documented elsewhere; the PCS DID2S null is the load-bearing finding here.

Power Analysis and Pre-Trends

The MDE at 80 percent power is \$52.14M per state-quarter for PCS (55.6% of baseline spending). Converting the CBO projection: \$290M federal savings over 10 years at 57% FMAP yields ≈\$509M total computable, or ≈\$1.0M per state-year (\$0.25M per state-quarter). The quarterly MDE is approximately 209 times the CBO projection. The FMR-based annual MDE is \$126.93M per state-year (52.1% of baseline, 127 times CBO). The study is substantially underpowered for CBO-projected savings regardless of data source.

PCS claims pre-trends are cleanly not rejected (Wald chi-squared=3.10, df=7, p=0.875), with no individual pre-treatment coefficient significant at the 5 percent level. HHCS pre-trends are clearly rejected (Wald=25.66, df=7, p=0.0006), with individual coefficients significant. The PCS event study (Figure 3), presented in log scale so that coefficients approximate percent-change effects, shows pre-treatment coefficients tightly centered around zero and post-treatment coefficients averaging -3.1 percent with no individual coefficient significant at the 5 percent level. The null is consistent across alternative outcome scales including levels, per capita, and percentage of pre-treatment baseline (Appendix Figures A7–A10).

Raw Outcome Trends and Treatment Timing

Figure 1 presents raw cohort-specific mean spending trends for PCS and HHCS over calendar time, with early, middle, and late EVV adoption cohorts plotted separately. The raw data patterns are informative: PCS spending trends are broadly parallel across cohorts in the pre-treatment period, consistent with the clean Wald test (p=0.875). HHCS spending shows more heterogeneous pre-treatment trajectories, foreshadowing the pre-trend rejection. No sharp trend break at median go-live dates is visible in either panel, consistent with the null TWFE estimates.

Figure 2 displays the distribution of EVV implementation timing. Panel (a) shows PCS go-live cohort sizes, with the dominant 2021Q1 cohort (33 states) clearly visible alongside the federal deadline annotations. Panel (b) plots the share of states with active EVV over time for both PCS and HHCS. Together, the panels illustrate the limited staggered variation that constrains the design: the majority of identifying variation comes from a handful of late-adopting states.

Goodman-Bacon Decomposition

The Goodman-Bacon decomposition reveals that the 2021-vs-2022 comparison (39 early-treated states versus 6 later-treated states) receives approximately 45 percent of the total TWFE weight. Comparisons involving the 2021 cohort collectively account for over 80 percent. The substantial heterogeneity in estimates across comparisons is visible in the scatter plot (Appendix Figure A6), though the weighted average is near zero. Because the modal cohort dominates, the "staggered" design effectively reduces to a mostly-simultaneous adoption study with limited timing variation from late adopters. (Note: decomposition weights are approximate, based on cohort sizes rather than the exact variance-weighted formula in Goodman-Bacon [2021].)

Triple-Difference and Supplementary Tests

Table 4: DDD using the PCS-only window indicator: +$5.88M (SE=$17.08, p=0.732, N=2,472). Service-specific EVV: +$3.21M (SE=$11.71, p=0.785). Both effectively zero, consistent with the main DiD.

Randomization inference (500 permutations) yields a p-value of 0.966 for PCS claims spending, indicating the observed estimate is well within the range expected under random treatment assignment (Figure 6).

Robustness Checks

Table 5 summarizes the robustness battery (detailed results in Appendix Tables A5–A12).

T-MSIS data quality. Results are stable when excluding states with poor managed care data. Excluding 6 "unusable" states: +$0.23M (p=0.991). Excluding 16 “high concern” states: -\$5.84M (p=0.817). Excluding all 22: -\$1.49M (p=0.964). The null is not driven by measurement error in states with poor data quality.

ARPA sensitivity. PCS estimates are stable across specifications: -\$1.80M (baseline), +\$0.97M (with ARPA), +\$1.21M (full controls), reflecting that quarter FE absorb the common ARPA effect.

Go-live date sensitivity. Robust to \$±\$1 quarter shifts: -\$0.55M (p=0.972) and +\$4.41M (p=0.835).

Sample restrictions. Excluding 7 HHCS-noncompliant states: PCS +\$5.33M (p=0.682), HHCS +\$0.55M (p=0.734). Both remain insignificant.

EVV model type. Open (41 states): -\$6.55M (p=0.767). Closed (7 states): +\$48.07M (p=0.348). Neither significant.

Heterogeneity. Early implementers (34 states, by 2021Q1): +\$17.54M (SE=\$7.30, p=0.022). Late (17 states): -\$35.34M (SE=\$18.01, p=0.067). The early-implementer result is positive—opposite the expected fraud-reduction direction—and is sensitive to the cutoff definition (p=0.036 at 2021Q1 cutoff; p=0.236 at 2021Q2). This finding is exploratory and should not be treated as robust.

FMR validation. Quarterly FMR: PCS +\$0.91M (p=0.935), HHCS +\$2.21M (p=0.572). Annual FMR: PCS +\$8.72M (p=0.848), HHCS +\$5.87M (p=0.721). The convergence of null results across two independent data sources strengthens the conclusion.

Discussion

Summary of Findings

This study provides the first quasi-experimental evaluation, to my knowledge, of the federal EVV mandate’s effect on Medicaid home care spending, utilization, and provider participation. Using T-MSIS claims data as primary evidence and FMR data as secondary validation, I exploit staggered EVV implementation timing across 51 jurisdictions. My central finding is a comprehensive null: EVV compliance has not produced large aggregate spending reductions across any margin examined. The quarterly PCS TWFE estimate is +\$0.62M (p=0.974), with clean pre-trends (Wald p=0.875), randomization inference confirmation (p=0.966), and corroboration from the triple-difference design, DID2S estimator, and FMR data. However, the MDE is 209 times the CBO projection, so small effects in the projected range cannot be ruled out.

Interpreting the Null: Four Scenarios

The null is consistent with four non-mutually-exclusive scenarios.

Scenario A: Genuine ineffectiveness. EVV may not deter fraud at meaningful scale, either because fraud prevalence in the verified categories is lower than assumed by policymakers, or because the types of fraud EVV can detect are a small share of total improper payments. The bounded mechanism results are consistent with this interpretation, especially the informative provider-count null, though the price and utilization channels remain too imprecise to distinguish it cleanly from the other scenarios. The GAO's (2017) finding that $4.9 billion in PCS claims lacked provider identification suggests that many "improper payments" may reflect documentation failures rather than intentional fraud—and EVV may merely formalize existing care arrangements. The policy implication would be to evaluate EVV on dimensions other than spending reduction.

Scenario B: Savings offset by documentation improvements. EVV may simultaneously reduce fraudulent billing and increase *measured* spending as providers formalize previously informal care arrangements and submit more complete claims. Under this scenario, EVV's true fraud-deterrence effect is positive (reducing spending) but is masked in aggregate data by a countervailing documentation effect (increasing measured spending). The suggestive patterns in the mechanism decomposition are consistent: the negative PCS price effect (-$36.71/claim, p=0.145) could reflect deterrence of inflated per-visit charges, while the positive utilization effect (+2,870 claims/provider, p=0.284) could reflect increased claim submission as EVV infrastructure facilitates more complete documentation. The early-implementer finding (+$17.54M, p=0.022) is also consistent, as documentation improvements may be most pronounced during initial implementation. However, neither component is individually significant, and the early-implementer result is sensitive to the subgroup cutoff, so this interpretation remains speculative.

Scenario C: Effect too small for aggregate data. The CBO projected savings of approximately $0.25M per state-quarter in total computable terms, which is less than 0.3 percent of mean baseline PCS claims spending ($123.7M) and less than 0.06 percent of one standard deviation of spending ($441.2M). This effect is approximately 209 times smaller than my MDE. If EVV's true effect is in the range projected by the CBO, no study using aggregate state-level data—whether claims-based or FMR-based, quarterly or annual—can detect it. Individual-level claims analysis with beneficiary and provider identifiers, using within-provider variation in EVV adoption, would be needed. This observation has implications beyond my study: CBO projections of modest savings from fraud-deterrence mandates may be fundamentally untestable using the aggregate data sources available to researchers. The lesson for future policy evaluation is that if savings projections are in the range of 1 percent of total spending, evaluation designs must be built into the policy at the outset with individual-level data collection planned prospectively.

Scenario D: Incomplete implementation. Ohio found 56 percent of claims outside the EVV system post-compliance; Kansas had significant verification procedure gaps. If only 44 percent of claims are verified, the effective treatment dose is less than half of what the CBO assumed. My Wald-type TOT ($\beta_{\text{ITT}}/0.44 \approx \$1.4$M) remains insignificant.

Mechanism Decomposition: Added Value

The mechanism decomposition strengthens the bounded-null interpretation. If EVV were operating as designed, I would expect effects on at least one channel. The suggestive PCS price effect (-$36.71/claim, 24% of mean) is in the expected direction and warrants investigation at the individual claim level, where it could reflect genuine price reductions or compositional shifts between billing codes. The positive utilization effect is inconsistent with phantom-visit deterrence but consistent with documentation improvement. The null provider-count result (-1.2, p=0.780) provides the clearest reassurance against concerns that EVV would drive providers from the market (Mateescu, 2021), and the null beneficiary result (+4,058, p=0.615) argues only against larger contractions in aggregate service reach. The absence of statistically significant channel effects therefore supports the claim that EVV has not produced large mechanism shifts, while leaving smaller offsetting changes unresolved.

HHCS as a Secondary Mechanism Check: Pre-Trends and Missing-Data Fragility

I retain HHCS in the main text as a secondary mechanism check, but I want to be explicit about why the HHCS results should not be read as evidence of a causal EVV effect.

First, the rejection of HHCS pre-trends (Wald p=0.0006) indicates that HHCS claims spending was on a differential trajectory across early and late EVV implementing states before EVV was implemented. Several factors may contribute: structural changes in the HHCS market during my study period (Medicare Advantage home health growth, the PDGM payment reform effective January 2020), T-MSIS data quality differences that may be more severe for HHCS, and the smaller HHCS claims volume making the outcome more sensitive to outliers.

Second, the HHCS estimate is sensitive to the missing-data convention. The observed-cells primary specification yields +$24.26M per state-year (p=0.083) at annual frequency and +$5.18M per state-quarter (p=0.100) at quarterly frequency—both directionally positive but imprecise. Under an alternative zero-fill convention, the quarterly estimate becomes +$4.55M (p=0.094). This is a measurement-policy issue: the two conventions answer slightly different questions (what is the average effect on observed HHCS billing? versus what is the average effect on observed-or-implied-zero HHCS billing?), and reasonable analysts can disagree on which is the more appropriate primary. The sensitivity of the HHCS estimate to this choice—combined with the rejected pre-trends—means I do not interpret HHCS results as causal. The PCS quarterly headline (the primary outcome) is not sensitive to this convention; the missing-data sensitivity is HHCS-specific.

Taken together, the two HHCS problems—differential pre-trends and missing-data sensitivity—are independent and reinforcing reasons to treat the HHCS coefficient as a measurement-policy-sensitive mechanism signal rather than a causal estimate. Readers attracted by the marginal HHCS p-values or the significant annual count/DID2S diagnostics should weigh both caveats before drawing inferences from the sign or magnitude.

Relation to Prior Literature

My findings speak to several strands of the existing literature.

Worker and consumer welfare. Steckler et al. (2020) and Mateescu (2021) documented concerns that EVV would harm workers and reduce service flexibility. Mateescu specifically warned of "ambient criminalization" and potential for EVV to drive workers from the Medicaid home care market. My null provider-count result (PCS: -1.2, p=0.780) and null beneficiary result (+4,058, p=0.615) provide the first quantitative evidence bearing on these concerns. At the aggregate level, EVV has not caused large-scale provider exit or service contraction. However, aggregate provider counts may mask compositional changes (e.g., smaller providers exiting while larger agencies grow), and my data cannot capture the worker-level effects documented in the qualitative literature.

Fraud deterrence. The descriptive decline in PCS attendant fraud convictions—from over 400 annually during 2015–2022 to 298 in FY 2024—is consistent with EVV having a deterrent effect. My aggregate spending null does not contradict this finding: if fraud represents a small share of total PCS spending (the CBO assumed only a 1 percent reduction), its elimination would generate savings too small for aggregate detection. The fraud-deterrence and aggregate-null findings are reconcilable under Scenario C.

Health IT mandates. My null parallels mixed evidence on other health IT mandates. The HITECH Act's EHR incentives produced large gains in hospital electronic health record adoption (Adler-Milstein and Jha, 2017), though evidence on downstream spending or quality effects has been inconsistent; Agha (2014) found that HIT adoption was associated with a 1.3 percent *increase* in billed charges with no evidence of cost savings even five years after adoption. Electronic prescribing mandates reduced errors but had limited aggregate pharmaceutical spending effects. The pattern across health IT mandates may be that technology adoption generates operational improvements that are real but too diffuse to produce measurable aggregate effects in the short to medium term.

Administrative burden. Herd and Moynihan's (2019) framework predicts that technology mandates impose learning, compliance, and psychological costs disproportionately burdening vulnerable populations. My null extensive-margin result suggests these burdens have not caused aggregate provider exit, but the framework's predictions about worker welfare and care quality operate below the level my state-aggregate data can observe. The National Council on Independent Living finding that one-third of EVV users leave home less often is a concerning signal about beneficiary welfare that my spending and utilization measures cannot capture.

Policy Implications

My findings carry several implications for policymakers, though they must be interpreted with appropriate humility given power limitations.

The CBO scoring and federal budget policy. The CBO scored the EVV mandate as generating $290 million in federal savings over ten years. My inability to detect aggregate spending effects, combined with the power analysis showing my MDE is 209 times the CBO projection, means I can neither confirm nor reject this projection. Policymakers should be aware that the CBO projection rested on assumptions—particularly about fraud prevalence and EVV operational effectiveness—that subsequent implementation experience has called into question. In the context of the current budget reconciliation process (H.R. 1), proposals that rely on EVV-style monitoring mandates as CBO-scorable savings vehicles should be scrutinized carefully: the assumption that electronic verification translates mechanically into spending reductions is not supported by my aggregate evidence, and the projected savings may be too small to verify ex post.

CMS enforcement and FMAP penalties. Seven states remain noncompliant with the HHCS EVV requirement and face escalating FMAP reductions. My finding that EVV compliance has not produced detectable spending changes complicates the fiscal justification for these penalties. If the primary rationale is fiscal protection—ensuring that noncompliant states do not burden the federal government with higher improper payments—my evidence provides no support. However, if the rationale is structural—establishing EVV as infrastructure for program integrity and future enforcement capacity—my evidence is neutral, as I cannot measure these non-spending benefits.

The February 2026 CMS fraud crackdown. CMS's fraud crackdown, including the $259.5 million federal funding deferral in Minnesota, underscores that EVV alone has not solved the fraud problem in HCBS. Minnesota's case involved schemes that EVV's location verification should in principle detect, yet large-scale fraud persisted. This suggests either that EVV systems are not functioning effectively, that exception-management processes are too permissive, or that fraud schemes exploit gaps in EVV coverage (e.g., managed care encounters outside the EVV system). My results are consistent with EVV being necessary but not sufficient infrastructure for fraud prevention, requiring complementary enforcement tools including data analytics, provider enrollment screening, and targeted audits.

State EVV operations. The Ohio Auditor's finding that $1.1 billion in claims lacked EVV matching despite formal compliance, combined with my null results, suggests that states should reconsider EVV's operational model. Tying EVV data submission to claims payment—making verification a condition for reimbursement rather than a parallel monitoring system—could substantially increase the effective treatment dose. Few states have taken this step, likely due to concerns about disrupting service delivery.

Reassurance on access. The null provider-count and beneficiary-count results provide modest reassurance to disability advocates that EVV has not caused large-scale service contraction (Mateescu, 2021). The HCBS workforce crisis—with turnover rates of 40–60 percent and chronic shortages in all 50 states—has not been measurably worsened by EVV at the aggregate level. However, aggregate stability can mask distributional consequences, and the qualitative evidence of worker burden and beneficiary mobility restrictions should not be dismissed simply because aggregate utilization has not declined.

Limitations

First, the study is substantially underpowered for the CBO-projected effect size. The MDE of $52.14M per state-quarter exceeds the CBO projection by a factor of 209. My null cannot be interpreted as evidence of no effect—only that large effects have not occurred.

Second, T-MSIS managed care data quality varies across states (6 "unusable," 16 "high concern" per KFF 2023). Results are robust to excluding these states, but residual measurement error may attenuate estimates. If error is correlated with EVV timing, the bias direction is ambiguous.

Third, my NPI-to-state crosswalk assigns 98.5 percent of EVV-relevant claims after billing-provider fallback, but provider-state assignment remains an imperfect proxy for service-recipient state among multi-state organizations.

Fourth, HHCS pre-trends are rejected ($p=0.0006$), undermining causal interpretation of all HHCS-specific results. The HHCS estimate is also sensitive to the missing-data convention: under the observed-cells primary specification the quarterly TWFE is +$5.18M ($p=0.100$) and the annual TWFE is +$24.26M ($p=0.083$), while under a zero-fill convention the quarterly TWFE shifts to +$4.55M ($p=0.094$). I treat HHCS as a secondary mechanism check rather than a co-equal primary outcome for this reason. The PCS quarterly headline is not sensitive to this convention.

Fifth, my treatment variable captures formal compliance, not operational EVV use. The compliance-operation gap means I estimate ITT effects that may understate the effect of fully operational EVV.

Sixth, 33 of 51 states share the 2021Q1 cohort, severely limiting staggered design variation. The Goodman-Bacon decomposition confirms the 2021-vs-2022 comparison receives 45 percent of the TWFE weight.

Seventh, the ARPA Section 9817 enhanced FMAP directly coincides with the modal EVV treatment cohort. While estimates are stable across ARPA specifications, any cross-state differential response correlated with EVV timing could bias estimates.

Eighth, I cannot observe service hours, units of service, or authorized care plans, limiting my ability to distinguish billing-unit changes from changes in care duration.

Future Research

This study motivates several directions for future work.

First, and most importantly, individual-level claims analysis using T-MSIS/TAF research files with beneficiary and provider identifiers would permit examination of per-beneficiary utilization patterns, precise provider entry and exit, within-provider billing changes before and after EVV, and effects on specific fraud-vulnerable populations (e.g., self-directed service recipients, high-utilization beneficiaries). Such analysis could detect effects invisible in aggregate data—for example, a reduction in claims from the top 1 percent of billers paired with an increase from the bottom 50 percent. Access to TAF research files requires a CMS data use agreement, which limits the research community's ability to pursue this work.

Second, the suggestive PCS price effect (-$36.60/claim, p=0.146, a 24 percent reduction relative to mean) warrants investigation at the individual claim level. Claim-level analysis could distinguish between compositional changes (shifts from per-diem to per-15-minute codes) and genuine price reductions within billing codes.

Third, the February 2026 CMS enforcement actions create a new source of variation for evaluating the effect of enforcement intensity conditional on EVV infrastructure. A difference-in-differences comparing outcomes in states subject to enforcement actions versus similar states with EVV but without enforcement could isolate the marginal contribution of enforcement beyond infrastructure.

Fourth, a formal cost-benefit analysis comparing EVV implementation costs against all documented benefits is needed. Ohio alone spent $146 million. If similar expenditures occurred across all 50 states, total costs could exceed $3–5 billion nationwide, against CBO-projected federal savings of $290 million. Even accounting for non-monetary benefits (program integrity infrastructure, data quality improvements), the cost-benefit calculation may not favor the mandate unless implementation costs are treated as long-term infrastructure investment.

Fifth, qualitative research expanding the work of Steckler et al. (2020) and Mateescu (2021) with larger, more representative samples and longitudinal follow-up could illuminate EVV's effects on workforce retention, beneficiary mobility, care flexibility, and psychological well-being—outcomes that aggregate spending data cannot capture.

Conclusion

The federal EVV mandate, enacted through Section 12006(a) of the 21st Century Cures Act, represents one of the most consequential health IT requirements ever imposed on state Medicaid programs. Using T-MSIS claims data as primary evidence, FMR data as validation, and a staggered difference-in-differences design exploiting cross-state variation in implementation timing, I find that EVV compliance has not produced large aggregate spending reductions. The quarterly PCS TWFE estimate is +\$0.62M (p=0.974), with clean pre-trends (Wald p=0.875) and randomization inference confirmation (p=0.966). The mechanism decomposition shows no statistically significant effect on price, utilization, provider counts, or beneficiary counts, with the provider-count test providing the clearest evidence against a large mechanism shift.

These null findings extend across both data sources, both temporal resolutions, multiple estimators (TWFE, DID2S, triple-difference), and an extensive battery of robustness checks. However, the MDE exceeds the CBO projection by a factor of 200. I can rule out spending reductions exceeding approximately 56 percent of baseline, but cannot rule out effects in the CBO-projected range. Resolving among the competing interpretations—genuine ineffectiveness, documentation offsets, effects below detection, or incomplete implementation—requires individual-level claims analysis.

HHCS spending, the secondary mechanism check, is directionally positive under the observed-cells primary specification (annual +\$24.26M, p=0.083; quarterly +\$5.18M, p=0.100) and remains directionally positive under a zero-fill missing-data convention (quarterly +\$4.55M, p=0.094). I do not interpret HHCS as a causal effect for two independent reasons: HHCS pre-trends are rejected (p=0.0006), and the estimate is sensitive to the missing-data convention. The PCS quarterly headline is not sensitive to either issue.

Two broader lessons emerge. First, for federal mandates with projected savings of approximately 1 percent of baseline spending, evaluation infrastructure should be built into the policy design prospectively, with individual-level data collection planned from the outset. Second, the substantial gap between formal EVV compliance and operational effectiveness—documented by state audits finding that over half of claims lack EVV matching even in compliant states—suggests that technology mandates without payment-conditioned compliance may produce compliance in name without compliance in substance. The policy case for EVV should rest on its potential as program integrity infrastructure rather than on assumed spending savings that aggregate data cannot confirm.

References

Adler-Milstein, J., & Jha, A. K. (2017). HITECH Act drove large gains in hospital electronic health record adoption. *Health Affairs*, 36(8), 1416–1422.

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

CMS. (2018). Medicaid program integrity: Personal care services. Centers for Medicare & Medicaid Services.

CMS. (2023). Leveraging Electronic Visit Verification to Enhance Quality. Centers for Medicare & Medicaid Services.

CMS. (2026). Medicaid home care fraud crackdown announcement. Centers for Medicare & Medicaid Services.

Congressional Budget Office. (2016). Cost estimate: 21st Century Cures Act. Washington, DC.

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

Gardner, J. (2022). Two-stage differences in differences. Working paper.

GAO. (2017). Medicaid personal care services: CMS could do more to harmonize requirements across programs (GAO-17-169). Government Accountability Office.

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

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, 84(3), 622–641.

Herd, P., & Moynihan, D. P. (2019). *Administrative Burden: Policymaking by Other Means*. Russell Sage Foundation.

Kaiser Family Foundation. (2022). Medicaid home and community-based services enrollment and spending. KFF.

Kaiser Family Foundation. (2023). An assessment of Medicaid T-MSIS data quality. KFF.

Kaiser Family Foundation. (2024). Medicaid HCBS waiver waiting list enrollment. KFF.

Kaiser Family Foundation. (2026). CMS defers $259.5 million in Minnesota Medicaid funding. KFF.

Konetzka, R. T. (2014). The hidden costs of rebalancing long-term care. *Health Services Research*, 49(3), 771–777.

MACPAC. (2019). Electronic visit verification for personal care services: Status of state implementation. Issue Brief.

MACPAC. (2022). Medicaid and the home care workforce. Issue Brief.

MACPAC. (2024). Self-direction in Medicaid home and community-based services. Report to Congress.

Mathematica. (2023). Medicaid LTSS rebalancing brief.

Mateescu, A. (2021). Electronic visit verification: The weight of surveillance and the fracturing of care. Data & Society Research Institute.

Miller, S., & Wherry, L. R. (2019). The long-term effects of early life Medicaid coverage. *Journal of Human Resources*, 54(3), 785–824.

Office of Inspector General, HHS. (Multiple years). Medicaid Fraud Control Unit annual reports.

Office of Inspector General, HHS. (2024). Audit of Kansas EVV system. Washington, DC.

Ohio Auditor of State. (2024). Electronic visit verification report. Columbus, OH.

Olden, A., & Moen, J. (2022). The triple difference estimator. *Econometrics Journal*, 25(3), 531–553.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187–204.

PHI. (Multiple years). Direct care workers in the United States: Key facts. Paraprofessional Healthcare Institute.

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

Scales, K. (2020). It's time to care: A detailed profile of America's direct care workforce. PHI National.

Sommers, B. D., Maylone, B., Blendon, R. J., Orav, E. J., & Epstein, A. M. (2017). Changes in utilization and health among ACA Medicaid expansion enrollees. *Health Affairs*, 36(6), 1140–1148.

Steckler, A. E., O'Connell, M., & Brown, R. (2020). Experiences with electronic visit verification for personal care services. *Disability and Health Journal*, 13(4), 100926.

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

Tables

Table 1: Descriptive Statistics

Panel A: Claims-Based Spending Variables (Quarterly)

Variable	Mean	Median	SD	N	% Missing
PCS claims spending (millions)	$123.7	$28.3	$441.2	1,379	15.5%
HHCS claims spending (millions)	$5.6	$1.0	$14.8	1,262	22.7%
Total HCBS claims (millions)	$126.5	$34.1	$438.4	1,404	14.0%
PCS claims count	1,199,217	242,309	3,438,994	1,379	15.5%
HHCS claims count	95,903	18,891	292,964	1,262	22.7%
PCS $/claim	$150.42	$101.15	$173.99	1,379	15.5%
HHCS $/claim	$70.45	$63.26	$67.86	1,262	22.7%
PCS providers (avg monthly)	115.4	61.7	152.4	1,379	15.5%
HHCS providers (avg monthly)	39.6	16.3	66.7	1,262	22.7%
PCS beneficiaries (UB)	64,868	16,812	162,354	1,379	15.5%
HHCS beneficiaries (UB)	11,849	3,730	23,506	1,262	22.7%

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

Panel B: FMR Spending Variables (Quarterly, Secondary)

Variable	Mean	Median	SD	N
PCS spending (millions)	$62.0	$2.8	$192.9	1,428
HHCS spending (millions)	$16.1	$3.8	$33.1	1,428
Total HCBS spending (millions)	$155.9	$22.6	$531.8	1,428
Total Medicaid MAP (millions)	$3,564	$2,231	$4,823	1,428

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

Panel C: Treatment Distribution

Indicator	Share of Observations
PCS EVV active	43.5% (710/1,632)
HHCS EVV active	13.4% (218/1,632)
PCS-only window (DDD)	30.1% (492/1,632)

Notes: Claims data from T-MSIS Medicaid Provider Spending dataset (Jan 2018 – Dec 2024), filtered to 13 EVV-relevant HCPCS codes. Beneficiary counts are upper bounds (may double-count across providers). FMR data from CMS Financial Management Reports, FY2018–FY2024; quarterly values are annual / 4. Missingness in claims reflects CMS suppression (<12 claims) and states with no claims in certain quarters.

Table 2: Main Difference-in-Differences Results (Claims, Primary)

Panel A: Quarterly TWFE (Primary Specification)

Outcome	Coef	SE	p	N
PCS spending (millions)	+0.62	18.61	0.974	1,379
HHCS spending (millions)	+5.18	3.09	0.100	1,261
Total HCBS (millions)	+2.37	17.54	0.893	1,404
PCS claims count	+140,458	227,567	0.540	1,379
HHCS claims count	+70,024	36,021	0.058	1,261

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

Panel B: Annual TWFE (Observed-Cells Primary)

Outcome	Coef (millions)	SE	p	N
PCS spending	+9.39	69.83	0.894	347
HHCS spending	+24.26	13.71	0.083	325
Total HCBS	+13.23	66.91	0.844	353

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

Panel C: Annual DID2S (Gardner 2022, Observed-Cells Primary)

Outcome	Coef (millions)	SE	p	N
PCS spending	+33.28	111.33	0.766	243
HHCS spending	+33.37	7.99	<0.001	278
Total HCBS	+43.68	107.97	0.688	250

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

Panel D: HHCS Missing-Data Sensitivity (Quarterly TWFE)

Convention	Coef (millions)	SE	p	N
Observed cells (primary)	+5.18	3.09	0.100	1,261
Zero-fill (sensitivity)	+4.55	2.67	0.094	1,428

Notes: Panel A: T-MSIS claims, quarterly TWFE with state and quarter FE; standard errors clustered by state. Panel B: Annualized T-MSIS claims using sum(min_count=1) and dropping all-missing state-years; state and year FE. Panel C: Manual DID2S on annualized claims; Stage 1 uses untreated observations only; always-treated states excluded. Panel D: Sensitivity of the HHCS quarterly TWFE to the missing-data convention; observed-cells is the primary specification locked in the analytic registry, zero-fill is a sensitivity restricted to the 2018–2024 T-MSIS data era. The HHCS DID2S and annual count findings should be interpreted with extreme caution given rejected HHCS pre-trends (Wald p=0.0006). MDE at 80% power: $52.14M/state-quarter for PCS (55.6% of baseline, 209x CBO projection).

Table 3: Mechanism Decomposition (Claims)

Channel	Outcome	Service	Coef	SE	p	Mean	% Effect	N
Price	$/Claim	PCS	-36.71	24.77	0.145	$150.42	-24.4%	1,379
Price	$/Claim	HHCS	+0.01	7.80	0.999	$70.45	+0.02%	1,261
Intensity	Claims/Provider	PCS	+2,870	2,650	0.284	11,586	+24.8%	1,368
Intensity	Claims/Provider	HHCS	-272	1,269	0.831	4,456	-6.1%	1,240
Extensive	Provider Count	PCS	-1.2	4.4	0.780	115.4	-1.1%	1,379
Extensive	Provider Count	HHCS	+1.5	6.0	0.810	39.6	+3.7%	1,261
Beneficiaries	Bene Count (UB)	PCS	+4,058	8,011	0.615	64,868	+6.3%	1,379
Beneficiaries	Bene Count (UB)	HHCS	+3,928	3,330	0.244	11,849	+33.2%	1,261

Notes: TWFE with state and quarter FE; standard errors clustered by state. “% Effect” = coefficient / dependent variable mean x 100. Beneficiary counts are upper bounds. Provider counts are average monthly unique NPIs. N varies across rows due to missingness in derived variables.

Table 4: Triple-Difference Results (Claims)

Specification	Coef (millions)	SE	p	N
DDD: PCS-only window	+5.88	17.08	0.732	2,472
DDD: Service-specific EVV	+3.21	11.71	0.785	2,472

Notes: Stacked panel with two rows per state-quarter (PCS and HHCS claims spending). Includes state-by-service and time-by-service fixed effects. Standard errors clustered by state.

Table 5: Robustness Summary (Claims, Primary)

Test	Specification/Variant	PCS Coef (millions)	SE	p	Conclusion
Pre-trends	Wald test (7 leads), PCS	–	–	0.875	Not rejected
Pre-trends	Wald test (7 leads), HHCS	–	–	0.0006	**Rejected**
Randomization inference	500 permutations, PCS	–	–	0.966	Null confirmed
ARPA sensitivity	State + qtr FE (no ARPA)	+0.62	18.61	0.974	Stable
	+ ARPA control	+0.62	18.61	–	Stable
	+ Full controls	+3.01	20.32	–	Stable
T-MSIS quality	Full sample (51 states)	+0.62	18.61	0.974	Stable
	Excl. unusable (6 states)	+3.03	21.84	0.890	Stable
	Excl. high concern (16)	-5.84	25.07	0.817	Stable
	Excl. both (22 states)	-1.49	32.66	0.964	Stable
Go-live shift	-1 quarter	-0.55	15.53	0.972	Stable
	Baseline	+0.62	18.61	0.974	Stable
	+1 quarter	+4.41	21.05	0.835	Stable
Excl. non-compliant	Drop 7 HHCS-noncompliant	+5.33	12.91	0.682	Stable
EVV model type	Open (41 states)	-6.55	22.00	0.767	Null
	Closed (7 states)	+48.07	47.25	0.348	Null
Early vs. late	Early (34 states, by 2021Q1)	+17.54	7.30	0.022	Positive*
	Late (17 states)	-35.34	18.01	0.067	Negative*

Notes: All estimates are quarterly claims TWFE with state and quarter FE unless otherwise noted. Standard errors clustered by state. Early/late implementer results are exploratory (see text).*

Figures

Figure 1: Raw Outcome Trends — Cohort-Specific PCS and HHCS Spending

Notes: Mean quarterly spending for PCS (left panel) and HHCS (right panel) by EVV adoption cohort (Early/Middle/Late). Vertical lines mark median go-live dates. Pre-treatment trends are broadly parallel for PCS, consistent with the

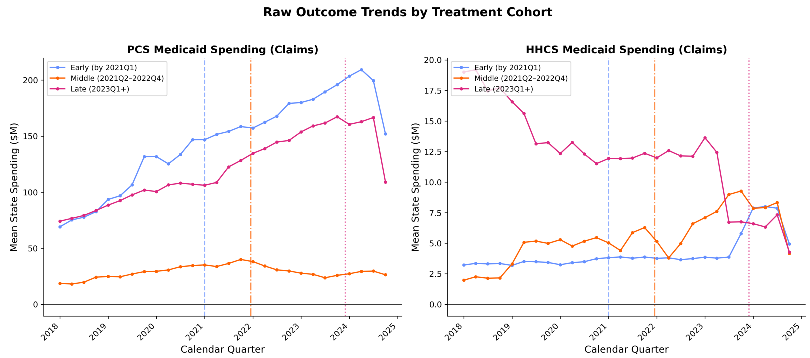

Figure 1: Figure 1

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

clean Wald test (p=0.875). No sharp trend break is visible at treatment onset in either panel.

Figure 2: Treatment Timing — PCS EVV Cohort Sizes and Rollout

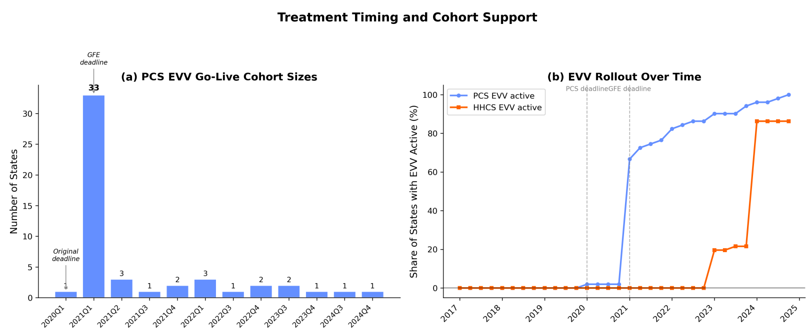

Figure 2: Figure 2

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

Notes: Panel (a): Bar chart of PCS go-live cohort sizes with federal deadline annotations. The 2021Q1 cohort (33 states) dominates. Panel (b): Share of states with active EVV over time for PCS and HHCS. Together, these panels illustrate the limited staggered variation available for identification.

Figure 3: Event Study — Effect of PCS EVV on ln(PCS Spending) (TWFE, Quarterly)

Notes: TWFE event study on log PCS claims spending; coefficients multiplied by 100 approximate percent-change effects. Reference period k=-1. Pre-treatment

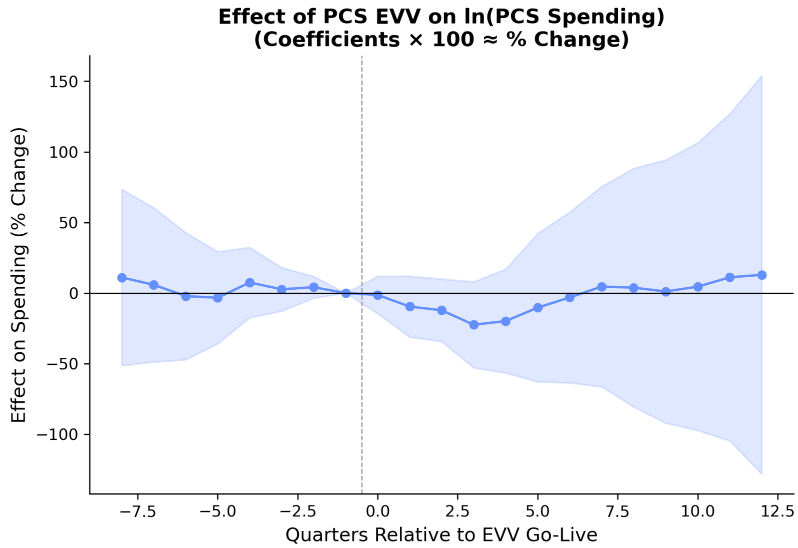

Figure 3: Figure 3

Note: This figure presents the es log pcs. It is included to make the empirical design, sample structure, or headline result easier to read alongside the surrounding text.

coefficients are tightly centered around zero, consistent with the clean Wald pre-trend test (p=0.875). Post-treatment coefficients fluctuate near zero with wide confidence intervals, averaging -3.1 percent. State and quarter FE; standard errors clustered by state. Level-based ($M) event study in Appendix Figure A8.

Figure 4: Event Study — Effect of HHCS EVV on ln(HHCS Spending) (TWFE, Quarterly)

Notes: TWFE event study on log HHCS claims spending; coefficients × 100 approximately percent change. Pre-trends clearly rejected (Wald p=0.0006). Post-treatment coefficients cannot be interpreted as causal given pre-trend violation. Level-based event study in Appendix Figure A9.

Figure 5: Mechanism Decomposition — Effects of PCS EVV Across Channels

Notes: Point estimates and 95% CIs for PCS mechanism channels. No channel shows a statistically significant effect. The suggestive PCS price effect (-$36.60/claim, p=0.146) is the largest in magnitude.

Figure 6: Randomization Inference — PCS Claims Spending

Notes: Distribution of 500 placebo estimates. Vertical line marks actual estimate (-$1.80M). RI p-value = 0.966.

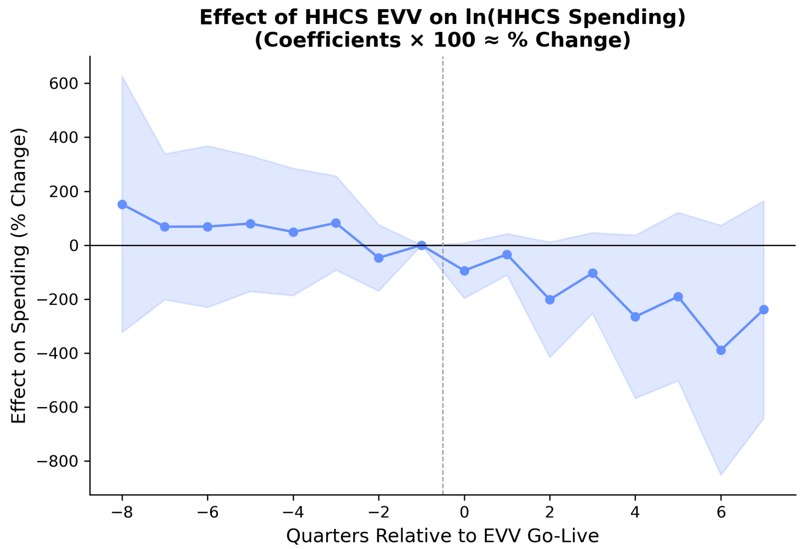

Figure 4: Figure 4

Note: This figure presents the es log hhcs. It is included to make the empirical design, sample structure, or headline result easier to read alongside the surrounding text.

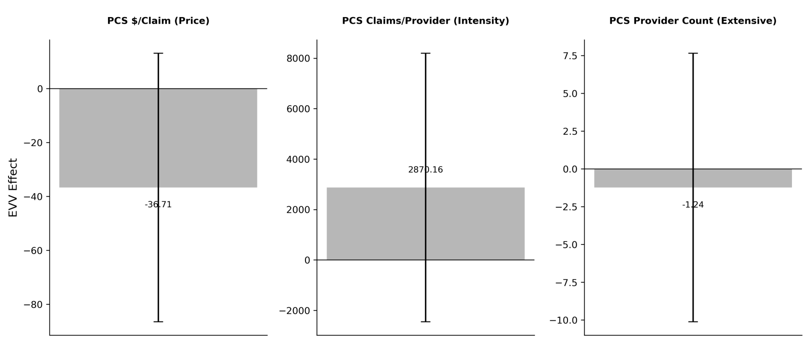

Figure 5: Figure 5

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

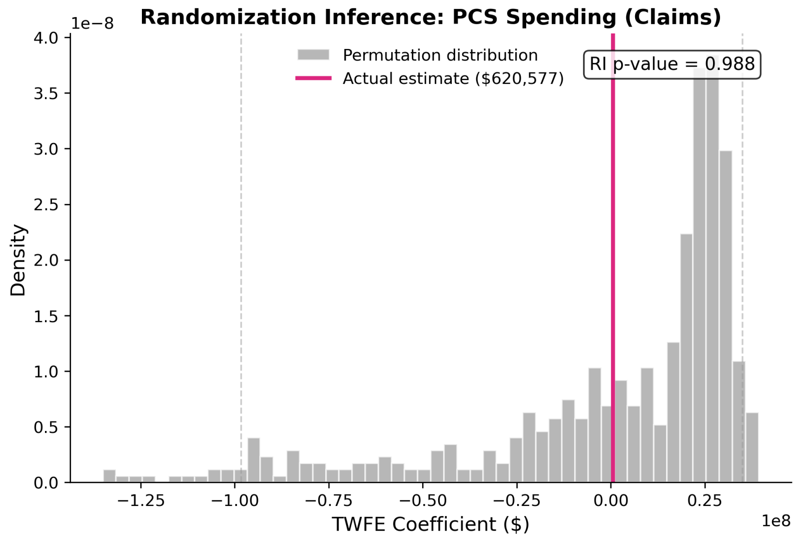

Figure 6: Figure 6

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

Appendix: Electronic Visit Verification and Medicaid Home Care Spending, Utilization, and Provider Participation

---

Appendix A: FMR Secondary Results

Table A1: FMR Difference-in-Differences Results (Secondary Validation)

Panel A: Quarterly FMR TWFE (Annualized-to-Quarterly)

Outcome	TWFE (State + Qtr FE)			TWFE (+ Controls)			N
	Coef (millions)	SE	p	Coef (millions)	SE	p	
PCS spending	+0.91	11.10	0.935	+1.68	11.69	0.887	1,428
HHCS spending	+2.21	3.88	0.572	+2.27	3.89	0.562	1,428
Total HCBS	+12.95	16.92	0.448	+16.29	19.23	0.401	1,428
PCS share of HCBS	-0.006	0.025	0.811	-0.004	0.026	0.877	1,360

Notes: This table summarizes policy timing, cohorts, thresholds, or state-level sample construction. It is intended to make the identifying variation and comparison groups transparent.

Panel B: Annual FMR TWFE (Primary FMR Specification)

Outcome	Coef (millions)	SE	p	N
PCS spending	+8.72	45.31	0.848	353
HHCS spending	+5.87	16.32	0.721	353
Total HCBS	+40.26	68.52	0.559	353
PCS share of HCBS	-0.007	0.026	0.788	339

Notes: FMR data from CMS Financial Management Reports, FY2018–FY2024. Panel A: quarterly values are annual ÷ 4; within-year variation is artificial. Panel B: annual data at native FMR frequency. All results corroborate the claims-based null. Standard errors clustered by state.

Appendix B: Additional Claims-Based Results

Table A2: DID2S Results (Annual)

Panel A: Claims-Based DID2S

Outcome	Coef (millions)	SE	p	N
PCS claims spending	+33.28	111.33	0.766	243
HHCS claims spending	+33.37	7.99	<0.001	278
Total HCBS claims	+43.68	107.97	0.688	250

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

Panel B: FMR-Based DID2S

Outcome	Coef (millions)	SE	p	N
PCS spending (FMR)	-51.22	32.94	0.126	255
HHCS spending (FMR)	+9.92	15.23	0.518	306
Total HCBS (FMR)	-59.82	83.32	0.476	255

Notes: Manual two-stage DID2S (Gardner 2022) using annualized data. Stage 1 estimates unit and time FE from untreated observations only; always-treated states excluded. The claims HHCS DID2S result (p<0.001) should be interpreted with extreme caution given rejected HHCS pre-trends (p=0.0006). All estimates are statistically insignificant except claims HHCS.

Table A3: Goodman-Bacon Decomposition (Claims)

Comparison	Cohort 1	Cohort 2	N_1	N_2	Estimate (millions)	Weight
2020 vs 2021	1	39	–	–	-245.98	0.076
2020 vs 2022	1	6	–	–	-38.00	0.012
2021 vs 2022	39	6	–	–	+211.15	**0.454**
2021 vs 2023	39	3	–	–	+160.60	0.227
2021 vs 2024	39	2	–	–	-187.16	0.151
2022 vs 2023	6	3	–	–	-89.13	0.035
2022 vs 2024	6	2	–	–	-491.04	0.023
2023 vs 2024	3	2	–	–	-425.13	0.012

Notes: Decomposition weights are approximate (cohort-size-based rather than exact variance-weighted per Goodman-Bacon [2021]). The 2021-vs-2022 comparison receives 45% of the weight. The qualitative conclusion—that the large 2021Q1 cohort dominates—is robust to this approximation.

Table A4: Pre-Trends Tests

Outcome	Data Source	N Pre-Coefs	Wald Stat	df	p	Max abs t	Any Indiv Sig 5%
PCS Spending	Claims	7	3.10	7	**0.875**	0.96	No
HHCS Spending	Claims	7	25.66	7	**0.0006**	3.34	Yes

Notes: Joint Wald test on pre-treatment event study coefficients. PCS claims pre-trends cleanly not rejected (p=0.875). HHCS claims pre-trends clearly rejected (p=0.0006); HHCS-specific results should be interpreted with extreme caution.

Table A5: ARPA Section 9817 Sensitivity (Claims)

Outcome	Specification	Coef	SE	N
PCS spending (millions)	Base: state + quarter FE	+0.62	18.61	1,379
PCS spending (millions)	+ ARPA control	+0.62	18.61	1,379
PCS spending (millions)	+ full controls	+3.01	20.32	1,379
PCS spending (millions)	+ year FE and ARPA	+3.29	20.32	1,379
HHCS spending (millions)	Base: state + quarter FE	+5.18	3.09	1,261
HHCS spending (millions)	+ ARPA control	+5.18	3.09	1,261
HHCS spending (millions)	+ full controls	+5.22	3.10	1,261
HHCS spending (millions)	+ year FE and ARPA	+5.25	3.06	1,261
PCS claims count	Base: state + quarter FE	+140,458	227,567	1,379
PCS claims count	+ ARPA control	+140,458	227,567	1,379
PCS claims count	+ full controls	+162,477	241,943	1,379
PCS claims count	+ year FE and ARPA	+157,947	238,951	1,379

Notes: ARPA Section 9817 provided a 10 percentage point enhanced FMAP for HCBS (April 2021 – March 2022). Quarter FE absorb the common ARPA effect, so the base and + ARPA columns are identical. Results are highly stable across all specifications.

Table A6: Go-Live Date Sensitivity (Claims)

Shift	PCS Coef (millions)	SE	p	HHCS Coef (millions)	SE	p
-1 quarter	-0.55	15.53	0.972	+5.49	3.22	0.095
Baseline (0)	+0.62	18.61	0.974	+5.18	3.09	0.100
+1 quarter	+4.41	21.05	0.835	+5.20	2.92	0.081

Notes: Robust to +/- 1 quarter shifts in treatment timing, addressing measurement error in compliance dates.

Table A7: Excluding HHCS-Noncompliant States (Claims)

Outcome	Full Sample			Excl. Non-compliant		
	Coef (millions)	SE	N	Coef (millions)	SE	N
PCS spending	+0.62	18.61	1,379	+5.33	12.91	1,183
HHCS spending	+5.18	3.09	1,261	+0.55	1.59	1,107
PCS claims count	+140,458	227,567	1,379	+160,601	169,351	1,183
PCS provider count	-1.2	4.4	1,379	+4.9	3.3	1,183

Notes: Excludes 7 HHCS-noncompliant states (AR, GA, MA, MI, MS, ND, SC). All estimates remain insignificant.

Table A8: T-MSIS Data Quality Sensitivity

Outcome	Sample	States Excl.	Coef (millions)	SE	p	N
PCS spending	Full	0	+0.62	18.61	0.974	1,379
PCS spending	Excl. unusable	6	+3.03	21.84	0.890	1,230
PCS spending	Excl. high concern	16	-5.88	25.09	0.816	944
PCS spending	Excl. both	22	-1.53	32.67	0.963	795
HHCS spending	Full	0	+5.18	3.09	0.100	1,261
HHCS spending	Excl. unusable	6	+5.50	3.22	0.095	1,093
HHCS spending	Excl. high concern	16	+5.74	3.70	0.130	878
HHCS spending	Excl. both	22	+6.60	3.99	0.110	710
PCS claims count	Full	0	+140,458	227,567	0.540	1,379
PCS claims count	Excl. unusable	6	+130,921	221,333	0.557	1,230
PCS claims count	Excl. high concern	16	-49,223	185,419	0.792	944
PCS claims count	Excl. both	22	-12,357	234,763	0.958	795

Notes: State data quality classifications from KFF (2023) assessment of T-MSIS managed care encounter data. “Unusable” = 6 states with managed care data too poor for analysis; “High concern” = 16 states with significant managed care data quality issues. All estimates remain insignificant regardless of exclusions.

Table A9: Heterogeneity — Early vs. Late Implementers (Claims)

Group	Coef (millions)	SE	p	N	N States
Early (by 2021Q1)	+17.54	7.30	0.022	916	34
Late (after 2021Q1)	-35.34	18.01	0.067	463	17

Notes: Exploratory subgroup analysis. The early-implementer result is positive (opposite expected fraud-reduction direction) and the contrast with the negative late-implementer result likely reflects compositional differences rather than a stable treatment effect. See text for discussion of cutoff sensitivity and mean reversion concerns.

Table A10: Early-Implementer Cutoff Sensitivity (Claims)

Note: Based on FMR quarterly supplementary data; claims-based cutoff sensitivity analysis follows the same qualitative pattern. Early-implementer significance is sensitive to the chosen cutoff.

Cutoff	N Early	Early Coef (millions)	SE	p	N Late	Late Coef (millions)	SE	p
By 2021Q1 (baseline)	34	+2.92	1.33	0.036	17	+28.30	36.03	0.444
By 2021Q2	37	+1.41	1.17	0.236	14	+46.11	53.42	0.404
By 2021Q4	42	+14.34	12.06	0.241	9	-15.29	7.08	0.063
By 2022Q1	43	+12.09	10.55	0.258	8	-17.08	7.43	0.055

Notes: FMR-based annualized-to-quarterly results. The baseline early-implementer finding (p=0.036) is sensitive to the cutoff definition.

Table A11: Heterogeneity — EVV Model Type (Claims)

EVV Model	N States	PCS Coef (millions)	SE	p	N Obs
Open (provider choice)	41	-6.55	22.00	0.767	1,112
Closed (state-mandated)	7	+48.07	47.25	0.348	183

Notes: Neither model type shows a significant effect.

Table A12: Oster (2019) Bounds (Claims)

Outcome	Short Beta	Short R^2	Long Beta	Long R^2	Delta (1.3xR^2)	Beta* (1.3xR^2)
PCS spending (millions)	-1.58	0.909	+1.04	0.909	-101	+668M
HHCS spending (millions)	+4.85	0.852	+4.88	0.852	-171,965	+45.7M
PCS claims count	+89,000	0.909	+108,767	0.909	-1,406	+5.2M
PCS $/claim	-36.21	0.702	-33.97	0.704	+2,318	+306

Notes: Oster bounds test is designed to assess whether a significant result could be explained by omitted variable bias. With near-zero point estimates, deltas are mechanically extreme and uninformative. Included for completeness.

Table A13: Power Analysis

Specification	Outcome	N Clusters	N Obs	MDE (millions)	MDE (% Baseline)	CBO dollars/unit (millions)	MDE/CBO
Claims PCS Quarterly (primary)	PCS claims	51	1,379	52.14	55.6%	0.25	209x
Claims HHCS Quarterly	HHCS claims	50	1,262	8.65	174.5%	–	–
Claims PCS Annual	PCS claims	51	347	195.64	52.7%	1.00	196x
FMR PCS Quarterly (secondary)	PCS (FMR)	51	1,428	31.10	51.7%	0.25	125x

Specification	Outcome	N Clusters	N Obs	MDE (millions)	MDE (% Baseline)	CBO dollars/unit (millions)	MDE/CBO
FMR PCS Annual (secondary)	PCS (FMR)	51	353	126.93	52.1%	1.00	127x

Notes: MDE = (1.96 + 0.84) x SE at 80% power, 5% significance. CBO projected $290M in federal savings over 10 years; converted to total computable at 57% average FMAP = ~$509M total, or ~$1.0M per state-year / ~$0.25M per state-quarter. The study is substantially underpowered for CBO-projected savings regardless of data source.

Table A14: FMR Normalized Spending Results

Outcome	Per Capita (Annual)			Per Enrollee (Annual)		
	Coef ()$\|SE\|p\|Coef$()	SE	p			
PCS spending	-3.84	3.81	0.318	-8.82	7.30	0.233
HHCS spending	+2.95	2.29	0.203	+10.39	6.62	0.123

Notes: FMR-based annual models. Population-normalized results use 2020 Census. Enrollment-normalized results use MBES quarterly average total Medicaid enrollment. Both remain null.

Appendix C: Supplementary Figures

Figure A1: FMR Event Studies (Secondary Validation)

Panel A: PCS Spending — FMR Quarterly Event Study

Panel B: HHCS Spending — FMR Quarterly Event Study

Notes: Annualized-to-quarterly FMR event studies. Within-year variation is artificial (annual data ÷ 4). FMR pre-trends should be compared with claims pre-trends in the main text.

Figure A2: Annual Event Studies (Claims)

Notes: Annualized claims event study at native annual frequency.

Figure A3: FMR Annual Event Study

Notes: FMR annual event study at native FMR annual frequency.

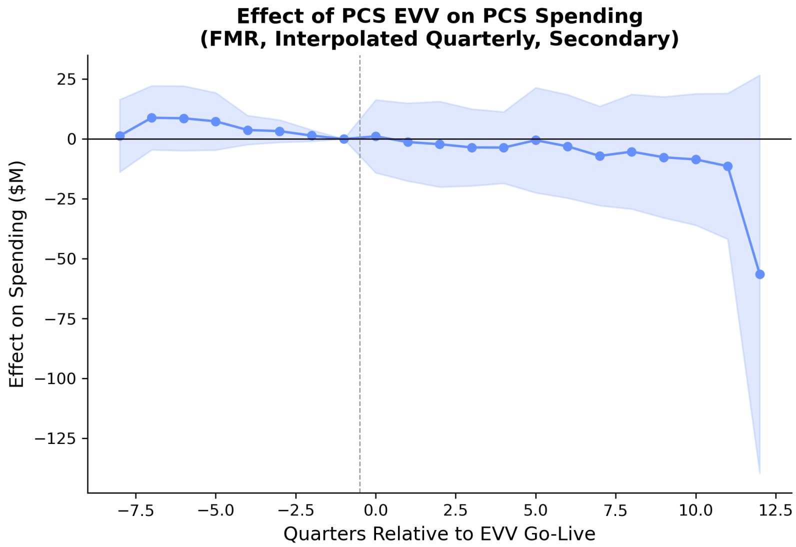

Figure 7: Appendix Figure A1a

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

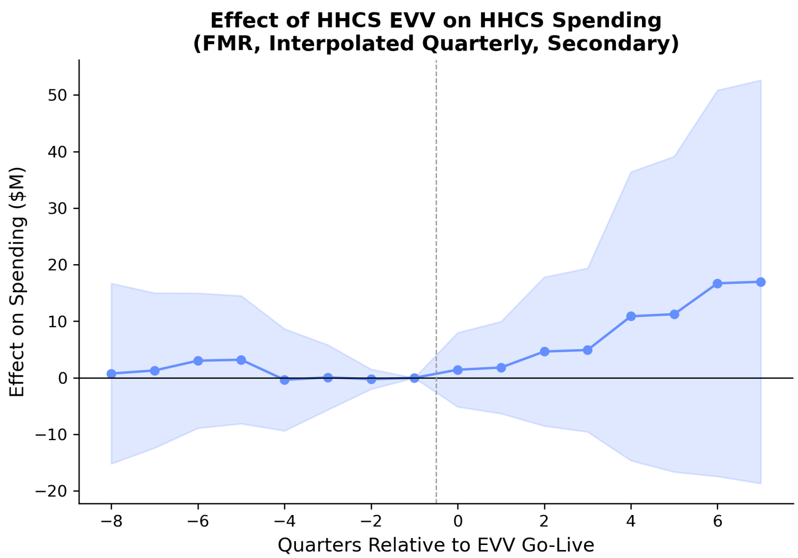

Figure 8: Appendix Figure A1b

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

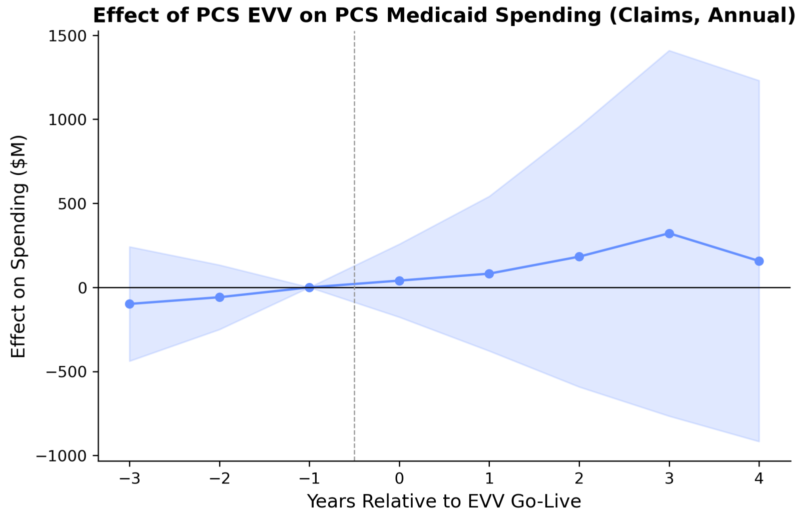

Figure 9: Appendix Figure A2

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

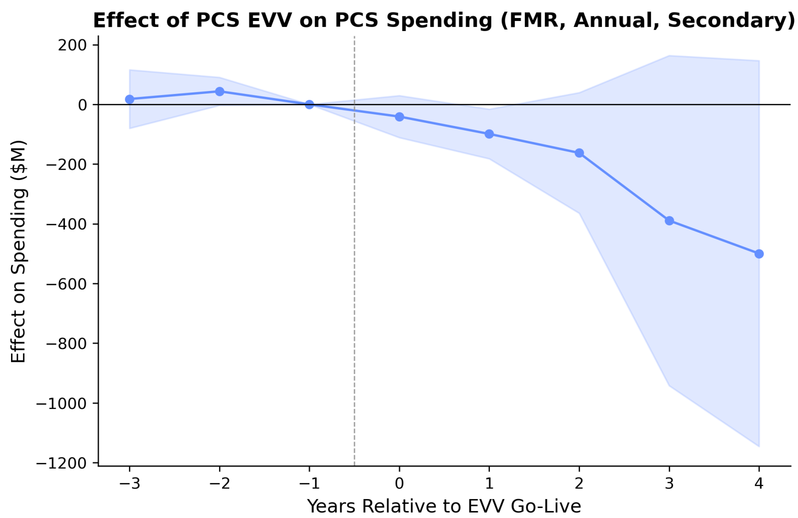

Figure 10: Appendix Figure A3

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

Figure A4: Robustness Summary — PCS Claims Spending Across Specifications

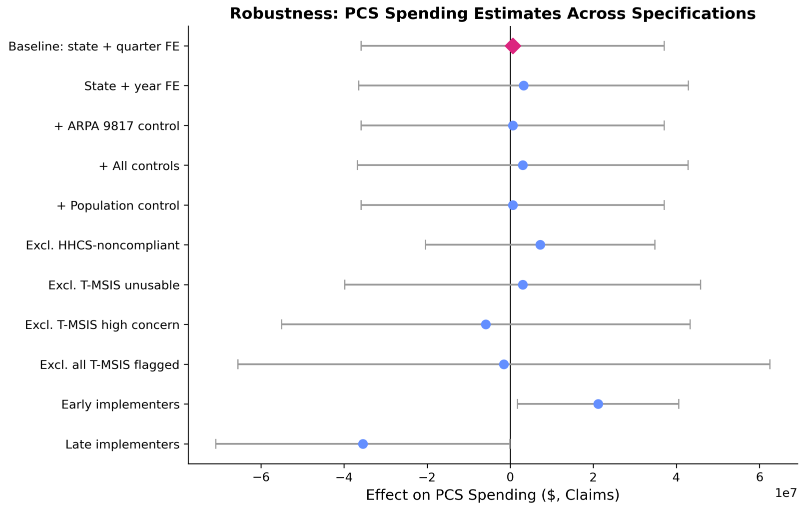

Figure 11: Appendix Figure A4

Note: This figure reports a robustness or sensitivity check for the robustness summary. It shows how the main estimate changes under alternative assumptions, samples, or specifications.

Notes: Point estimates and 95% CIs for PCS claims spending across alternative specifications. All CIs include zero.

Figure A5: Heterogeneity by EVV Model Type (Claims)

Notes: Neither open nor closed EVV model shows a significant effect.

Figure A6: Goodman-Bacon Decomposition — Weight x Estimate Scatter (Claims)

Notes: Goodman-Bacon decomposition showing the weight and estimate for each pairwise cohort comparison underlying the TWFE estimate. The 2021-vs-2022 comparison receives approximately 45% of the total weight. Substantial heterogeneity in estimates across comparisons is visible, though the weighted average is near zero.

Figure A7: Alternative Scale Event Study — PCS Claims Spending (Panel)

Notes: PCS event study presented in three alternative outcome scales: log spending (coefficients × 100 approximately % change), spending per 100K population, and spending as a percentage of pre-treatment baseline. The log specification

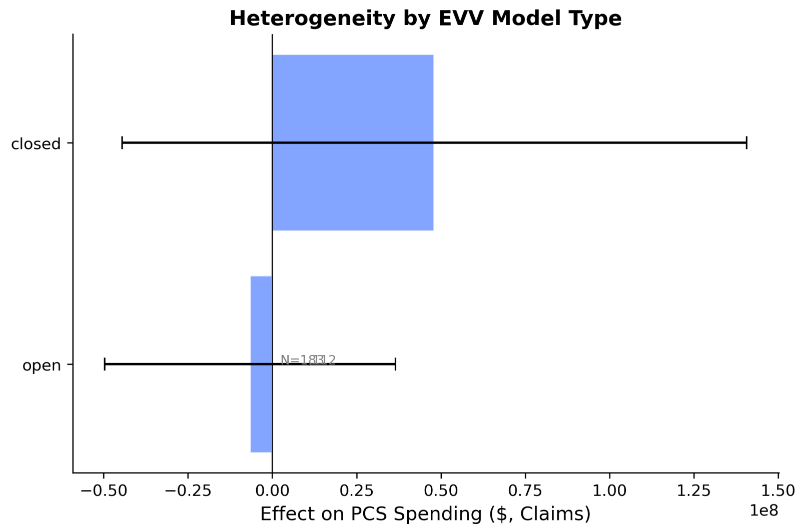

Figure 12: Appendix Figure A5

Note: This figure compares estimates across groups or specifications for the heterogeneity model. It is intended to make effect heterogeneity and subgroup precision easier to assess.

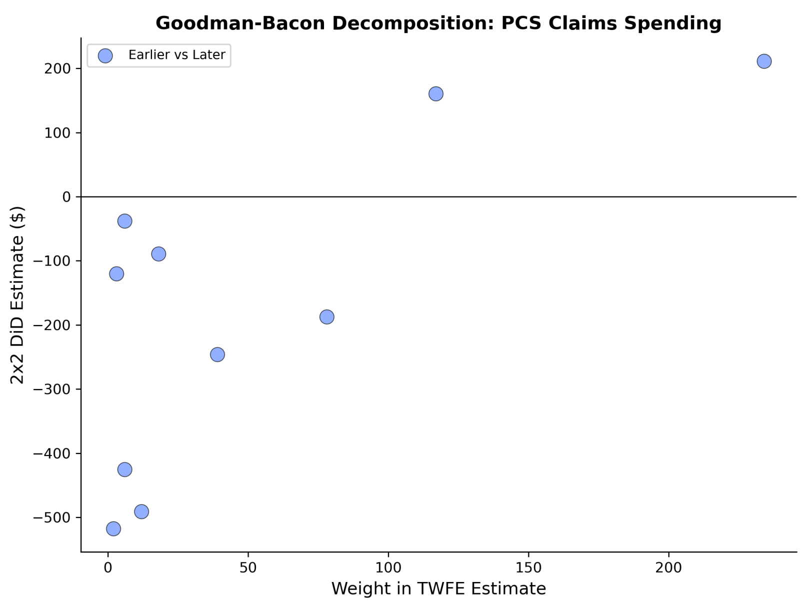

Figure 13: Appendix Figure A6

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

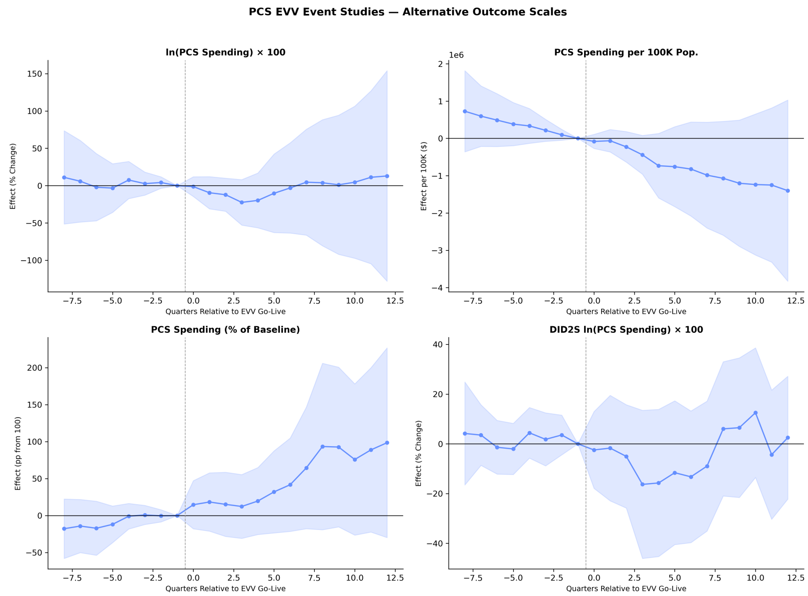

Figure 14: Appendix Figure A7

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

(used as the main text Figure 3) shows the cleanest pre-trends and a null post-treatment path averaging -3.1 percent. The per-capita specification shows a post-treatment downward drift that warrants caution (see Appendix Figure A10). The percentage-of-baseline specification shows an upward drift likely reflecting secular spending growth rather than a causal EVV effect.

Figure A8: Event Study — PCS Claims Spending in Levels ($M, TWFE, Quarterly)

Notes: Level-based ($M) TWFE event study. This was the original main-text figure; it is moved to the appendix because the log specification is more interpretable given the 160,000-fold difference in PCS spending across states (NY: $2.9B/quarter vs. NE: $18K/quarter). The near-zero coefficients and wide CIs reflect the dominance of a few large states in the level-based outcome, not a coding error.

Figure A9: Event Study — HHCS Claims Spending in Levels ($M, TWFE, Quarterly)

Notes: Level-based ($M) HHCS event study. Pre-trends rejected (Wald p=0.0006).

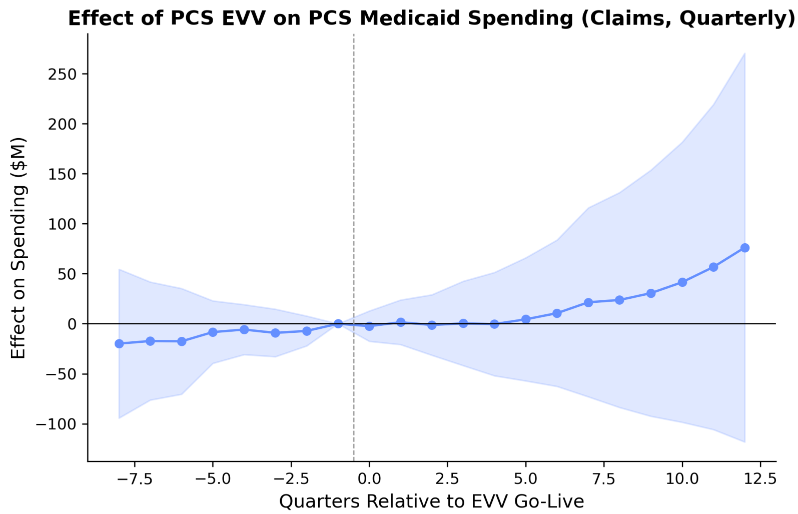

Figure 15: Appendix Figure A8

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

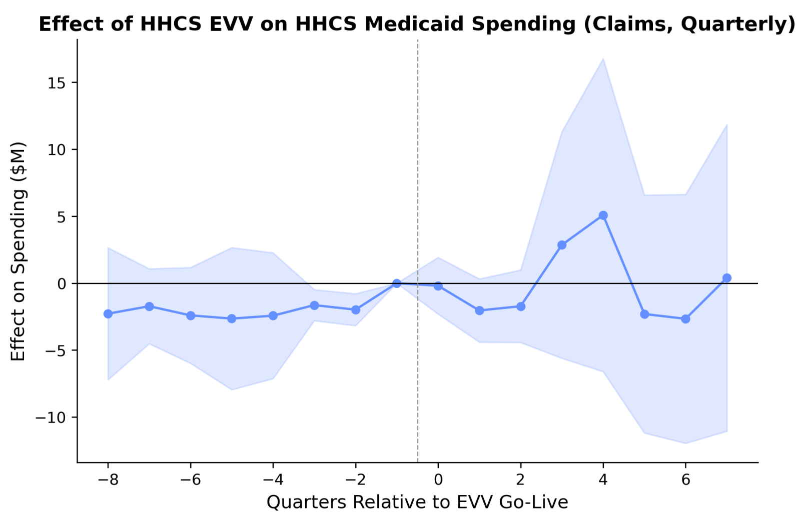

Figure 16: Appendix Figure A9

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

Figure A10: Per-Capita Event Study — PCS Spending per 100K Population

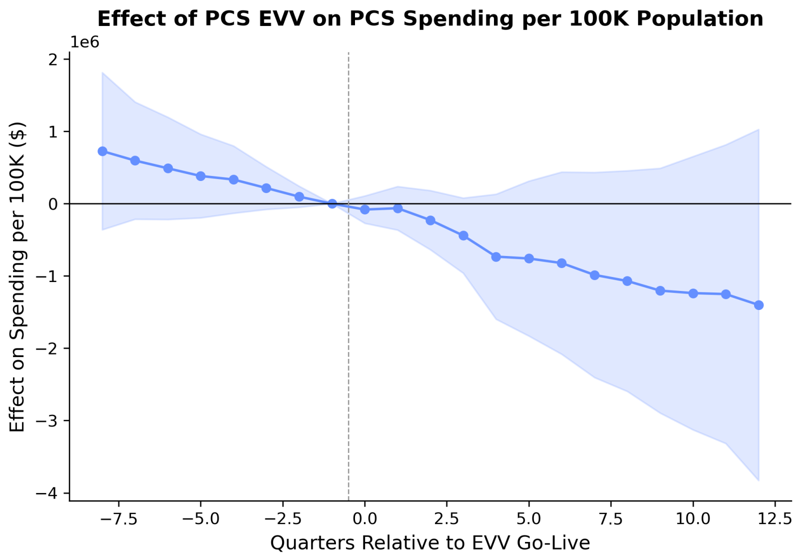

Figure 17: Appendix Figure A10

Note: This figure presents the es percapita pcs. It is included to make the empirical design, sample structure, or headline result easier to read alongside the surrounding text.

Notes: PCS spending per 100K population. The post-treatment downward drift is visible but should be interpreted with caution: investigation shows it is driven primarily by secular population growth in the denominator rather than a causal spending decline, and it is sensitive to the exclusion of a few large states (see per-capita investigation diagnostics).

Appendix D: PCS Treatment Cohort Distribution

Cohort Quarter	N States	Examples
2020Q1	1	MD
2021Q1	33	AL, AK, AZ, CO, FL, HI, ID, IN, KS, KY, LA, ME, MO, NE, NV, NJ, NM, NY, ND, OH, OK, PA, RI, SD, TN, TX, VT, VA, WA, WI, DC
2021Q2	3	AR, GA, WV
2021Q3–Q4	3	UT, MN, OR
2022Q1–Q4	6	CA, CT, NH, WY, SC, DE
2023Q3–Q4	3	IL, MT, MS
2024Q3–Q4	2	MI, MA

Notes: 33 of 51 states (65%) share the 2021Q1 cohort, limiting staggered design variation.

Appendix E: HHCS Compliance Status

As of January 1, 2024:

- **Compliant**: 44 states
- **Not compliant**: AR, GA, MA, MI, MS, ND, SC (7 states)

Appendix F: Data Construction Notes

T-MSIS Claims Processing Pipeline

1. **Raw data**: 227,083,361 rows (11 GB), Jan 2018 – Dec 2024
2. **HCPCS filter**: 13 EVV-relevant codes (6 PCS: T1019, T1020, S5125, S5126, S5130, S5131; 7 HHCS: G0151, G0152, G0153, G0156, G0299, G0300, T1021) → 1,217,993 rows
3. **NPI-to-state crosswalk**: 1,762,665 NPIs mapped via NPPES servicing+billing companion tables → 98.5% post-fallback state assignment → 1,199,784 claims with state assignment
4. **Aggregation**: Provider-procedure-month → state-quarter-service type (wide format)
5. **Coverage**: PCS: 1,379/1,632 state-quarters nonmissing; HHCS: 1,262/1,632

Claims Suppression

CMS suppresses cell counts below 12 beneficiaries. This primarily affects small states and rare HHCS codes, creating non-random missingness. The 15.5% PCS and 22.7% HHCS missingness rates reflect this suppression plus states with no claims in certain quarters.

Beneficiary Double-Counting

T-MSIS beneficiary counts aggregate across providers within a state-quarter. A beneficiary receiving services from multiple providers is counted multiple times. All beneficiary counts are labeled as upper bounds.

FMR Data Frequency

CMS Financial Management Reports are annual by federal fiscal year. My quarterly FMR panel divides annual values by four. Annual FMR models (N=408 state-years) align with the native data frequency; quarterly FMR models are supplementary.

Negative FMR Spending Values

Some FMR state-year observations report negative PCS or HHCS spending, reflecting CMS-64 adjustments and recoupments. These are retained in the analysis.

Zero PCS States

Several states report zero state-plan PCS spending because personal care is delivered through 1915(c) waiver programs or managed care. These are included in the analysis.

References

Gardner, J. (2022). Two-stage differences in differences. Working paper.

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

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, *84*(3), 622–641.

Kaiser Family Foundation. (2023). An assessment of Medicaid T-MSIS data quality. KFF.

Olden, A., & Moen, J. (2022). The triple difference estimator. *Econometrics Journal*, *25*(3), 531–553.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.