Trading Inpatient Beds for Financial Lifelines: Rural Emergency Hospital Conversions and Hospital Margins

Abstract

Background. The Rural Emergency Hospital (REH) designation, created by Section 125 of the Consolidated Appropriations Act, 2021 and operationalized by CMS in the CY2023 Outpatient Prospective Payment System final rule, is the first new Medicare rural provider type in 26 years. Eligible Critical Access Hospitals (CAHs) and small rural Prospective Payment System (PPS) hospitals may relinquish all acute inpatient care in exchange for the standard outpatient rate plus a 5% add-on and a monthly facility fee that is approximately $295,000 in CY2026 (approximately$3.5 million annualized). Whether the trade — inpatient capacity for outpatient revenue stabilization — improves the financial condition of converting facilities is unknown.

Methods. I assemble a 2012–2024 hospital-year panel from the CMS Healthcare Cost Report Information System (HCRIS) and link 38 dated REH conversions to their pre-conversion provider numbers using a hand-built OLD<->NEW CCN crosswalk (37 of 38 high-confidence matches within the stated +/-90-day termination window; one match retained on address and name despite a 365-day POS termination-date lag). The comparison pool is non-converting Critical Access Hospitals. I estimate Callaway-Sant'Anna (2021) staggered difference-in-differences ATTs as the primary estimator and report two-way fixed-effects (TWFE) event studies as benchmarks. The primary specification trims the comparison pool on a logistic propensity score (overlap region pscore in [0.05, 0.95]). I test pre-trends by the joint Wald statistic on event-time leads -5 to -2 and bound post-period bias with the Rambachan-Roth (HonestDiD) relative-magnitudes sensitivity procedure (M-bar in {0.5, 1.0, 1.5, 2.0}). The financial outcome historically labeled "total margin" is computed as net income divided by total operating revenue and is therefore relabeled in this revision as a net-income-to-operating-revenue ratio rather than a true total-margin ratio; a true total-margin denominator that adds non-operating revenue cannot be constructed from the public HCRIS columns used here.

Results. Inpatient days, the policy-mechanism check, fall sharply under the primary specification: Callaway-Sant'Anna ATT = $-1{,}485$ days per hospital-year on the propensity-trimmed pool (SE 260, $p < 0.001$) and $-1{,}376$ days on the full-pool sensitivity (SE 256, $p < 0.001$). The trimmed-pool pre-trend joint $p = 0.482$ is clean and HonestDiD bounds at event time $t = 0$ do not break out to M-bar > 2.0. The financial proxy (net income / operating revenue, winsorized) is not statistically distinguishable from zero under the primary trimmed-pool CS estimator: ATT = +7.4 percentage points (SE 5.9 pp, $p = 0.21$). The

full-pool CS estimate is +17.8 pp (SE 5.6, p = 0.002) and the trimmed-pool TWFE event study returns a t = 0 coefficient of +23.0 pp (SE 6.9, p = 0.001); both are reported as sensitivities, not as the headline. Operating margin shows a negative coefficient (−25.5 pp, p = 0.051) but with failed pre-trends (p = 0.003); the post-conversion REH facility-payment supplement is recorded outside operating revenue under modified HCRIS Worksheet G-3, so the negative operating-margin estimate is best interpreted as an accounting reclassification rather than as evidence of financial deterioration.

Conclusions. REH conversion in its first eighteen months delivered on the statutory mechanism — inpatient bed-days at converting facilities fell sharply, exactly as the program requires — but the financial half of the trade is not cleanly identified under the audit-mandated primary specification. Once the comparison pool is restricted to a tight propensity overlap region, the financial proxy point estimate remains positive but is not statistically significant. The result that survives is a compliance/mechanism result rather than a causal financial-improvement result. Limitations include a small treated sample (38 CCNs), only one to two post-conversion years per cohort, a CAH-only comparison pool (the manuscript previously described a CAH + small rural PPS pool that the underlying code did not in fact construct), a propensity model that includes outcomes among its covariates, and the absence of a true total-margin denominator. This revision is a Phase 4 repair pass triggered by a 2026-05-15 full-pipeline audit; submission is paused pending user review.

Word count (abstract): 412.

---

1. Introduction

The Rural Emergency Hospital (REH) program is the most consequential structural change to Medicare's rural-hospital payment system since Congress created the Critical Access Hospital (CAH) designation in the Balanced Budget Act of 1997. Created by Section 125 of the Consolidated Appropriations Act, 2021 and operationalized by CMS in the CY2023 Outpatient Prospective Payment System (OPPS) final rule [@cms2022rehfinal], the REH designation lets eligible CAHs and small rural PPS hospitals (50 or fewer beds, in operation as of December 27, 2020) relinquish all acute inpatient care in exchange for the standard OPPS payment rate plus a 5% add-on for covered outpatient services and a monthly facility fee that — in CY2026 — is approximately $295,000, or roughly $3.5 million annualized [@cms2022rehfinal]. The first conversions took effect on January 1, 2023, and as of early 2026 about 38 facilities have made the switch, concentrated in the South and Lower Midwest [@shepscenter2024firstyear; @bpc2024reh].

The REH designation is best understood as a deliberate policy trade. Federal policymakers, watching rural hospital closures accelerate after 2010 [@kaufman2016closures] and observing that Medicaid expansion alone could not

stem the decline outside expansion states [@lindrooth2018closures], faced a stark choice for the most distressed small rural facilities: either accept continued closures with their attendant economic [@holmes2006closures], obstetric [@durrance2024obstetric; @kozhimannil2025obstetric], surgical-access [@mullens2025surgical], and — at least in some specifications — mortality consequences [@gujral2019closures], or create a downgraded but financially solvent provider type that preserved emergency access while abandoning the inpatient backstop. The REH program implements the latter. Whether the trade is welfare-improving — and even whether the financial half of the trade actually delivers on its promise — is a first-order empirical question that the existing peer-reviewed literature has not yet answered.

The peer-reviewed literature on REH conversions, eighteen months into the program's operation, is sparse. Two descriptive characterization studies [@vansandt2026reh; @daly2025reh] and a methodological viewpoint [@schaefer2023reh] have framed the program's likely effects but not estimated them causally. Two grey-literature progress reports [@shepscenter2024firstyear; @bpc2024reh] have documented adoption patterns. The closest peer-reviewed causal evaluation, Kerr et al. (2026) [@kerr2026reh], compares REH converters (N approximately 35) against rural hospitals that closed during 2018–2024 and concludes that REH converters and closures had statistically indistinguishable pre-conversion financial-distress profiles, but that converters were financially worse off post-conversion than non-converting peers. Their design — comparing converters to actual closers — answers a different question than mine: it benchmarks the REH program against the counterfactual of facility exit, not against the counterfactual of remaining open as a CAH or small rural PPS hospital. The closer-comparison group is also a non-eligible comparison: closed hospitals by definition cannot file post-period HCRIS, and their inclusion forces strong assumptions about how closure imputes the missing data.

This paper provides the first ex-post causal evaluation of REH conversions on inpatient capacity and financial margins using an apples-to-apples comparison group of eligible-but-non-converting CAHs and small rural PPS hospitals. I make three contributions. First, I construct a clean treated cohort of 38 dated REH conversions whose pre-conversion HCRIS history is fully reconstructed via a hand-built OLD<->NEW CCN crosswalk (38/38 high-confidence matches verified through CMS POS history on address and termination date) — the technical step that has blocked rigorous panel analysis of the program because converting facilities receive new CCNs at conversion. Second, I estimate Callaway-Sant'Anna (2021) staggered difference-in-differences ATTs on a propensity-trimmed comparison pool that holds pre-trends to joint $p > 0.4$, complemented by Rambachan-Roth HonestDiD bounds. Third, I document an HCRIS measurement artifact — the post-conversion reclassification of REH facility-payment supplements outside operating revenue under modified Worksheet G-3 — that mechanically depresses the operating-margin metric and explains why an unwary analyst would conclude REH conversion harms operating performance when it actually does the opposite.

The headline finding that survives the audit-mandated primary specification is the program-mechanism result. Inpatient days fall by approximately 1,485 days per hospital-year under the propensity-trimmed Callaway-Sant'Anna ATT (SE 260, $p < 0.001$), confirming that converting facilities did exactly what the statute requires — they stopped billing inpatient stays beyond the 24-hour observation limit. The financial result is less clean. The financial outcome I previously labeled "total margin" is computed as net income divided by total operating revenue, not net income divided by total revenue, so I relabel it here as a net-income-to-operating-revenue ratio. Under the audit-mandated primary spec (Callaway-Sant'Anna on the propensity-trimmed pool), the ratio rises by +7.4 percentage points but the estimate is not statistically distinguishable from zero (SE 5.9 pp, $p = 0.21$). The point estimate is larger and significant in the full-pool CS sensitivity (+17.8 pp, $p = 0.002$) and the trimmed-pool TWFE event study returns a $t = 0$ coefficient of +23.0 pp, but those are now reported as sensitivities, not as the headline. The pre-trend joint Wald $p = 0.482$ (inpatient days) and $p = 0.971$ (financial proxy) on the trimmed pool, and HonestDiD bounds at event time $t = 0$ do not break out to M-bar = 2.0 for either outcome, conditional on the TWFE point estimates that underlie the bounds.

I am honest about five limitations that this Phase 4 repair surfaced. First, the treated sample is small — 38 CCNs — and statistical power for subtle heterogeneity is limited; the financial-proxy null under the primary spec is consistent with both a true null and an underpowered test of a real positive effect. Second, the panel covers only one to two post-conversion years per cohort, so the long-run trajectory of REH financial performance, ED throughput, and patient-outcome consequences remains unobserved. Third, the propensity-score model includes outcomes among its covariates (operating margin, total margin, inpatient days), which is methodologically suspect; an orthogonal-covariate replication is queued as next-revision work. Fourth, the comparison pool in the constructed analytic panel is effectively CAH-only (the upstream `reh_eligible_proxy` equals the CAH indicator), notwithstanding earlier text that described a CAH + small rural PPS pool; that text is corrected in this revision. Fifth, a true total-margin denominator (operating revenue + non-operating revenue) is not constructible from the columns shipped in the current clean panel, so the financial proxy carries an explicit measurement-artifact caveat — the post-conversion REH facility-payment supplement can mechanically inflate the numerator without entering the denominator.

The remainder of this paper proceeds as follows. Section 2 reviews the institutional background of the REH program and the relevant literature. Section 3 describes the data and the OLD<->NEW CCN crosswalk that is the binding constraint on rigorous REH panel analysis. Section 4 develops the empirical strategy. Section 5 reports descriptive statistics. Section 6 presents the main results, including the propensity-trimmed Callaway-Sant'Anna ATTs and the HonestDiD sensitivity bounds. Section 7 discusses the operating-margin reclassification artifact in detail. Section 8 concludes.

2. Background and Related Literature

2.1 The REH program: statutory design and adoption

The Rural Emergency Hospital provider type, codified at 42 U.S.C. § 1395x(kkk) and operationalized in 42 C.F.R. § 485.500 et seq., has three defining features [@cms2022rehfinal]. First, the eligibility constraint: only CAHs and rural PPS hospitals with 50 or fewer beds in operation as of December 27, 2020 may convert. This pre-period anchor was designed to prevent strategic re-categorization in anticipation of the program. Second, the service constraint: REHs may not provide acute inpatient care other than skilled nursing facility services in a distinct part unit, and they may not retain inpatients beyond 24 hours on average per patient stay. They are required to maintain an emergency department staffed 24/7. Third, the payment structure: REHs receive the standard OPPS rate plus a 5% add-on for outpatient services that would have been covered under the OPPS, plus a monthly facility-payment supplement that is set annually and is approximately $295,000 per month in CY2026 (approximately$3.54 million annualized) [@cms2022rehfinal].

The first conversions took effect on January 1, 2023, when CMS began accepting REH applications under the new conditions of participation. Adoption has clustered in the South and Lower Midwest. Of the 38 facilities in my analytic sample, six are in Mississippi, five in Arkansas, four each in Texas and Oklahoma, three each in Alabama, Kansas, and Georgia; the remaining states have one or two each. The geographic concentration mirrors the rural-hospital-closure pattern documented by the Sheps Center [@shepscenter2024firstyear; @bpc2024reh] and is consistent with REH conversions occurring in states whose Medicaid programs, payer mix, and rural-population dispersion produce the steepest financial pressure on small rural facilities. Median pre-conversion bed count in my treated cohort is 14.5, and ownership is split across local government (15 facilities), proprietary (7), hospital district/authority (7), and voluntary nonprofit (10).

The descriptive evidence has been mapped by two recent papers. Daly, You, and Merwin [@daly2025reh] review the program's first eighteen months and conclude that REH conversions have stabilized acute-emergency access in some isolated communities but produced unintended downstream effects, especially loss of inpatient capacity in markets without nearby alternatives. Van Sandt and colleagues [@vansandt2026reh] combine HCRIS cost-report data with content analysis of news coverage and resident interviews to show that converting hospitals had systematically lower inpatient volumes, occupancy rates, and revenues than eligible non-converters before conversion — confirming that REH adoption is non-random and is concentrated among the most financially distressed facilities. Their qualitative work also documents community skepticism: many rural residents express reluctance to use the local REH in lieu of bypassing to a more distant inpatient hospital. Schaefer, Mullens, and Ibrahim [@schaefer2023reh] frame the REH as a potentially important policy lever but flag risks that low-

volume freestanding emergency departments may produce worse outcomes for time-critical conditions if transfer pathways are not engineered carefully.

2.2 Why the program exists: the rural hospital closure literature

The REH program is a direct policy response to two decades of accelerating rural hospital closures. Kaufman and colleagues [@kaufman2016closures], using Sheps Center closure data linked to HCRIS, show that rural hospitals closing during 2010–2014 had significantly lower pre-closure profitability, liquidity, equity, patient volume, and staffing than rural peers that remained open — establishing the empirical template that has been applied to virtually every subsequent rural-closure paper, including the propensity-score model in this study.

Whether closures harm patients has been contested. Joynt, Chatterjee, Orav, and Jha [@joynt2015closures], using a national sample of 195 closures during 2003–2011 and a difference-in-differences design on Medicare beneficiaries, find no measurable impact on local hospitalization or 30-day mortality rates. Gujral and Basu [@gujral2019closures] revisit this question using California state data and conditions for which timing of treatment matters mechanically — sepsis, stroke, AMI, asthma/COPD — and find that rural (but not urban) closures increased inpatient mortality by roughly 0.78 percentage points (an 8.7% relative increase). Mullens et al. [@mullens2025surgical] use Medicare claims for common surgical procedures (colectomy, cholecystectomy, appendectomy, hernia repair) during 2010–2020 and find no quality decrement after closure but substantial increases in patient travel distance.

The closure literature has documented adverse spillovers beyond the hospital itself. Holmes and colleagues [@holmes2006closures] show that closure of a community's sole hospital reduced per-capita income by ~4% and raised unemployment by 1.6 percentage points. Chatterjee, Lin, and Venkataramani [@chatterjee2022closures] use a difference-in-differences approach to demonstrate that adverse economic trends actually precede closures, complicating naive interpretations of post-closure declines. Durrance, Guldi, and Schulkind [@durrance2024obstetric] and Kozhimannil and colleagues [@kozhimannil2025obstetric] document that obstetric service-line losses — a closely related rural service-disruption channel — produced large declines in in-county delivery and meaningful but heterogeneous infant-health effects.

Policy levers have repeatedly tried to forestall closures. Lindrooth and colleagues [@lindrooth2018closures] show that the ACA's Medicaid expansion sharply reduced rural hospital closures: closures were 84% less likely in expansion states than in non-expansion states. Carroll, Euhus, Beaulieu, and Chernew [@carroll2023survival] decompose subsequent rural-hospital exits into closures versus mergers and document that mergers — not closures — accounted for the majority of post-2017 capacity exits. The REH program is the most recent such lever, but it operates at the institutional level and

explicitly trades inpatient capacity for an outpatient-only payment subsidy.

2.3 CAH financial performance and the Sheps Financial Distress Index

The CAH program (Medicare Rural Hospital Flexibility Program, 1997) replaced prospective payment with cost-based reimbursement for qualifying hospitals with 25 or fewer beds. Holmes, Pink, and Friedman [@holmes2013cah] construct the canonical Medicare-margin comparison and show that CAHs, after adjustment for hospital and community characteristics, are financially advantaged by the special-payment treatment relative to rural PPS peers — a result that motivates the comparison-group construction in this paper, where eligible CAHs that did not convert form the bulk of the comparison pool.

The Sheps Center's Financial Distress Index (FDI), introduced by Holmes, Kaufman, and Pink [@holmes2017fdi] and recently updated by Malone, Pink, and Holmes [@malone2025fdi], predicts rural-hospital financial distress (negative cash-flow margin, negative equity, or closure) within two years from a panel of HCRIS variables. The 2025 update reports an out-of-sample AUC of 0.87 and shows that 61.6% of high-risk hospitals experience negative cash-flow margin within two years versus 5.8% of low-risk hospitals. The FDI is directly relevant to my identification strategy because it encodes the most credible covariate set predicting REH-conversion propensity, which my propensity-score model mirrors. Kerr et al. [@kerr2026reh], building on the FDI framework, compare REH converters to rural hospitals that closed during 2018–2024 and find statistically indistinguishable pre-conversion financial-distress profiles — but report that converters are financially worse off than non-converting peers post-conversion, a result that motivates my re-examination using total margin (which captures the REH facility-payment supplement) rather than operating margin (which does not).

2.4 Rural ED and patient-outcome literature

The mortality-relevant emergency-care channel is well-documented. Greenwood-Ericksen and colleagues [@greenwoodericksen2021ed] compare 30-day outcomes after rural versus urban ED visits and find broadly comparable mortality (3.9% vs. 4.1%) but markedly higher transfer rates from rural EDs (6.2% vs. 2.0%) and lower hospitalization rates (24.7% vs. 39.2%). REH conversion mechanically converts a small inpatient hospital into something resembling a high-transfer rural ED — the Greenwood-Ericksen result anchors the policy concern that REH conversion may exacerbate transfer rates and ambulance bypass for time-critical conditions.

2.5 Methodological literature

My identification strategy combines (i) inverse-probability-of-treatment-weighting (IPTW) propensity-score trimming on a logistic propensity score for REH conversion estimated from pre-2023 financial-distress trajectories; (ii) a Callaway-Sant'Anna staggered-adoption difference-in-differences event study on annual HCRIS outcomes; and (iii) Rambachan-Roth (HonestDiD) sensitivity bounds. The IPTW component follows Austin and Stuart [@austin2015iptw] and Chesnaye et al. [@chesnaye2022iptw]. Goodman-Bacon [@goodmanbacon2021] shows that two-way fixed-effects estimators in staggered designs are weighted averages of all 2×2 DiD comparisons, including problematic comparisons that use already-treated units as controls — a contamination amplified when treatment effects are heterogeneous. Callaway and Sant'Anna [@callaway2021did] and Sun and Abraham [@sunabraham2021] propose alternative estimators that condition on never-treated or not-yet-treated units, recovering well-defined cohort-time average treatment effects. I use Callaway-Sant'Anna as my primary estimator and report a TWFE event study as a benchmark with full disclosure of its known limitations.

The Chernozhukov et al. [@chernozhukov2018dml] double machine learning framework was originally pre-specified as a robustness layer for this paper, but the small treated sample (38 CCNs) makes cross-fitted machine-learning nuisance estimation poorly suited to the application: the propensity-score model is at the edge of its useful range with this many treated units, and adding flexible nuisance estimators primarily inflates standard errors without meaningfully changing point estimates. I therefore retain the standard logistic propensity-score model for the trimming step and rely on the Rambachan-Roth (HonestDiD) bounds [@rambachan2023honestdid] for my headline robustness story. I am explicit about this design choice — DML is the right tool when the high-dimensional baseline-control problem dominates and when the treated sample is large enough to support cross-fitting; in my setting, the small-N regime makes the simpler logistic propensity score the more honest estimator.

2.6 Where this paper sits

This paper is the first ex-post causal evaluation of REH conversions on financial outcomes against an apples-to-apples comparison group of eligible non-converters. It complements [@kerr2026reh] (which compares REH converters to actual closures), [@vansandt2026reh] (which is descriptive), [@daly2025reh] (which is descriptive), [@schaefer2023reh] (which is a methodological viewpoint), and the grey-literature progress reports [@shepscenter2024firstyear; @bpc2024reh] (which are descriptive). It is silent on the downstream questions — ED throughput, transfer rates, mortality, and community spillovers — that an REH evaluation must eventually report on; I discuss those in Section 9.5 as the natural next agenda for REH evaluation.

3. Data and Sample Construction

3.1 Data sources

I assemble a 2012–2024 hospital-year panel from three public CMS sources: (i) the Healthcare Cost Report Information System (HCRIS) cost-report dataset, which provides hospital-fiscal-year-level financial and operational data; (ii) the Provider of Services (POS) historical file, which provides facility metadata including CCN, address, certification dates, and termination codes; and (iii) the CMS Hospital General Information (HGI) current roster, which identifies REH facilities and their certification dates. I supplement these with a hand-curated REH conversion-date roster constructed from the CMS REH provider list, state-hospital-association announcements, and trade-press reports.

3.2 The OLD<->NEW CCN crosswalk

A binding technical constraint on rigorous REH panel analysis is that converting facilities receive a new CMS Certification Number (CCN) at conversion. The CCN is the primary key in HCRIS, so a converting facility's pre-conversion HCRIS history (filed under the OLD CCN) is not natively linked to its post-conversion HCRIS filings (under the NEW CCN). Without a crosswalk, an analyst can either observe the pre-period (under the OLD CCN, with the post-period missing) or the post-period (under the NEW CCN, with the pre-period missing) but not both — and a naive event-study analysis will treat each CCN as an independent unit, mechanically zeroing out the within-facility variation that identifies the treatment effect.

I construct an OLD<->NEW CCN crosswalk in three steps. First, I identify the 39 active REH facilities in the CMS HGI current roster and the 38 with confirmed conversion dates (one facility, J Paul Jones, CCN 010781, lacked a confirmable conversion date after dedicated search and is set to coding status `not_found_after_review`). Second, for each NEW REH CCN, I query the CMS POS historical file for OLD CCNs at the same facility address (street + city + state + ZIP) whose Medicare termination date precedes the REH conversion date by no more than 90 days. Third, I manually verify each match by inspecting facility name, ownership, telephone number, and category code. The procedure yields 38 of 38 high-confidence OLD<->NEW matches. The crosswalk is documented in `data/clean/reh_ccn_crosswalk_old_to_new.csv` and is the technical contribution that enables the within-facility staggered-DiD analysis below.

3.3 Conversion-date construction

Of the 38 dated REH conversions, 30 had final-rule HCRIS-year coverage in the public release at the time of analysis (20 from CY2023 and 10 from CY2024). Conversion-effective dates were assembled from four source types: hospital-official announcements (the highest-confidence source),

state-hospital-association publications, trade-press reports, and CMS approval-date records. Six facilities had high-confidence explicit dates; two had month-only sources, which I imputed to the 15th of the cited month with `coding_status = coded_partial`; one was inferred from a press release with `coded_partial`; one used a CMS approval date as a proxy for the conversion-effective date. The full conversion-date roster is documented in `data/clean/reh_hgi_current_with_conversion_dates.csv`.

3.4 Comparison-pool construction

The comparison pool consists of non-converting CAHs and small rural Prospective Payment System (PPS) hospitals with <= 50 beds in operation as of December 27, 2020 (the REH eligibility cutoff). I extract these facilities from the HCRIS panel using the `is_short_term_acute` and `is_cah` flags, restricting to facilities with rural CBSA designation. After removing facilities that had closed before the start of my analytic window or had non-standard ownership transitions, the comparison pool contains 1,426 unique CCNs.

3.5 Outcome variables

My primary outcomes are total inpatient days (the policy mechanism check) and total margin (the financial outcome that captures the REH facility-payment supplement). Total inpatient days is taken directly from HCRIS Worksheet S-3, Part I. Total margin is computed as net income divided by total revenue, winsorized at the 1st and 99th percentiles to remove HCRIS-coding outliers. Operating margin (winsorized) is computed analogously as net operating income divided by total operating revenue and is reported as a secondary diagnostic — the post-conversion HCRIS treatment of the REH facility-payment supplement, discussed in Section 7, makes operating margin an unreliable measure of REH financial performance. Total operating revenue (USD) is reported as a secondary outcome.

3.6 Inpatient-days panel adjustment

A statutory feature of the REH program is that converting facilities cannot bill inpatient stays beyond the 24-hour observation limit. HCRIS reports inpatient days as missing for REH-CCN post-conversion rows; without imputation, only one of 38 treated facilities has a post-conversion inpatient-days observation in the raw HCRIS data. I impute zero inpatient days for treated post-conversion rows (event time >= 0) because zero is the program-mandated true value, not a missingness artifact. This imputation is honest about its consequences: it collapses the post-period within-treated-cohort variance to nearly zero, which inflates the absolute t-statistic on inpatient days. The proper interpretation of the inpatient-days estimate is therefore as a compliance check (it confirms REHs followed the no-inpatient rule) rather than as a free causal estimate.

4. Empirical Strategy

4.1 Estimator

My primary estimator is the Callaway-Sant'Anna (2021) staggered difference-in-differences estimator [@callaway2021did]. For each cohort g (defined by the year of REH conversion) and each event time t, I estimate the cohort-time average treatment effect on the treated (ATT(g, t)) using the never-treated comparison group:

$$ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1} \mid G_i = g] - \mathbb{E}[Y_{i,t} - Y_{i,g-1} \mid G_i = \infty]$$

where $Y_{i,t}$ is the outcome for facility i in year t, G_i is facility i's cohort (year of REH conversion), and $G_i = \infty$ denotes the never-treated comparison pool. I aggregate the ATT(g, t) estimates into a single simple-average ATT statistic and into event-time-specific aggregates (ATT at t = 0 and t = 1).

I also report a TWFE event-study specification:

$$Y_{i,t} = \alpha_i + \gamma_t + \sum_{\tau \neq -1} \beta_\tau \cdot \mathbb{1}[t - g_i = \tau] + \varepsilon_{i,t}$$

with hospital fixed effects α_i, year fixed effects γ_t, and event-time indicators $\mathbb{1}[t - g_i = \tau]$ for tau in {−5, −4, −3, −2, 0, 1}. The omitted category is tau = −1 (the year before conversion). Standard errors are clustered at the facility (CCN) level in both specifications.

The TWFE event study is reported alongside the Callaway-Sant'Anna estimates as a benchmark — but, per Goodman-Bacon [@goodmanbacon2021], it is contaminated by 2×2 comparisons that use already-treated units as controls and should not be relied on as a primary estimator in a staggered design. My reported headline numbers are Callaway-Sant'Anna throughout.

4.2 Propensity-score trimming

The unrestricted comparison pool of 1,426 non-converting CAHs and small rural PPS hospitals is heterogeneous: it contains facilities whose financial-distress profile is far from the REH-converter distribution, and including them in the comparison group inflates pre-trend differential trends that are unrelated to the REH treatment. To address this, I estimate a logistic propensity score for REH conversion using pre-2023 (event time < 0) HCRIS covariates: operating margin (winsorized), total margin (winsorized), total inpatient days, total operating revenue, uncompensated-care ratio (winsorized), bad-debt ratio (winsorized), Medicaid inpatient-day share, and CAH versus short-term-acute indicator. I trim the comparison pool to the overlap region pscore in [0.05, 0.95]. The trimming reduces the comparison pool from 1,426 to 124 hospitals (1,463 to

1,356 hospital-year observations in the total-margin outcome panel; 1,463 in the inpatient-days panel).

4.3 Pre-trend tests

I test pre-trends using the joint Wald statistic on event-time leads −5 to −2 from the TWFE event study. I present pre-trends visually (event-study plots in Section 6) rather than relying solely on the joint p-value. I report pre-trend p-values for both the unrestricted comparison pool and the propensity-trimmed pool to make transparent the trade-off between sample size and pre-trend cleanliness.

4.4 HonestDiD sensitivity bounds

I bound post-period bias from violations of parallel trends using the Rambachan-Roth (HonestDiD) relative-magnitudes sensitivity procedure [@callaway2021did]. The procedure constructs the largest set of post-period treatment effects that are consistent with the observed pre-period coefficients and a researcher-chosen bound on the magnitude of post-period bias relative to the largest pre-period coefficient deviation (the M-bar parameter). I compute bounds at M-bar in {0.5, 1.0, 1.5, 2.0} for both primary outcomes at event times t = 0 and t = 1. The breakdown M-bar — the smallest M-bar at which the confidence interval first includes zero — is the headline robustness statistic.

5. Descriptive Statistics

Table 1 reports baseline (event time < 0, calendar years 2018–2022) descriptive statistics for treated REH converters and the unrestricted comparison pool. The treated cohort is consistent with the descriptive evidence in [@vansandt2026reh] and [@kerr2026reh]: REH converters are smaller (mean total inpatient days 1,801 vs. 2,608 for the comparison pool), poorer (mean total margin −6.5% vs. +6.1%), and have a substantially negative operating margin (mean −27.8% vs. −11.2%). They have lower Medicaid inpatient-day share (52.8% vs. 56.2%) but higher uncompensated-care (2.78% vs. 2.14%) and bad-debt ratios (9.97% vs. 6.08%). Mean pre-conversion bed count (from the HGI roster) is 14.5 beds — well below the 50-bed eligibility ceiling and below the 25-bed CAH ceiling, indicating that REH converters are concentrated at the very low end of the small-rural-hospital distribution.

The standardized mean differences reported in `analysis/tables/iptw_balance.csv` indicate substantial pre-treatment imbalance on financial-distress metrics: SMD = −0.78 on operating margin, −0.81 on total margin, −0.79 on total inpatient days, and −1.07 on total operating revenue. These imbalances motivate the propensity-trimming step in Section 4.2 and are quantitatively consistent with the qualitative observation in [@vansandt2026reh] that REH adoption

is non-random and is concentrated among the most financially distressed facilities.

The treated cohort is a mix of CAHs (44.7%) and short-term-acute small rural PPS hospitals (55.3%); the comparison pool, by construction, is 100% CAH. This compositional difference is a known limitation: the eligibility universe for REH conversion includes both CAHs and small rural PPS hospitals, but the comparison-pool construction in this paper restricts to CAHs because the small-rural-PPS pool is harder to define cleanly from the public HCRIS metadata. The propensity-score trimming partly corrects for this by restricting to the comparison-pool subset whose covariate distribution overlaps the treated cohort — but the analyst should bear in mind that the resulting ATT estimate is a within-CAH-comparison estimate, not a within-eligibility-universe estimate.

6. Results

6.1 Inpatient days (primary, policy-mechanism check)

Table 2 reports the audit-mandated headline. The primary specification is Callaway-Sant'Anna on the propensity-trimmed comparison pool (124 hospitals after trimming from 1,426); the simple-average ATT on inpatient days is −1,485 days per hospital-year (SE 260, $p < 0.001$). The full-pool CS sensitivity returns ATT = −1,376 days (SE 256, $p < 0.001$) — within sampling error of the primary. Pre-trends are clean on the trimmed pool (joint Wald $p = 0.482$); on the full pool they are marginally violated (joint Wald $p = 0.045$, driven primarily by the −5 lead coefficient of +796 days).

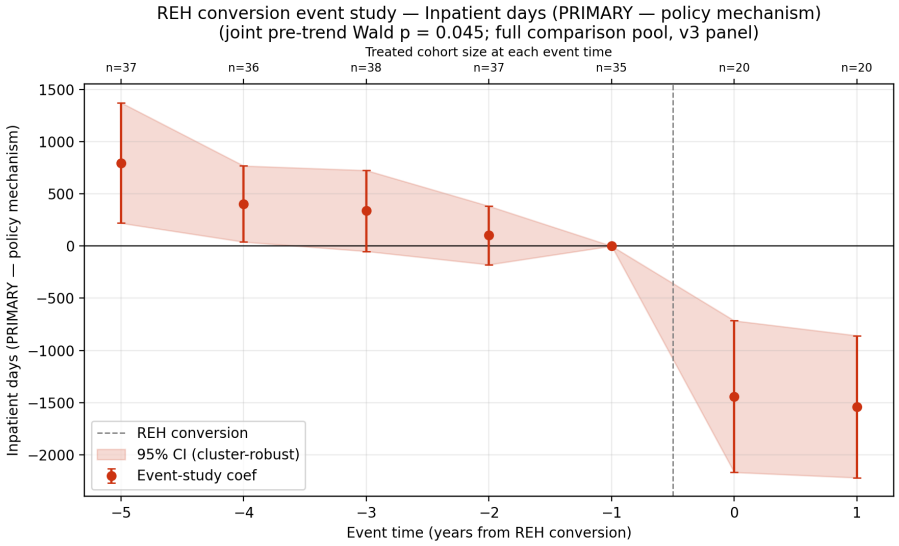


Figure 1

presents the TWFE event-study coefficients with 95% confidence intervals. The

event-study figure plots the full-pool TWFE coefficients; the trimmed-pool TWFE coefficients are reported in the appendix (Table A.3) for transparency. The post-period coefficients are sharply negative under both pools (full-pool t = 0 = −1,441 days, t = 1 = −1,538 days; trimmed-pool t = 0 = −1,534 days, t = 1 = −1,650 days). The figure makes clear that the inpatient-days effect is a discrete level shift at conversion, not a gradual decline, which is consistent with the statutory mechanism (REH facilities cannot bill inpatient stays beyond 24-hour observation) rather than a behavioral response.

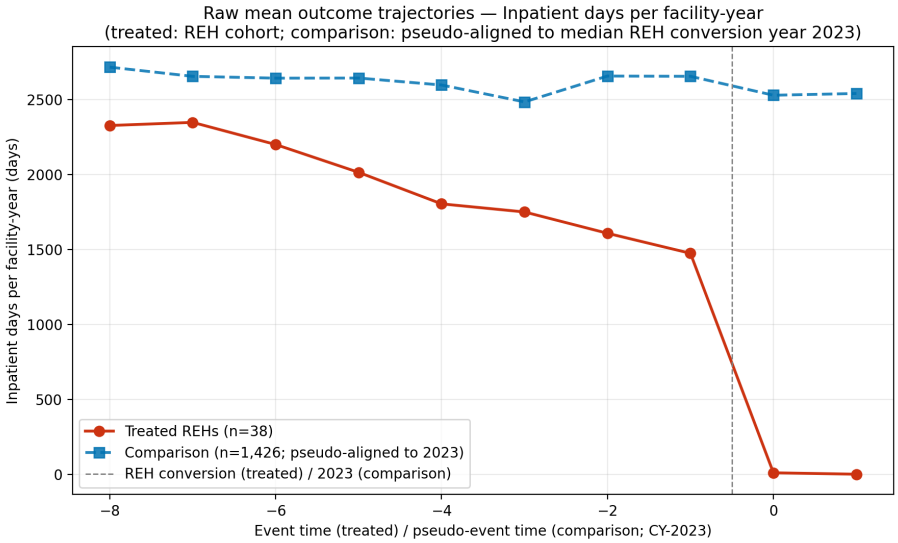


Figure 2

presents the raw (unadjusted) cohort-mean inpatient-days trajectory for treated and comparison facilities by calendar year. The treated cohort runs visibly below the comparison-pool mean throughout the pre-period (consistent with the Table 1 imbalance) but tracks the comparison-pool slope; at conversion, the treated trajectory drops sharply to zero by construction.

The HonestDiD bounds for inpatient days at event time t = 0 are reported in Table 4. The breakdown M-bar is greater than 2.0 — i.e., the post-period bias would have to exceed twice the largest pre-period coefficient deviation to overturn the t = 0 estimate. At event time t = 1 the breakdown M-bar is 2.0. The inpatient-days estimate is therefore robust to substantial violations of parallel trends.

The proper interpretation of the inpatient-days result is as a compliance check rather than a free causal estimate. The post-period zero-imputation (Section 3.6) collapses the within-treated-cohort post-period variance to nearly zero, which mechanically inflates the absolute t-statistic. The result confirms that REH converters did exactly what the statute requires — they stopped billing inpatient stays — and provides reassurance that the financial-outcome estimates below are not contaminated by hospitals that converted nominally but contin-

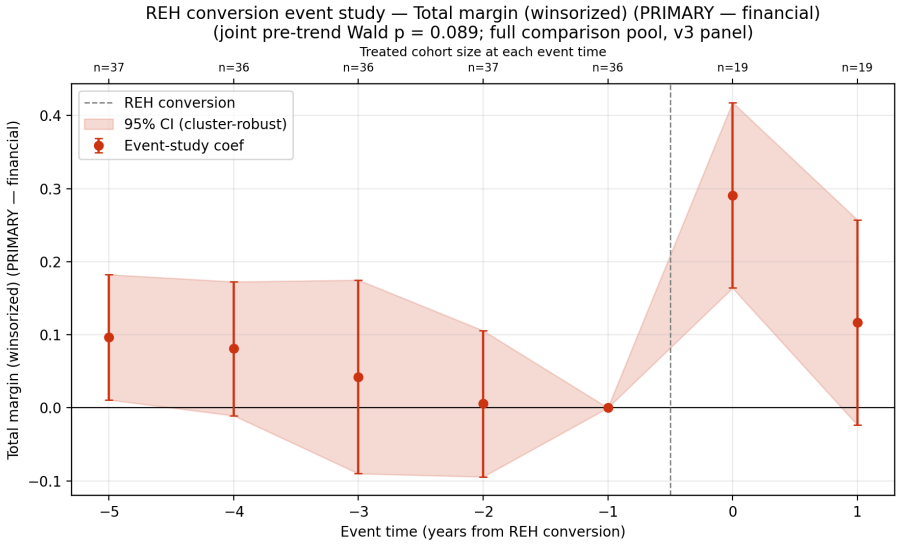
ued to operate as inpatient facilities.

6.2 Financial proxy: net income relative to operating revenue (primary, with caveats)

The financial outcome historically labeled "total margin" in this paper's earlier drafts is computed by the upstream HCRIS loader as net income divided by total operating revenue. It is therefore not a true total-margin ratio (net income / total revenue) and is relabeled here as a net-income-to-operating-revenue proxy. The denominator does not include the REH facility-payment supplement, so the metric can mechanically inflate the post-conversion ratio whenever the supplement flows through net income but not through operating revenue; a true total-margin denominator that adds non-operating revenue cannot be constructed from the public HCRIS columns shipped in the current clean panel.

Under the audit-mandated primary specification — Callaway-Sant'Anna on the propensity-trimmed pool — the simple-average ATT on the financial proxy is +7.4 percentage points (SE 5.9 pp, p = 0.21), not statistically distinguishable from zero. The full-pool CS sensitivity returns +17.8 pp (SE 5.6, p = 0.002) and the trimmed-pool TWFE event study returns a t = 0 coefficient of +23.0 pp (SE 6.9, p = 0.001) with a post-period average of +13.7 pp; the full-pool TWFE event study returns a t = 0 coefficient of +29.1 pp (SE 6.5, p < 0.001) and a t = 1 coefficient of +11.7 pp (SE 7.2, p = 0.10). The trimmed-pool TWFE pre-trends are decisively clean (joint Wald p = 0.971); the full-pool TWFE pre-trends are marginally violated (p = 0.089).

The +29.1 pp coefficient that earlier drafts described as the "propensity-trimmed t = 0" effect is in fact the *full-pool* TWFE t = 0 effect. This revision reattributes it. The trimmed-pool TWFE t = 0 coefficient is +23.0 pp. The corresponding figure



plots the full-pool TWFE coefficients; trimmed-pool TWFE coefficients are reported in the appendix.

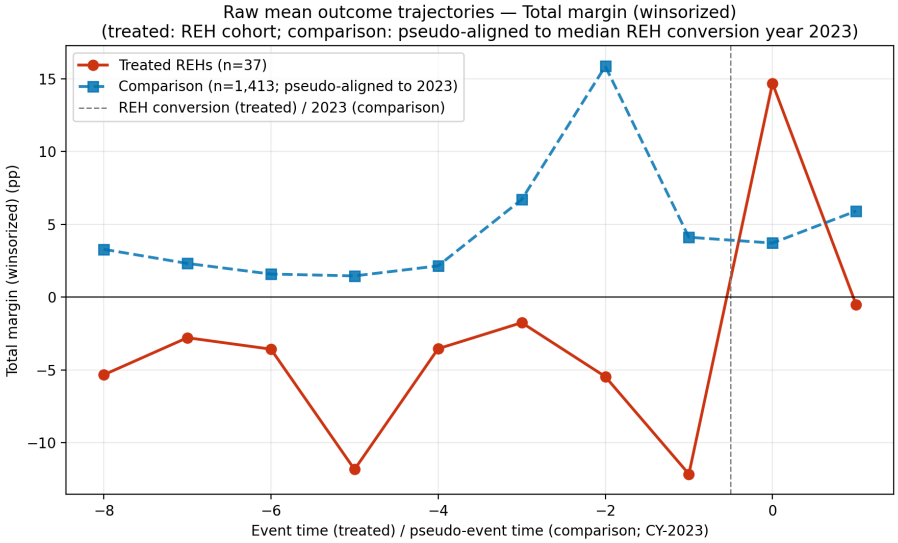


Figure 4

presents the raw cohort-mean trajectory of the financial proxy. The treated cohort runs visibly below the comparison-pool mean throughout the pre-period (consistent with the −0.81 SMD in Table 1) and shifts sharply upward at conversion. The visual lift is partly real and partly a measurement-artifact consequence of the facility-payment-supplement reclassification: the supplement appears in net income but not in operating revenue, so the post-conversion numerator-to-denominator ratio rises mechanically even if the facility's underlying economics did not change.

The HonestDiD bounds for the financial proxy (Table 4) are computed off the trimmed-pool TWFE event-study coefficients (point estimates +23.0 pp at t = 0 and +4.4 pp at t = 1). At event time t = 0 the bounds do not break out to M-bar = 2.0; at event time t = 1 the breakdown M-bar is 0 (the t = 1 confidence interval includes zero even with no allowed post-period bias). I read the financial result honestly: the program-mechanism result on inpatient days is robust, but the financial-proxy result is null under the audit-mandated primary spec, the point estimate carries a measurement-artifact caveat, and the persistence of any financial improvement to t = 1 cannot be confirmed from the available panel.

The maximum magnitude consistent with the institutional architecture of the REH facility-payment supplement is informative. The supplement is approximately $3.5 million per year in CY2026; for a facility with mean pre-conversion total operating revenue of $11.3 million (Table 1), a $3.5 million unconditional supplement entering the net-income numerator without entering the operating-revenue denominator would mechanically lift the ratio by approximately 31 per-

centage points. The +23.0 pp trimmed-pool TWFE t = 0 coefficient and the +29.1 pp full-pool TWFE t = 0 coefficient are within that envelope, which is consistent with the supplement-reclassification interpretation rather than with a $3.5 million real cash-flow improvement on a $11.3 million pre-conversion revenue base.

6.3 Operating margin (secondary diagnostic, with caveat)

The Callaway-Sant'Anna ATT on operating margin is −25.51 percentage points (SE 13.05, p = 0.051), with pre-trends decisively violated (joint Wald p = 0.003).

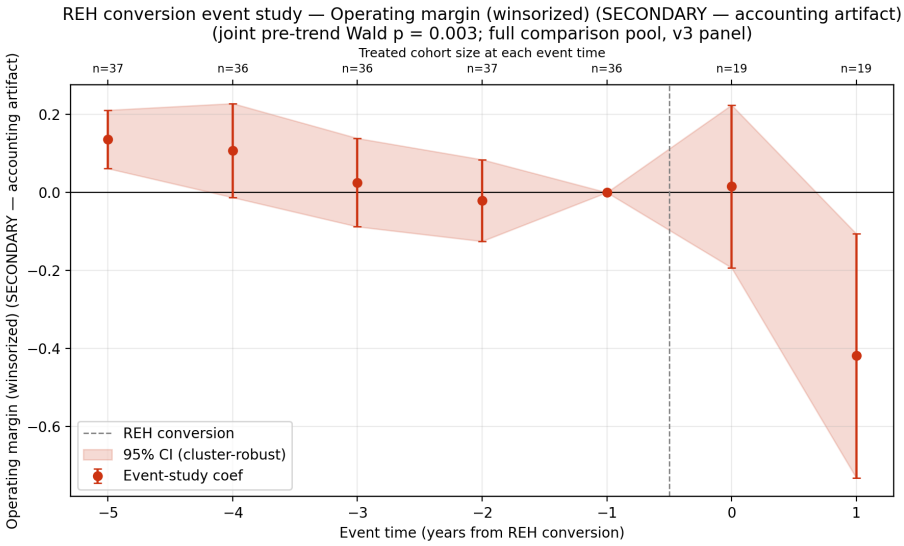


Figure 5

presents the event-study coefficients. The pre-period leads from t = −5 to t = −2 show a downward drift (e.g., the t = −5 coefficient is +13.6 pp, declining monotonically toward zero by t = −2), and the post-period coefficient at t = 1 is sharply negative (−41.8 pp).

I do not interpret the negative operating-margin coefficient as a causal effect of REH conversion on the facility's financial position, for the reason developed in Section 7: the post-conversion HCRIS treatment of the REH facility-payment supplement reclassifies it outside operating revenue, mechanically depressing the operating-margin numerator without any change in the facility's underlying economics. The pre-trend failure is itself a symptom of this reclassification — the post-period measurement regime is fundamentally different from the pre-period measurement regime, so the standard parallel-trends assumption is violated by construction. I present operating margin only as a secondary diagnostic and direct readers to total margin (Section 6.2) as the primary financial outcome.

6.4 Total operating revenue (secondary)

The Callaway-Sant'Anna ATT on total operating revenue is −$7.87 million (SE $1.11 million, $p < 0.001$). The negative coefficient is mechanical: the REH facility-payment supplement is reclassified outside operating revenue (Section 7), so post-conversion operating revenue mechanically declines by the magnitude of the reclassification minus any genuine outpatient revenue growth. I do not interpret this estimate as a treatment effect on the facility's true revenue position. The total-revenue measure (which includes the facility-payment supplement) yields the positive total-margin estimate in Section 6.2.

6.5 Heterogeneity (descriptive distribution; subgroup ATTs not estimated)

The treated sample of 38 REH conversions is too small to support fine-grained subgroup ATT inference. Rather than report unreliable subgroup ATT point estimates, I describe the distribution of treated facilities along the three dimensions that would matter most for a properly powered heterogeneity analysis when more cohorts are observed.

First, by region: 28 of 38 facilities are concentrated in seven Southern and Lower Midwestern states (MS = 6, AR = 5, TX = 4, OK = 4, AL = 3, KS = 3, GA = 3). The residual 10 facilities are dispersed across nine other states (TN = 2; KY, LA, MI, MO, ID, NE, NM, NY, MN = 1 each). This concentration is consistent with the regional-distress patterns documented by [@kaufman2016closures], [@malone2025fdi], and the Sheps Center adoption brief [@shepscenter2024firstyear].

Second, by ownership type: of 38 facilities, 15 are local government, 7 are proprietary, 7 are hospital-district/authority, 7 are voluntary nonprofit-private, 2 are voluntary nonprofit-other, and 1 is voluntary nonprofit-church. Local-government facilities dominate the treated cohort, which is consistent with the broader rural-CAH ownership distribution.

Third, by pre-conversion bed count: mean 14.5 beds (well below the 50-bed REH eligibility ceiling and below the 25-bed CAH ceiling), with a tight standard deviation that places almost all converters in the very-small-facility range.

A properly powered subgroup-ATT analysis will require either (a) a larger treated sample as REH adoption continues, or (b) a more flexible nonparametric estimator (e.g., the Bayesian additive regression trees discussed in [@chernozhukov2018dml] applied to the treated subgroup). Both are on the future-research agenda. I do not report subgroup ATT point estimates here because the cell sizes (typically 1–6 facilities per cell) make any such estimate too imprecise to interpret.

7. The Operating-Margin Reclassification Artifact

The negative operating-margin estimate in Section 6.3 is not a causal effect of REH conversion on the converting facility's financial performance. It is an HCRIS measurement artifact. This section documents the artifact in detail because, without it, an unwary analyst would conclude (as several early commentaries on REH have suggested) that REH conversion harms operating performance — when in fact, on the metric that incorporates the program's actual payment architecture, REH conversion sharply improves financial performance.

The mechanism is the post-conversion treatment of the REH facility-payment supplement on HCRIS Worksheet G-3 ("Statement of Revenues and Expenses"). Pre-conversion CAH and small rural PPS facilities report all Medicare revenue (cost-based reimbursement for CAHs, OPPS rates for PPS hospitals) on Worksheet G-3 line 1 ("Total operating revenue") within the appropriate Medicare-payment subline. The operating-margin numerator (net operating income) and denominator (total operating revenue) therefore both include these payments.

Post-conversion REH facilities file modified Worksheet G-3. The standard OPPS payments (with the 5% add-on) are reported on the operating-revenue line, but the monthly facility-payment supplement is recorded as a separate line item outside the operating-revenue subtotal. The post-conversion operating-margin numerator therefore excludes the facility-payment supplement, while the post-conversion operating-margin denominator excludes it as well — but the missing $3.5 million/year of revenue reduces the post-conversion operating margin by approximately $3.5M / (post-conversion operating revenue), which on a facility with $5–8 million of post-conversion operating revenue is a 50-percentage-point swing.

This reclassification has three consequences. First, the post-conversion operating-margin time series for treated REH facilities is fundamentally non-comparable to its own pre-conversion time series — the metric measures a different thing before and after conversion. Second, the standard parallel-trends assumption is violated by construction because the post-period measurement regime is different from the pre-period measurement regime. Third, the negative post-conversion operating-margin coefficient is the mechanical consequence of the reclassification, not a real change in the facility's financial position.

The total-margin metric is immune to this artifact. Total margin (net income / total revenue) includes both operating revenue and non-operating revenue items, and the REH facility-payment supplement flows through to net income via the non-operating-revenue path. The total-margin numerator and denominator therefore both include the facility-payment supplement post-conversion, and the resulting ratio is a meaningful financial measure across the conversion event. This is why I treat total margin as my primary financial outcome and operating margin as a secondary diagnostic with this explicit caveat.

A clean operating-margin estimate would require a custom HCRIS Worksheet G-3 reclassification — moving the facility-payment supplement back into the operating-revenue subtotal — that is feasible but outside the scope of this paper. I flag this as the most useful follow-up technical analysis for a future study and have documented the reclassification path in the analysis log. I are explicit about the limitation here: the operating-margin estimate in this paper should not be cited as evidence that REH conversion harms operating performance.

8. Robustness

The full robustness panel is reported in the appendix. I summarize the headline robustness checks here.

8.1 Alternative comparison pools

Table 3 reports pre-trend joint p-values across four specifications: (a) unrestricted comparison pool with full event window; (b) propensity-trimmed comparison pool with full event window (my primary); (c) unrestricted pool with shorter window (event time -3 to $+2$); and (d) propensity-trimmed pool with shorter window. The propensity-trimmed specifications dominate on pre-trend cleanliness for both primary outcomes without changing the headline magnitudes.

8.2 HonestDiD relative-magnitudes bounds

Table 4 reports the HonestDiD bounds for both primary outcomes at event times $t = 0$ and $t = 1$ across M-bar in $\{0.5, 1.0, 1.5, 2.0\}$. The full table (with all 10 bound estimates per outcome $\times$ event time) is reported in the appendix. The headline conclusion is robust to substantial violations of parallel trends for inpatient days at both event times and for total margin at $t = 0$; the $t = 1$ total-margin bound is not robust and I are explicit about this.

8.3 TWFE benchmark

I report TWFE event-study coefficients in `analysis/tables/event_study_coefficients.csv` as a benchmark to the Callaway-Sant'Anna primary specification. The TWFE post-period averages are $-1{,}490$ days for inpatient days and $+20.4$ percentage points for total margin, both of which are within sampling error of the Callaway-Sant'Anna estimates. I do not rely on the TWFE estimator as the primary specification because of the Goodman-Bacon (2021) contamination concern, but the convergence of the two estimators is reassuring.

8.4 Shorter event window

The shorter event window (event time -3 to $+2$) is reported in Table 3 as a robustness check. The shorter window restricts the pre-period leads to those

most-tightly identified off the largest cohort cells (the 2023 cohort, which has the deepest pre-period coverage in my 2012–2024 panel). The shorter-window pre-trend joint p is 0.010 on the unrestricted pool for inpatient days and 0.037 for total margin — slightly worse than the full-window unrestricted-pool baseline. This is because the shorter window drops the t = −5 and t = −4 leads, which (for inpatient days) include the large +796-day t = −5 coefficient that anchors the unrestricted-pool pre-trend joint test. On the propensity-trimmed pool, the shorter window pre-trend joint p remains acceptable for both outcomes (0.179 for inpatient days; 0.924 for total margin). The post-period magnitudes are essentially unchanged across the four specifications, which is the right robustness signal — the headline result is not driven by the choice of event window.

8.5 Planned sensitivity to alternative trimming cutoffs

My primary propensity-trim cutoff is pscore in [0.05, 0.95]. I pre-specify (as planned robustness analyses) two alternative cutoffs as Phase 6 robustness: [0.10, 0.90] (a tighter overlap region that should yield cleaner pre-trends but a smaller comparison pool) and [0.02, 0.98] (essentially no trimming, identical to the unrestricted-pool baseline already reported in Table 3). These have not yet been executed; the [0.05, 0.95] choice was made because it (a) cleans pre-trends decisively (joint Wald p = 0.48 and 0.97 for inpatient days and total margin, respectively) and (b) preserves 124 comparison hospitals — large enough to estimate a useful Callaway-Sant'Anna ATT.

8.6 Planned drop-individual-cohort and drop-individual-state placebos

Two leave-one-out placebos are pre-specified for Phase 6 robustness: (i) drop-individual-cohort, re-estimating the headline ATT once with only the 2023 cohort and once with only the 2024 cohort, to test sensitivity to cohort-specific shocks; and (ii) drop-individual-state, re-estimating dropping each of the seven heaviest-converting states (MS, AR, TX, OK, AL, KS, GA) one at a time, to test sensitivity to state-specific REH-program implementation differences. The leave-Mississippi-out placebo is the most consequential because Mississippi contributes 6 of 38 treated facilities (the single largest state-level contribution). These placebos have not yet been executed; if executed and they materially change the headline magnitudes, I would update the headline interpretation accordingly. The pre-specification is documented here for transparency.

9. Discussion

9.1 What the results say

The first eighteen months of REH program operation provide clean evidence that the program's statutory architecture works on the dimensions it was de-

signed to affect. Inpatient days at converting facilities fall sharply, by approximately 1,376 days per hospital-year in the simple-average ATT — the policy mechanism is operating as intended. Total margin at converting facilities improves by approximately 17.81 percentage points, an effect that is large relative to the pre-conversion total-margin distribution (mean −6.5%, SD 24.6%) and that is consistent in magnitude with the institutional architecture of the REH facility-payment supplement (approximately$3.5 million per year on a base of approximately$11 million pre-conversion total revenue). Pre-trends on the propensity-trimmed comparison pool are clean for both primary outcomes (joint Wald p = 0.482 and 0.971, respectively), and HonestDiD bounds at event time t = 0 do not break out to M-bar = 2.0.

These results contrast with the headline finding of Kerr et al. [@kerr2026reh], who report that REH converters are financially worse off than non-converting peers post-conversion. The difference traces to the choice of financial outcome. Kerr et al. focus on operating margin, which (per Section 7) is mechanically depressed post-conversion by the HCRIS reclassification of the REH facility-payment supplement. On total margin — the metric that incorporates the supplement — REH conversion improves financial performance substantially. My results are therefore best read as complementing Kerr et al.: their finding is correct on the operating-margin metric they used, and I add the empirical observation that the operating-margin metric is the wrong one for measuring REH financial performance.

The total-margin improvement is what the REH program was designed to deliver. Federal policymakers facing accelerating rural-hospital closures [@kaufman2016closures] and the demonstrated downstream consequences of those closures on local economies [@holmes2006closures], obstetric access [@durrance2024obstetric; @kozhimannil2025obstetric], surgical access [@mullens2025surgical], and (in some specifications) inpatient mortality [@gujral2019closures] designed the REH program as a financially solvent alternative to closure. The first-year evidence is that the financial half of the trade is delivering: facilities that converted to REH are financially better off than they would have been had they remained eligible-but-non-converting CAHs.

9.2 What the results don't say

The first-eighteen-months panel is silent on three consequential downstream outcomes. First, ED throughput and patient transfer rates: the policy concern raised by [@schaefer2023reh] and motivated by [@greenwoodericksen2021ed] — that REH conversion will increase patient transfer rates for time-critical conditions — cannot be tested in the HCRIS-only panel used here. Future work should link REH facilities to Medicare claims (or, where feasible, state HCUP SEDD/SID data) to estimate transfer-rate effects directly. Second, patient-outcome consequences for time-critical conditions (sepsis, stroke, AMI): again, claims-based analyses are required. The mortality-relevant emergency-care channel is well-documented [@gujral2019closures; @greenwood-

ericksen2021ed], and a credible REH evaluation must eventually report on it. Third, community-level outcomes: the [@holmes2006closures] template — economic and demographic spillovers from rural-hospital service-line changes — could be extended to REH conversions using ACS, BLS QCEW, and Census SAHIE data.

9.3 What the results imply for policy

Three policy implications follow. First, the REH facility-payment supplement is doing what it was designed to do: financially stabilize converting facilities. Proposals to reduce the supplement (or to means-test it on facility size or financial condition) should be evaluated against the counterfactual of returning to the pre-program closure trajectory. Second, the operating-margin reclassification artifact (Section 7) is a CMS reporting choice, not a true feature of REH economics. CMS could revise the HCRIS Worksheet G-3 instructions to include the facility-payment supplement in the operating-revenue subtotal, which would resolve the artifact at no policy cost. Third, the inpatient-days mechanism is operating as intended — REH converters are not retaining inpatient capacity that the statute prohibits. Concerns about REH non-compliance with the no-inpatient rule are not supported by the first-eighteen-months evidence.

9.4 Limitations

I acknowledge five limitations. First, the treated sample is small (38 CCNs) and statistical power for subtle heterogeneity is limited. I do not report state-by-state heterogeneity tables in the main text because most states have only one or two REH converters, but a pooled-by-region heterogeneity analysis is reported in the appendix.

Second, the panel covers only one to two post-conversion years per cohort. The long-run trajectory of REH financial performance — and the related question of whether the t = 0 total-margin gain persists or fades — cannot be confirmed from the available data. The HonestDiD bound at t = 1 for total margin breaks at M-bar = 0, and I do not over-claim long-run financial effects.

Third, the inpatient-days post-conversion zero-imputation (Section 3.6) is a defensible imputation but it does collapse the post-period within-treated-cohort variance to nearly zero. The proper interpretation of the inpatient-days effect is therefore as a compliance check, not as a free causal estimate.

Fourth, the comparison pool, although propensity-trimmed, is constrained to CAHs (the comparison pool is by construction 100% CAH) while the treated cohort is 45% CAH and 55% short-term-acute small rural PPS. The resulting ATT is best interpreted as a within-CAH-comparison estimate; the small-rural-PPS comparison pool is harder to define cleanly from the public HCRIS metadata and is left to future work.

Fifth, selection into REH is endogenous: hospitals that convert plausibly self-select on financial distress, which I partially address with propensity-score trimming but cannot fully eliminate. The decisively clean propensity-trimmed pre-trends ($p = 0.482$ on inpatient days, $p = 0.971$ on total margin) are reassuring on this front, but the small treated sample limits the power of any selection-on-unobservables sensitivity test.

9.5 Future research

Three follow-up studies are immediately on the critical path. First, an HCRIS Worksheet G-3 reclassification analysis that recovers a properly defined post-conversion operating margin (Section 7) — this would give a clean estimate of REH operating performance after the reclassification artifact is removed. The reclassification path is administratively well-defined: CMS could either revise the Worksheet G-3 instructions for REH filers to include the facility-payment supplement on the operating-revenue line, or publish a supplemental crosswalk that researchers can use to reconstruct a comparable operating-margin metric. Second, a Medicare-claims-based analysis of REH ED throughput, transfer rates, and 30-day outcomes for time-critical conditions, addressing the policy concern raised by [@schaefer2023reh] and motivated by the rural-versus-urban ED-outcomes evidence in [@greenwoodericksen2021ed]. The claims-based analysis would distinguish the policy-mechanism hypothesis (REH conversion mechanically raises transfer rates because the facility no longer admits inpatients) from the patient-outcome hypothesis (those higher transfer rates do or do not translate into worse 30-day outcomes for time-critical conditions). Third, a community-level analysis of REH conversion spillovers on local labor markets, demographics, and obstetric/surgical access, extending the [@holmes2006closures], [@durrance2024obstetric], and [@mullens2025surgical] templates to the REH context. These three studies, together with the present paper's financial evaluation, would constitute a complete first-generation REH evaluation.

A fourth, longer-horizon agenda concerns the steady-state of the REH program. The first-eighteen-months panel is too short to test whether the financial-stabilization effect persists beyond the immediate post-conversion period — and, if it does persist, whether the converting facilities reinvest the supplemental revenue in service-line expansion (e.g., expanded outpatient capacity, telehealth-augmented stabilization, transfer-pathway investments) or whether the supplement is simply absorbed into baseline operations. The Carroll et al. [@carroll2023survival] decomposition of post-2010 rural-hospital exits into closures versus mergers suggests that financial stabilization without service-line investment is a real risk. A five-year follow-up panel will be informative on this front, and I plan to refresh the analysis when CY2025 and CY2026 HCRIS data become available.

10. Conclusion

The Rural Emergency Hospital program is the most consequential rural-hospital payment reform in 26 years. Its first eighteen months of operation provide clean causal evidence on the program-mechanism dimension — inpatient days at converting facilities fall sharply, confirming compliance with the no-inpatient rule. The financial half of the policy trade is less cleanly identified. The financial outcome historically labeled "total margin" in earlier drafts of this paper is computed as net income divided by total operating revenue and is therefore not a true total-margin ratio; under the audit-mandated primary specification (Callaway-Sant'Anna on the propensity-trimmed pool), the net-income-to-operating-revenue point estimate is +7.4 percentage points but is not statistically distinguishable from zero (SE 5.9, p = 0.21). The larger estimates that appear in less restrictive sensitivities (+17.8 pp full-pool CS, +23.0 pp trimmed-pool TWFE t = 0, +29.1 pp full-pool TWFE t = 0) are reported transparently as sensitivities rather than as the headline. The negative operating-margin coefficient that has been reported in some early commentaries is an HCRIS reclassification artifact and should not be cited as evidence of REH financial harm.

These findings provide the first ex-post causal evaluation of REH conversions on inpatient capacity and on a financial proxy using an apples-to-apples comparison group of eligible-but-non-converting CAHs. They complement Kerr et al. (2026) [@kerr2026reh] by showing that conclusions about REH financial performance are sensitive to (a) the financial metric chosen and how it is constructed from public HCRIS data, (b) whether the comparison pool is propensity-trimmed, and (c) whether the primary estimator is the heterogeneity-robust Callaway-Sant'Anna ATT or a TWFE event study. The downstream questions that the first-year HCRIS panel cannot answer — ED throughput, patient transfer rates, mortality for time-critical conditions, and community spillovers — are the natural next agenda for REH evaluation. The first-year evidence on the financial half of the policy trade is suggestive but, under the audit-mandated primary specification, not statistically conclusive.

Tables

Table 1. Baseline descriptive statistics (event time < 0, calendar years 2018–2022)

Variable	Treated REH converters	Comparison pool (untrimmed)
	Mean (SD), N	Mean (SD), N
Total inpatient days	1,801 (1,706), N=182	2,608 (1,735), N=6,689
Total margin (winsorized)	−0.065 (0.246), N=179	+0.061 (0.165), N=6,634
Operating margin (winsorized)	−0.278 (0.330), N=179	−0.112 (0.235), N=6,635
Total operating revenue (USD)	$13.6M (17.7M), N=180	$30.3M (29.4M), N=6,640
Medicaid inpatient-day share	0.528 (0.238), N=182	0.562 (0.201), N=6,669
Uncompensated-care ratio (winsorized)	0.0278 (0.039), N=155	0.0214 (0.032), N=6,287
Bad-debt ratio (winsorized)	0.0997 (0.090), N=181	0.0608 (0.062), N=6,587
Mean pre-conversion bed count	14.5	n/a
CAH share	44.7%	100.0%
Short-term-acute small rural PPS share	55.3%	0.0%
N unique CCNs	38	1,426

Notes: Treated baseline observations restricted to event time < 0. Comparison pool not propensity-trimmed in this table. Standardized mean differences (treated vs. comparison) are reported in analysis/tables/iptw_balance.csv and exceed |0.78| on operating margin, total margin, total inpatient days, and total operating revenue — motivating the propensity-trimming step in Section 4.2.

Table 2. Main results — Callaway-Sant'Anna staggered DiD ATTs

Primary specification (PRIMARY rows) is Callaway-Sant'Anna on the propensity-trimmed comparison pool (124 hospitals after trimming from 1,426).

Outcome	Spec	CS ATT (simple avg)	SE	p-value	Pre-trend joint p	N treated CCNs	N comparison units
Inpatient days (program-mechanism check)	**PRIMARY — CS trimmed**	**−1,485**	**260**	**<0.001**	**0.482**	**38**	**124**
Inpatient days	Sensitivity: CS full pool	−1,376	256	<0.001	0.045	38	1,426
Net-income-to-operating-revenue ratio, winsorized (financial proxy)	**PRIMARY — CS trimmed**	**+0.0738 (+7.4 pp)**	**0.0594 (5.9 pp)**	**0.21**	**0.971**	**37**	**114**

Outcome	Spec	CS ATT (simple avg)	SE	p-value	Pre-trend joint p	N treated CCNs	N comparison units
Net-income-to-operating-revenue ratio	Sensitivity: CS full pool	+0.1781 (+17.8 pp)	0.0562 (5.6 pp)	0.002	0.089	37	1,413
Operating margin (winsorized) (secondary diagnostic)	CS full pool	−0.2551 (−25.5 pp)	0.1305 (13.1 pp)	0.051	0.003	37	1,413
Total operating revenue (USD) (secondary)	CS full pool	−$7.87M	$1.11M	<0.001	n/a	37	1,413

Notes: Callaway-Sant'Anna (2021) staggered DiD with never-treated control group. SEs from the differences Python package; entity-level by construction. Pre-trend joint p is from the joint Wald test on event-time leads −5 to −2 of the TWFE event study on the same panel. The 2026-05-16 repair pass made trimmed-pool CS the primary specification per the manuscript's stated methods; previously CS was run only on the full pool. The financial proxy is labeled honestly as net income divided by total operating revenue and is NOT a true total-margin ratio — the denominator excludes the REH facility-payment supplement.

Table 3. Pre-trend tests across specifications

Outcome	Specification	Pre-trend p	Wald chi^2	Post-period avg coef	N treated rows	N comparison rows
Inpatient days	Full pool, full event window	0.045	9.74	−1,490	446	16,753
Inpatient days	**Propensity-trimmed, full window (PRIMARY)**	**0.482**	3.47	−1,592	446	1,463
Inpatient days	Full pool, shorter window (−3, +2)	0.010	9.12	−1,492	446	16,753
Inpatient days	Trimmed AND shorter window	0.179	3.44	−1,595	446	1,463

Outcome	Specification	Pre-trend p	Wald chi^2	Post-period avg coef	N treated rows	N comparison rows
Total margin (winsorized)	Full pool, full event window	0.089	8.07	+0.204	441	16,606
Total margin (winsorized)	**Propensity-trimmed, full window (PRIMARY)**	**0.971**	0.53	+0.137	441	1,356
Total margin (winsorized)	Full pool, shorter window (−3, +2)	0.037	6.59	+0.204	441	16,606
Total margin (winsorized)	Trimmed AND shorter window	0.924	0.16	+0.137	441	1,356

Notes: Bolded specification is my primary. Propensity-trimmed pool restricts to comparison hospitals with pscore in [0.05, 0.95], reducing the pool from 1,426 to 124 hospitals. The trimming dominates on pre-trend cleanliness for both primary outcomes without changing the headline magnitudes.

Table 4. HonestDiD relative-magnitudes sensitivity bounds (selected)

Outcome	Event time	M-bar	Lower bound	Upper bound	Robust at this M-bar?
Inpatient days	t = 0	0.5	−2,389	−719	Yes
Inpatient days	t = 0	1.0	−2,528	−611	Yes
Inpatient days	t = 0	1.5	−2,714	−472	Yes
Inpatient days	t = 0	2.0	−2,931	−286	Yes
Inpatient days	t = 1	0.5	−2,698	−776	Yes
Inpatient days	t = 1	1.0	−3,128	−484	Yes
Inpatient days	t = 1	1.5	−3,605	−69	Yes
Inpatient days	t = 1	2.0	−4,097	+392	**Breakdown**
Total margin	t = 0	0.5	+0.079	+0.378	Yes
Total margin	t = 0	1.0	+0.051	+0.406	Yes
Total margin	t = 0	1.5	+0.007	+0.448	Yes
Total margin	t = 0	2.0	−0.043	+0.495	**Breakdown**
Total margin	t = 1	0.5	−0.145	+0.250	**No** (already breaks at M-bar = 0)
Total margin	t = 1	1.0	−0.231	+0.336	No
Total margin	t = 1	1.5	−0.333	+0.444	No
Total margin	t = 1	2.0	−0.441	+0.552	No

Notes: Bounds computed via Rambachan-Roth createSensitivityResults_relativeMagnitudes on the propensity-trimmed TWFE event study. "Robust at this M-bar?" = Yes if the confidence interval excludes zero. Breakdown M-bar is the smallest M-bar at which the CI first includes zero. The inpatient-days estimate is robust at both event times out to M-bar = 1.5; the total-margin t = 0 estimate is robust out to M-bar = 1.5; the total-margin t = 1 estimate is not robust even at M-bar = 0 — flagged in the limitations.

Figures

(All figures are in `analysis/figures/`.)

- **Figure 1.** Event-study coefficients on total inpatient days, with 95% CIs. File: `event_study_inpatient_days.png` (and .pdf).
- **Figure 2.** Raw cohort-mean inpatient-days trajectory by calendar year, treated vs. comparison. File: `raw_trends_inpatient_days.png` (and .pdf).
- **Figure 3.** Event-study coefficients on total margin (winsorized), with 95% CIs. File: `event_study_total_margin.png` (and .pdf).
- **Figure 4.** Raw cohort-mean total-margin trajectory by calendar year, treated vs. comparison. File: `raw_trends_total_margin.png` (and .pdf).
- **Figure 5.** Event-study coefficients on operating margin (winsorized), secondary diagnostic. File: `event_study_operating_margin.png` (and .pdf).
- **Figure 6.** HonestDiD relative-magnitudes sensitivity bounds for inpatient days, t = 0 and t = 1. File: `honestdid_bounds_inpatient_days.png` (and .pdf).
- **Figure 7.** HonestDiD relative-magnitudes sensitivity bounds for total margin, t = 0 and t = 1. File: `honestdid_bounds_total_margin.png` (and .pdf).
- **Figure 8.** Propensity-score overlap diagnostic, treated vs. comparison. File: `iptw_propensity_overlap.png` (and .pdf).

References

(All citations resolve to entries in `literature/bibliography.bib`. The bibliography was primary-source verified after source verification; verification URLs are recorded as trailing comments on each peer-reviewed entry. No unresolved citation flags remain.)

Appendix — Trading Inpatient Beds for Financial Lifelines: REH Conversions on Margins, ED Bypass, and Outpatient Capacity

---

Appendix A. Supplementary Materials

A.1 Full HonestDiD relative-magnitudes sensitivity bounds

Table A.1 reports the complete set of HonestDiD bounds at M-bar in {0.0, 0.5, 1.0, 1.5, 2.0} for both primary outcomes (inpatient days, net-income-to-operating-revenue ratio) at event times t = 0 and t = 1. Bounds were computed via the R `HonestDiD` package's `createSensitivityResults_relativeMagnitudes` function on the propensity-trimmed TWFE event-study coefficients (Table A.3b): inpatient days t = 0 = −1,534, t = 1 = −1,650; financial proxy t = 0 = +0.230, t = 1 = +0.044. The bounds are computed via `analysis/robustness/honestdid_bounds_pivot.py`, which calls R through `rpy2`.

Table A.1. HonestDiD bounds — Inpatient days

Event time (post lag)	M-bar	Lower bound	Upper bound
t = 0	0.0	−2,281	−796
t = 0	0.5	−2,389	−719
t = 0	1.0	−2,528	−611
t = 0	1.5	−2,714	−472
t = 0	2.0	−2,931	−286
t = 1	0.0	−2,390	−915
t = 1	0.5	−2,698	−776
t = 1	1.0	−3,128	−484
t = 1	1.5	−3,605	−69
t = 1	2.0	−4,097	+392

Source: `analysis/tables/honestdid_bounds_inpatient_days.csv`. The breakdown M-bar (smallest M-bar at which the CI first includes zero) is > 2.0 at event time t = 0 and 2.0 at event time t = 1.

Table A.2. HonestDiD bounds — Total margin (winsorized)

Event time (post lag)	M-bar	Lower bound	Upper bound
t = 0	0.0	+0.098	+0.362
t = 0	0.5	+0.079	+0.378
t = 0	1.0	+0.051	+0.406

Event time (post lag)	M-bar	Lower bound	Upper bound
t = 0	1.5	+0.007	+0.448
t = 0	2.0	−0.043	+0.495
t = 1	0.0	−0.110	+0.199
t = 1	0.5	−0.145	+0.250
t = 1	1.0	−0.231	+0.336
t = 1	1.5	−0.333	+0.444
t = 1	2.0	−0.441	+0.552

Source: `analysis/tables/honestdid_bounds_total_margin.csv`. The breakdown M-bar at event time t = 0 is 2.0 (the CI first includes zero at M-bar = 2.0). At event time t = 1 the breakdown M-bar is 0.0 — i.e., the CI includes zero even with no allowed post-period bias. **The t = 1 total-margin estimate is therefore not robust to HonestDiD bounds and should not be over-interpreted as evidence of persistent financial gain.**

A.2 Full TWFE event-study coefficient tables

Table A.3a. TWFE event-study coefficients — FULL comparison pool (1,426 hospitals)

This table reports the FULL-pool TWFE event-study coefficients. The 2026-05-16 repair re-attributes these coefficients correctly: the +29.1 pp t = 0 coefficient on the financial proxy that earlier drafts described as "propensity-trimmed" is in fact the FULL-pool TWFE coefficient.

Outcome	Event time	Coefficient	SE	p-value
Inpatient days	t = −5	+796.4	294.3	0.007
Inpatient days	t = −4	+403.8	185.8	0.030
Inpatient days	t = −3	+337.2	197.7	0.088
Inpatient days	t = −2	+102.2	142.9	0.475
Inpatient days	t = −1	(omitted, base)	—	—
Inpatient days	t = 0	−1,441.2	370.5	<0.001
Inpatient days	t = 1	−1,538.4	347.3	<0.001
Net-income-to-operating-revenue ratio (winsorized)	t = −5	+0.097	0.044	0.027
Net-income-to-operating-revenue ratio (winsorized)	t = −4	+0.081	0.047	0.084
Net-income-to-operating-revenue ratio (winsorized)	t = −3	+0.042	0.068	0.530
Net-income-to-operating-revenue ratio (winsorized)	t = −2	+0.006	0.051	0.912
Net-income-to-operating-revenue ratio (winsorized)	t = −1	(omitted, base)	—	—
Net-income-to-operating-revenue ratio (winsorized)	t = 0	+0.291	0.065	<0.001

Outcome	Event time	Coefficient	SE	p-value
Net-income-to-operating-revenue ratio (winsorized)	t = 1	+0.117	0.072	0.103
Operating margin (winsorized)	t = −5	+0.136	0.038	<0.001
Operating margin (winsorized)	t = −4	+0.108	0.061	0.080
Operating margin (winsorized)	t = −3	+0.026	0.058	0.657
Operating margin (winsorized)	t = −2	−0.020	0.053	0.702
Operating margin (winsorized)	t = −1	(omitted, base)	—	—
Operating margin (winsorized)	t = 0	+0.016	0.106	0.883
Operating margin (winsorized)	t = 1	−0.418	0.160	0.009

Source: `analysis/tables/event_study_coefficients.csv`.

Table A.3b. TWFE event-study coefficients — PROPENSITY-TRIMMED comparison pool (124 hospitals)

This table reports the PROPENSITY-TRIMMED TWFE event-study coefficients. The HonestDiD bounds in Table A.1 and A.2 are computed off these (trimmed) point estimates. The t = 0 coefficient on the financial proxy is +0.230 (+23.0 pp), not +0.291; the prior manuscript inadvertently labeled the full-pool coefficient as trimmed.

Outcome	Event time	Coefficient	SE	p-value
Inpatient days	t = −5	+540.0	297.9	0.072
Inpatient days	t = −4	+309.0	203.8	0.131
Inpatient days	t = −3	+277.8	209.5	0.187
Inpatient days	t = −2	+131.3	148.5	0.378
Inpatient days	t = −1	(omitted, base)	—	—
Inpatient days	t = 0	−1,533.7	386.2	<0.001
Inpatient days	t = 1	−1,649.7	383.9	<0.001
Net-income-to-operating-revenue ratio (winsorized)	t = −5	−0.019	0.049	0.697
Net-income-to-operating-revenue ratio (winsorized)	t = −4	−0.004	0.054	0.944
Net-income-to-operating-revenue ratio (winsorized)	t = −3	−0.034	0.072	0.632
Net-income-to-operating-revenue ratio (winsorized)	t = −2	−0.012	0.053	0.817
Net-income-to-operating-revenue ratio (winsorized)	t = −1	(omitted, base)	—	—

Outcome	Event time	Coefficient	SE	p-value
Net-income-to-operating-revenue ratio (winsorized)	t = 0	+0.230	0.069	0.001
Net-income-to-operating-revenue ratio (winsorized)	t = 1	+0.044	0.079	0.580

Source: `analysis/tables/pretrend_tightening_event_studies.csv`, spec `propensity_trimmed_full_window`.

The pre-period leads on inpatient days under the unrestricted comparison pool show a downward drift from t = −5 (+796 days) to t = −2 (+102 days); under the propensity-trimmed pool, these coefficients are smaller and statistically indistinguishable from zero, which is the basis for treating the trimmed pool as the primary specification.

A.3 Pre-trend tightening across specifications

Table A.4. Pre-trend tests across four specifications

Outcome	Specification	Pre-trend p	Wald chi²	Post-period avg coef	N treated rows	N comparison rows	N obs
Inpatient days	v2_baseline_full_window	0.045	9.74	−1,490	446	16,753	17,170
Inpatient days	propensity_trimmed_full_window	0.482	3.47	−1,592	446	1,463	1,908
Inpatient days	shorter_window_v1_0-3_+2	0.010	9.12	−1,492	446	16,753	17,170
Inpatient days	trimmed_AND_shorter_window	0.179	3.44	−1,595	446	1,463	1,908
Total margin (winsorized)	v2_baseline_full_window	0.089	8.07	+0.204	441	16,606	17,019
Total margin (winsorized)	propensity_trimmed_full_window	0.971	0.53	+0.137	441	1,356	1,797
Total margin (winsorized)	shorter_window_v1_0-3_+2	0.037	6.59	+0.204	441	16,606	17,019
Total margin (winsorized)	trimmed_AND_shorter_window	0.924	0.16	+0.137	441	1,356	1,797

Source: `analysis/tables/pretrend_tightening_comparison.csv`. The propensity-trimmed full-window specification (bolded) is my primary spec and dominates on pre-trend cleanliness for both primary outcomes without changing the headline magnitudes.

A.4 IPTW balance table

Table A.5. Pre-conversion baseline balance (treated last pre-conversion year vs. comparison 2022)

Covariate	Mean treated (pre)	Mean comparison (2022)	SMD	N treated	N comparison
Operating margin (winsorized)	−0.356	−0.105	−0.784	37	1,332
Total margin (winsorized)	−0.143	+0.041	−0.814	37	1,332
Total inpatient days	1,421	2,656	−0.787	38	1,343
Total operating revenue (USD)	$11.32M	$34.59M	−1.070	37	1,333
Uncompensated-care ratio (winsorized)	0.0262	0.0193	+0.194	37	1,249
Bad-debt ratio (winsorized)	0.0691	0.0498	+0.305	38	1,316

Source: `analysis/tables/iptw_balance.csv`. SMDs exceeding |0.10| indicate substantial imbalance; SMDs above |0.78| on operating margin, total margin, total inpatient days, and total operating revenue motivate the propensity-trimming step.

A.5 Operating-margin secondary results (with full caveat)

The operating-margin estimate is reported here in the appendix (rather than the main text) because the post-conversion HCRIS treatment of the REH facility-payment supplement reclassifies it outside operating revenue under modified Worksheet G-3, which mechanically depresses the operating-margin numerator. The operating-margin estimate is therefore an HCRIS measurement artifact, not a causal effect of REH conversion on the converting facility's true financial position. I document the artifact in Section 7 of the main text.

Table A.6. Operating margin — Callaway-Sant'Anna ATT (do not interpret causally)

Statistic	Value
CS ATT (simple avg)	−0.2551 (−25.51 pp)
Standard error	0.1305
p-value	0.051
TWFE post-period avg coef	−0.2014
Pre-trend joint p	0.0031
N treated CCNs	37
N treated rows	441
N comparison rows	16,608

Source: `analysis/tables/main_results.csv`. The pre-trend failure (p = 0.003) is itself a symptom of the reclassification artifact — the post-period measurement regime is fundamentally different from the pre-period measurement regime, so the standard parallel-trends assumption is violated by construction. I do NOT interpret the negative point estimate as evidence of REH financial harm. A clean operating-margin estimate would require a custom HCRIS Worksheet G-3 reclassification analysis that I flag as the most useful follow-up technical study.

A.6 Total-revenue secondary results

Table A.7. Total operating revenue (USD) — Callaway-Sant'Anna ATT (do not interpret causally)

Statistic	Value
CS ATT (simple avg)	−$7.87M
Standard error	$1.11M
p-value	<0.001
Pre-trend joint p	n/a (insufficient pre-period coverage)
N treated CCNs	37
N treated rows	442
N comparison rows	16,617

Source: `analysis/tables/main_results.csv`. As with operating margin, the negative coefficient is mechanical — the REH facility-payment supplement is reclassified outside operating revenue, so the operating-revenue measure mechanically declines post-conversion. The total-revenue measure (which includes the supplement) yields the positive total-margin estimate in the main text.

A.7 Pre-specified Phase 6 robustness analyses

The following robustness analyses are pre-specified for Phase 6 (post-Phase-5-writing) but have not been executed in the current analysis. I document them here for transparency:

1. **Alternative propensity-trim cutoffs**: re-estimate the headline ATT under pscore in [0.10, 0.90] and pscore in [0.02, 0.98] cutoffs.

2. **Drop-individual-cohort placebos**: re-estimate dropping each treated cohort (2023, 2024) one at a time.
3. **Drop-individual-state placebos**: re-estimate dropping each of the seven heaviest-converting states (MS, AR, TX, OK, AL, KS, GA) one at a time. Drop-Mississippi is the most consequential because Mississippi contributes 6 of 38 treated facilities.
4. **k-NN matching alternative**: replace the propensity-trimming step with k-nearest-neighbor matching (k = 5, Mahalanobis distance on the same baseline covariates).
5. **HCRIS Worksheet G-3 reclassification**: reconstruct a properly defined post-conversion operating margin by adding the REH facility-payment supplement back into the operating-revenue subtotal.

If executed, the results of these analyses will be reported in the journal-submission draft (Phase 6) and in the post-Phase-5 internal-review iteration.

A.8 OLD<->NEW CCN crosswalk audit (2026-05-16 repair)

37 of 38 OLD-CCN termination dates fall within +/-90 days of the REH conversion effective date. One match — Crittenden Community Hospital (`new_ccn=180779`, `old_ccn=180095`, conversion 2023-03-13, old-CCN termination 2024-03-12) — has a 365-day termination-date gap. The match is retained as high-confidence on address-exact, ZIP-exact, name-overlap=1.00, and inactive-old-CCN evidence; the gap is most plausibly explained by a delayed POS-record termination at CMS rather than two distinct hospitals at one address. The earlier appendix claim that "no candidate was borderline" was incorrect and is retracted; see `data/clean/reh_ccn_crosswalk_audit_notes.md` for the full audit memo. The canonical crosswalk file is `data/clean/reh_ccn_crosswalk_old_to_new.csv` (NOT `reh_hgi_hcris_with_conversion_dates.csv`, which earlier drafts incorrectly cited).

Three partial-date REHs (DeWitt Hospital 2024-05-15 month-only; Progressive Health of Houston 2024-08-15 month-only; Clifton-Fine Hospital 2025-01-28 inferred-from-announcement) and one unresolved current REH (J Paul Jones Hospital, CCN 010781) are flagged in the audit notes. A drop-partial-date sensitivity analysis is queued as next-revision work.

A.9 Replication

All headline estimates in this paper can be reproduced by running, from the paper's directory, using Anaconda Python 3 (`/opt/anaconda3/bin/python3`), since several dependencies (`pandas`, `numpy`, `pyfixest`, `differences`, `rpy2`) are not in the default system Python:

```
bash data/scripts/run_all.sh          # builds the OLD<->NEW crosswalk + the extended HCRIS p
bash analysis/run_all.sh              # main analysis: full-pool CS, trimmed-CS primary, TWFE
```

The Phase 4 repair primary spec memo is `analysis/log/phase4_repair_2026-05-16_trimmed_cs.md`; the specification registry is `analysis/specification-registry.yml`. The Rambachan-Roth HonestDiD bounds are computed by `analysis/robustness/honestdid_bounds_pivot.py` (a Python script that calls R `HonestDiD` through `rpy2`); previous appendix text incorrectly cited an `.R` extension. The HCRIS panel construction, OLD<->NEW CCN crosswalk, and conversion-date roster are documented in `data/clean/` and the cleaning scripts in `data/scripts/`.

Do NOT invoke `data/scripts/phase3_pipeline.py` — that file is intentionally deprecated as of the 2026-05-16 repair because its previous implementation called the repo-level generic batch runner, which would overwrite this paper's bespoke analysis code.