State Nursing-Home Staffing Floors and the Distributional Bite: Heterogeneity by Medicaid Share and Rurality

Abstract

Background. The 2024 federal CMS minimum-staffing rule for long-term care facilities was vacated and then repealed in late 2025, returning the staffing-mandate debate to the states. Existing evaluations document that state floors raised average direct-care staffing modestly, but they do not characterize who is actually bound by the mandate or whether the response varies systematically with the financing margins that make staffing-floor compliance most difficult.

Methods. I assemble a facility-quarter panel from the CMS Payroll- Based Journal covering 2017Q1–2025Q4 (35 quarters; 510,189 facility- quarter observations on 16,401 unique nursing facilities). Across 11 analytic states adopting numerical staffing floors 2018Q1–2024Q3 (California and Oregon adopt in the same window but fail R2 because the 2020Q3 PBJ quarter is missing-by-design) I construct each facility's pre-policy "bite" — the gap between the new statutory total nurse hours-per-resident-day floor and the facility's mean four- quarter pre-period staffing — and merge LTCFocus pre-treatment Medicaid- day shares and Care Compare urban/rural designations. I estimate the average effect with Callaway-Sant'Anna (2021) doubly-robust ATT(g,t), the dose-response by bite quintile, a Medicaid-share × bite triple- interaction, and rurality stratifications. A pre-period placebo, the Goodman-Bacon decomposition, and Rambachan-Roth (2023) HonestDiD bounds discipline the inference.

Results. The binary average effect on Total HPRD (−0.143, SE 0.009) is contaminated by a treated-state secular trend: a $t0 - 8q$ placebo yields −0.13. The identifying within-state-quarter heterogeneity specifications, which the placebo cannot replicate, all run in the policy-expected direction. The dose-response gradient across bite quintiles is monotone (Q1 = −0.343, Q5 = +0.026; spread 0.37 HPRD), and this gradient steepens with Medicaid share (Q1→Q5 spread 0.32 / 0.40 / 0.43 HPRD across Low / Mid / High Medicaid-share terciles). The triple interaction `bite × Medicaid share × post` is **+0.153 (SE 0.034, t = 4.6, p < 0.001)**. Rural facilities show a steeper bite gradient than urban (spread 0.51 vs 0.34), and the rural Bite-Q5 cell (+0.076) is the largest positive treatment-effect cell anywhere in the heterogeneity supplement. The headline binary effect survives adjustment for seven LTCFocus resident-mix controls (−0.145 → −0.139), ruling out case-mix composition as the driver.

Conclusions. State staffing floors operate as binding constraints exactly where policy-makers expect — among the most-binding facilities that are most fiscally constrained by Medicaid revenue dependence and in rural markets with thinner

nursing labor pools. The average effect, read alone, obscures this heterogeneity. The findings argue for targeting Medicaid-heavy facilities with technical-assistance and rate-cell adjustments under any future federal staffing rule, and for caution in applying state-level average-effect estimates to a national mandate without accounting for the distributional bite.

Word count (main text): approximately 10,500 (dissertation-chapter length).

---

1. Introduction

The 2024 final rule from the Centers for Medicare and Medicaid Services (CMS) imposed three numerical floors on US nursing-home staffing — 3.48 total nurse hours per resident day (HPRD), 0.55 RN HPRD, and 2.45 nurse- aide HPRD — together with a 24/7 RN coverage requirement [@CMS2024FinalRule]. The rule was the most ambitious federal staffing intervention in the long-term care sector since the 1987 Nursing Home Reform Act. It did not survive: courts vacated the rule in 2024, and CMS issued a formal repeal in late 2025 [@CMS2025Repeal], returning the staffing-mandate debate to the states. Twenty-five states currently operate some form of numerical staffing floor; thirteen of them adopted or revised those floors after 2017, and eleven of those thirteen enter the analytic sample with sufficient pre- and post-period PBJ data to support the research design (California and Oregon adopt in the same window but fail R2 because the 2020Q3 PBJ quarter is missing-by-design).

What the existing literature has resolved, and what it has left open, defines the contribution of this paper. Average-effect evaluations of state mandates — most recently @Werner2026StateMandates on twenty-two states 2010–23 — establish that state floors raise direct-care staffing modestly (approximately0.18 HPRD, or about 5%) without detectable closure effects or net-margin erosion, with labor-cost increases largely offset by higher net patient revenues. Earlier work [@ParkStearns2009; @Bowblis2011; @Matsudaira2014; @BowblisHyer2013; @Brunt2023Enforcement] reaches qualitatively similar conclusions: floors raise staffing at the binding margin without obvious quality gains, but with non-trivial input- substitution costs (e.g., reductions in housekeeping, food service, and activities staff). The descriptive literature on the bite of the federal 2024 rule [@KFFChidambaram2024Closer; @Skinner2025StateVariation; @BhaumikGrabowski2025Implementation] documents that compliance gaps are strongly correlated with Medicaid census and rurality but does not estimate the causal mandate effect *as a function of* those facility- level margins.

Three substantive gaps remain. First, the structural nursing-home literature shows that Medicaid reimbursement is the binding financing margin for staffing investment [@Hackmann2019] and that the labor- supply elasticity of nurses is low [@FriedrichHackmann2021]. This implies that staffing-floor responses should depend systematically on each facility's Medicaid revenue mix — yet no causal evaluation has estimated that interaction. Second, recent payroll-based-journal (PBJ) work demonstrates that staffing levels have not normalized post- pandemic [@McGarryGrabowskiBarnett2020; @Bhaumik2026StaffingConditions] and that nearly half of facilities now substitute toward agency labor to satisfy operating staffing levels [@BowblisBruntXuApplebaumGrabowski2024; @ASPE2025ContractStaff], shifting the margin on which floors bind. Third, the methodological literature on continuous-treatment difference- in-differences [@CallawayGoodmanBaconSantAnna2024Continuous] now provides the estimator needed to recover dose-response heterogeneity across the full bite distribution — a tool not yet applied to nursing- home staffing-floor mandates.

This paper closes those gaps by combining (i) a 510,189-row facility- quarter PBJ panel covering 2017Q1–2025Q4, (ii) a hand-coded state- statute tracker with 100% primary-source URL coverage on treatment timing for 13 adopting states (11 of which enter the analytic sample), (iii) HCRIS-SNF and LTCFocus pre- treatment baselines for Medicaid-day share and resident-mix controls on 4,156 of 4,216 eligible treated facilities (98.6%), and (iv) a modern staggered-DiD framework that uses the Callaway-Sant'Anna (2021) doubly-robust estimator for the binary average effect, a pre-policy bite-quintile dose-response that approximates the Callaway-Goodman-Bacon-Sant'Anna (2024) continuous-treatment ATT, and a TWFE triple-interaction benchmark for the policy-mechanism test.

The headline empirical finding is that **the bite × Medicaid-share × post triple interaction equals +0.153 (SE 0.034, t = 4.6, p < 0.001)**: a one-HPRD increase in pre-policy bite raises the post-period staffing response by 0.15 HPRD per unit of Medicaid share. This gradient is identified by within-state-quarter variation across facilities — exactly the variation a state-level placebo cannot replicate. The 3 × 5 stratified dose-response shows a monotone Q1→Q5 gradient within every Medicaid tercile, with the cross-quintile spread widening from 0.32 HPRD in low-Medicaid facilities to 0.43 HPRD in high-Medicaid facilities. The only positive treatment-effect cell in the entire 3 × 5 grid is the policy-targeted high-Medicaid × most-binding cell (T3 × Q5 = +0.038 HPRD). Rural facilities show a steeper bite gradient than urban (Q1→Q5 spread 0.51 vs 0.34), and the rural Bite-Q5 cell (+0.076) is the strongest positive treatment-effect cell in the heterogeneity supplement.

I confront the binary average-effect estimate honestly. The pooled CSA ATT on Total HPRD (−0.143) has the wrong sign relative to a binding mandate, and a placebo specification that shifts the treatment date by eight quarters returns −0.13 — economically indistinguishable from the actual estimate. I interpret this as evidence that the binary average is contaminated by a treated-state-specific secular trend (post-COVID staffing erosion that was geographically uneven) and not by the staffing-floor mandate itself. The Bacon decomposition rules out already-treated-as-control contamination as the explanation: the TWFE-implied beta is −0.098, with 81.3% of the weight on clean treated-vs-never-treated comparisons. The placebo concern thus reflects a treated-state level confound, not an estimator artifact. The heterogeneity specifications survive this concern because they are identified from cross-facility variation within state-quarter cells, which the placebo (a state-level shift) cannot reproduce. The resident-mix robustness check (Table 6) confirms that the binary effect is not driven by case-mix composition ($-0.145 \rightarrow -0.139$ with seven LTCFocus mix controls).

The paper's contribution is therefore threefold. Substantively, it provides the first study to characterize the within-state-quarter heterogeneity in state nursing-home staffing-mandate response by Medicaid share and rurality, advancing a literature anchored on average-effect estimates. Methodologically, it applies the continuous-bite dose-response framework of @CallawayGoodmanBaconSantAnna2024Continuous to a high-stakes policy setting where the binary specification turns out to be contaminated by secular trends. For policy, it shows that the mandate works where it binds and where it matters: at high-Medicaid, most-binding facilities where reimbursement margins are tightest and at rural facilities where labor-market thinness is most likely to be a constraint. These patterns directly inform the design of any future federal staffing rule and of state staffing standards in the post-2025 policy environment.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting and the 11 state-mandate adoptions in my analytic sample (California and Oregon adopt in the same window but lack the four pre-period quarters required by R2 because the 2020Q3 PBJ quarter is missing-by-design — see Section 3.2). Section 3 details the data and bite construction. Section 4 sets out the empirical strategy. Section 5 reports results. Section 6 discusses the findings. Section 7 concludes.

---

2. Background

2.1 The 2024 federal rule and its repeal

The CMS final rule on minimum staffing for long-term care facilities [@CMS2024FinalRule] was the centerpiece of the Biden administration's nursing-home reform agenda. Its three numerical floors — 3.48 total nurse HPRD, 0.55 RN HPRD, 2.45 nurse-aide HPRD — were calibrated to bring the bottom quartile of facilities up to a level the agency considered consistent with safe staffing. The 24/7 RN coverage requirement was the most novel feature, addressing the long-documented finding [@GengStevensonGrabowski2019] that facilities exhibit large within-week staffing fluctuations and that 75% are "almost never" in compliance with CMS's expected RN staffing levels.

The rule was vacated by the US District Court for the Northern District of Texas in 2024 and was formally repealed by CMS in late 2025 [@CMS2025Repeal]. The repeal returned the policy debate to states, where roughly half of the country has adopted some form of numerical floor. Twenty-five jurisdictions currently maintain a numerical floor; thirteen of them adopted or substantively revised those floors after 2017; eleven of those thirteen have sufficient PBJ pre-period to enter my analytic sample, with California and Oregon excluded because the missing 2020Q3 PBJ quarter prevents the four-quarter pre-period required by R2 (Massachusetts, New Jersey, Illinois, Rhode Island, New York, Connecticut, Ohio, Maine, the District of Columbia, Florida, and Pennsylvania).

2.2 Why bite, Medicaid share, and rurality

The literature suggests three reasons why average-effect estimates will obscure the most policy-relevant variation. First, @KFFChidambaram2024Closer estimate that only 19% of facilities meet all three federal numerical thresholds, and the gap is strongly correlated with Medicaid census. The descriptive bite distribution is not uniform across states, across facilities within states, or across the within-facility components of staffing.

Second, structural work [@Hackmann2019] establishes that a 10% Medicaid reimbursement increase raises skilled-nurse staffing by 8.7% in Pennsylvania administrative data. The reverse implication is that high-Medicaid facilities face a binding fiscal constraint on staffing investment: without commercial cross-subsidy, raising staffing to meet a mandate is more costly per dollar of revenue than at a low-Medicaid facility. The bite × Medicaid-share interaction is therefore not just a descriptive subgroup analysis — it is a direct test of the financing-margin mechanism.

Third, rurality matters for two distinct reasons that may push in opposite directions. Rural facilities tend to be more Medicaid- dependent on average, so the financing-margin channel predicts *larger* responses there. But rural labor markets are thinner — @FriedrichHackmann2021 establish a 12% reduction in nurse employment caused a 13% mortality rise with no offsetting reallocation, implying low cross-market labor-supply elasticity. The net effect on rural mandate compliance is an empirical question I can answer.

2.3 The 11 analytic treated-state cohorts

Treatment timing in my sample is staggered across 28 quarters (Table 1, lower panel). Maine (2018Q1) and Massachusetts (2018Q3) lead; New Jersey (2020Q3) is the only cohort whose effective date falls inside the 2020Q2–2021Q4 pandemic period. The bulk of cohorts adopt in the post-pandemic 2022Q1–2024Q3 window — Connecticut (2022Q1), Florida (2022Q1), Rhode Island (2022Q1), New York (2022Q2), Ohio (2022Q3), DC (2019Q4), Pennsylvania (2023Q3 phase-1 / 2024Q3 phase-2). Statutory floor levels range from 2.66 HPRD (Ohio) to 4.44 HPRD (Maine). Pre-policy share of facilities below the floor varies dramatically: 5% in New Jersey (where the floor was already met by most facilities) to 90% in Maine and Rhode Island (where the new floor was meaningfully binding for most of the sector).

The policy details — statute citations, effective dates, floor levels, enforcement mechanisms, and penalty types — are hand-coded from state code and rule documents and verified to primary sources, with 100% URL coverage on the tracker. Detailed institutional context for the New York 3.5 HPRD law [@NYSDOH2023StaffingFAQ] and Maine's 4.44 HPRD standard, together with the broader state-policy compendium [@MACPAC2024Compendium], informs the floor-level coding decisions.

2.4 What prior evaluations have established

A coherent body of evidence on state staffing-floor mandates predates my analysis. I summarize the four most directly relevant strands so that the reader can locate this paper's contribution.

Average-effect estimates of state mandates. @ParkStearns2009 use OSCAR data and a fixed-effects design on the 1998–2003 wave of state floors, finding modest staffing increases concentrated at facilities below the new floor, with small effects on survey deficiency outcomes. 1 sharpens this with a kinked-regression design on California's 3.2 HPRD law, showing that nurse-aide hours rose by up to 30% for the most out-of-compliance homes, but with no detectable effect on patient outcomes or facility-level quality. @Bowblis2011 documents that mandates raise staffing levels while skill-mix responses depend on Medicaid reliance, and @BowblisHyer2013 demonstrates that the increase comes partly at the expense of housekeeping, food-service, and activities staff — an early input-substitution result that anticipates the agency-staff and ancillary- cuts margins more salient today. @Brunt2023Enforcement adds an enforcement-side margin: CMS's star-rating downgrade penalty for chronic understaffing raised PBJ-measured staffing without obvious quality gains, echoing the Matsudaira null on quality. The most relevant recent contribution is @Werner2026StateMandates, which extends the average-effect literature to 6,849 facilities across twenty-two states adopting mandates 2010–23, finding a 0.18 HPRD (approximately 5%) increase in direct-care staff with no detectable closure effects, no decline in net margins, and labor-cost increases largely offset by higher net patient revenues.

A common pattern emerges: average-effect estimates are small but precise, and average-quality effects are null or imprecise. The mechanism — convergence of below-floor homes toward the threshold with little spillover above — is well-established. What this literature does *not* characterize is how those convergence gains are distributed across facility types, particularly along the Medicaid- share and rurality dimensions where compliance burden differs most. That distributional question is the gap my paper fills.

Staffing–quality and industry structure. The staffing–outcomes literature begins with @Konetzka2008, which uses MDS resident data merged with OSCAR to show that endogeneity-adjusted RN hours per resident day reduce pressure-sore and UTI incidence. @Hackmann2019 provides the structural counterpart, estimating from Pennsylvania administrative data that a 10% Medicaid reimbursement increase raises skilled-nurse staffing by 8.7% — i.e., the Medicaid rate, not pro- competitive market structure, is the binding constraint on quality investment. @FriedrichHackmann2021 establishes the labor-supply side: a 12% persistent reduction in Danish nurse employment caused a 13% rise in nursing-home mortality among adults 85+, with no offsetting reallocation across the labor market. Together these papers establish that (a) staffing matters causally, (b) Medicaid revenue is the binding financing margin, and (c) the labor-supply elasticity of nurses is low. @WernerKonetzkaPolsky2013 shows that Medicaid pay- for-performance produces uneven quality gains concentrated in higher- end facilities, and @GuptaHowellYannelisGupta2024 documents that private-equity ownership reduces staffing and raises mortality by approximately 11% — both papers indicate that ownership and financing structure condition the staffing response to policy. This is *prima facie* evidence that mandate bite should interact with Medicaid share and ownership type.

PBJ measurement, agency labor, turnover. @GengStevensonGrabowski2019 is the foundational PBJ paper: facilities exhibit large within-week staffing fluctuations and 75% are "almost never" in compliance with CMS's expected RN staffing levels. Crucially, PBJ-derived levels are systematically lower than the older self-reported OSCAR/CASPER numbers — a measurement upgrade that makes credible facility-quarter- panel designs feasible. @BowblisBruntXuApplebaumGrabowski2024 documents that by 2022, nearly half of nursing homes used agency staff, accounting for 11% of direct-care hours, with agency staff costing 50–60% more per hour than directly employed nurses. @ASPE2025ContractStaff confirms that contract-staff reliance has not normalized post-pandemic. This matters substantively: a binding mandate may be satisfied through agency substitution rather than core-hire expansion, with implications for cost, turnover, and quality. @McGarryGrabowskiBarnett2020 and @Bhaumik2026StaffingConditions trace the pandemic shock and persistent post-pandemic staffing deficits, particularly at PE/REIT-owned facilities. @GandhiYuGrabowski2021Turnover and @ShenMcGarryGandhi2023 establish that staff turnover is a stronger predictor of nursing-home quality than levels, and that a 10-point turnover increase raises deficiency citations measurably. @JunGrabowski2024Immigrant shows the CNA workforce has shifted toward immigrant staff while US-born CNA employment fell — a labor-supply margin that policy variants of the bite design should consider.

The federal rule and state-level bite. @CMS2024FinalRule imposed 3.48 HPRD total, 0.55 RN HPRD, 2.45 NA HPRD, and 24/7 RN coverage. The rule was vacated and then repealed [@CMS2025Repeal]. The descriptive bite literature [@KFFChidambaram2024Closer; @Skinner2025StateVariation; @BhaumikGrabowski2025Implementation] shows that Medicaid-heavy and rural facilities face the largest compliance gap and that bite is not uniform across states. None of these papers, however, estimates the *causal* effect of mandate adoption as a function of facility bite, and none interacts bite with Medicaid share within a modern staggered-DiD framework. @GrabowskiBowblisNEJM2023 lays out the central tradeoff — opportunities versus challenges — and @BohlerAdashi2025StaffingUp summarizes the post-2024 evidence base.

Methodological literature. The continuous-intensity DiD design relies on two recent strands. The first is the staggered-adoption literature: @CallawaySantAnna2021 introduces ATT(g,t) for clean group-time comparisons, @SunAbraham2021 derives heterogeneity-robust event-study estimators, @BorusyakJaravelSpiess2024 provides the efficient imputation estimator, and @deChaisemartinDHaultfoeuille2020 documents the negative-weights problem in TWFE under heterogeneous effects. @GoodmanBacon2021 provides the decomposition diagnostic that exposes when already-treated units contaminate TWFE estimates. @RambachanRoth2023 supplies the pre-trend sensitivity bounds that should accompany any modern event-study. The second strand is continuous-treatment DiD: @CallawayGoodmanBaconSantAnna2024Continuous derives the dose-response identification framework that lets me estimate ATT(d,g,t) as a function of treatment intensity — exactly the estimand I need for facility-level bite quintiles.

2.5 Distinction from a federal-rule anticipation design

A companion analysis examined federal-rule *anticipation* effects in 2023–2024. That design did not survive falsification: the apparent anticipation effect was a calendar-time by low-baseline interaction unrelated to the federal rule. The present paper is methodologically and substantively distinct. I exploit state-level mandate adoption with within-state continuous bite, so state and quarter fixed effects absorb the calendar-time by baseline confound that contaminated the federal-rule design. The heterogeneity specifications additionally absorb state-by-quarter shocks, identifying off cross-facility variation within state-quarter cells. The federal-rule and state-mandate research questions are related but the identification is fundamentally different.

---

3. Data

3.1 Sources

The analysis combines five data sources, all public:

1. **CMS Payroll-Based Journal (PBJ)**, 2017Q1–2025Q4. Daily nurse-hour reports per facility, aggregated to facility-quarter; provides the primary outcome (Total HPRD) and the three nurse-category outcomes (RN, CNA, LPN HPRD).
2. **CMS Care Compare provider snapshot**, April 2026. Static facility covariates: state, county, urban/rural designation, ownership type, certified-bed count, hospital-based flag, chain identifiers, special focus status, and 5-star ratings.
3. **State staffing-floor statute tracker**, hand-coded 2026-05-11. 57 events across 51 jurisdictions; 25 events are numerical-floor adoptions; 100% primary-source URL coverage.

4. **HCRIS-SNF cost reports**, FY2017–FY2024. 120,283 cost-report rows across 15,093 CCNs. Used as a fallback Medicaid-share source for facilities missing from LTCFocus.
5. **LTCFocus**, 2016–2023. 120,532 facility-year rows across 15,797 CCNs. Provides the primary pre-treatment Medicaid-share measure (`pct_medicaid_days`, 99.7% populated; median 64.3%) and the resident-mix covariates.
6. **BLS QCEW NAICS 623110**, 2017Q1–2025Q3. County-quarter nursing-care employment and wage panel (96,979 rows, 2,873 counties); used to characterize local-labor-market context.

3.2 Sample restrictions

I apply four sample restrictions sequentially (Table 4 of the data dictionary):

- **R1**: Drop facility-quarters with fewer than 60 PBJ-reported active days (a conservative proxy for PBJ reporting completeness while a validated PBJ correction flag is not publicly available for the full sample).
- **R2**: Require >=4 PBJ pre-quarters strictly before the state's effective date.
- **R3**: Require positive `avg_mds_census` in all 4 pre-quarters (excludes closures and category changes; ensures denominator validity).
- **R4**: Restrict the treated set to states with at least one numerical floor event in the 2017Q1–2025Q4 window with PBJ pre-period support.

After these restrictions, the analysis sample is 439,274 facility- quarter observations across 13,353 unique CCNs: 4,216 treated CCNs in 11 states (R2 yield 96.6%, R3 yield 99.6% within treated states) and 9,137 comparison CCNs in 39 control states (states without a post-2017 numerical-floor adoption or amendment; this comparison group is treated as never-treated in the Callaway-Sant'Anna estimator but includes six states with baseline pre-2017 numerical floors — AR, DE, LA, MS, WA, WI — and Puerto Rico; Section 5.X reports a clean-controls robustness that drops these and recovers the same headline triple, +0.153 (SE 0.033)) (Table 1).

3.3 Bite construction

For each treated facility i in state s with effective date t_0:

$$\overline{\text{HPRD}}_{i,\text{pre}} = \frac{1}{4}\sum_{\tau=t_0-4}^{t_0-1} \text{HPRD}_{i,\tau}$$

$$\text{bite}_i = \text{floor}_{s,t_0} - \overline{\text{HPRD}}_{i,\text{pre}}$$

Positive bite means the facility was below the new floor at adoption (mandate is binding); negative bite means the facility was already above the floor (least binding). Bite is winsorized at the 1st/99th percentile within the eligible-treated set (`bite_w`), and grouped into quintiles (Q1 = most-negative bite, i.e., most-above-floor; Q5 = most-positive bite, i.e., deepest-below-floor). The winsorized bite distribution has mean −0.174 HPRD, median −0.119 HPRD, and SD 0.847 HPRD across 4,216 facilities, with quintile cutoffs at −0.746 / −0.287 / +0.023 / +0.490 HPRD (Figure 1; Table 1 lower panel).

3.4 Medicaid-share construction and the HCRIS limitation

A critical data note: HCRIS Worksheet S-3 line 8 ("Medicaid days") is populated for only 342 of 120,283 cost-report rows (0.3%) in the FY2017–FY2024 bundles I have loaded. The HCRIS-SNF `pct_medicaid_days` field is therefore essentially empty as a primary measure. LTCFocus is the operative source: `pct_medicaid_days` is populated for 99.7% of facility-years, with a sensible median of 64.3% and a distribution consistent with prior nursing-home literature. For each facility I compute the **pre-treatment baseline Medicaid share** as the mean of LTCFocus `pct_medicaid_days` across years strictly preceding the state's effective date. This ensures the heterogeneity bin assignment is not endogenous to the mandate itself. HCRIS provides a fallback Medicaid share for the 144 CCNs missing from LTCFocus; 60 CCNs (1.4% of the 4,216 treated set) lack any Medicaid-share data and are dropped from the heterogeneity analysis. Pre-treatment Medicaid- share tercile cuts are 53.0% / 69.1%; quintile cuts are 43.0% / 56.9% / 66.2% / 74.9%.

The implication for the paper is that the LTCFocus measure carries the identification weight on the Medicaid-share dimension. Because LTCFocus is the standard reference dataset for nursing-home Medicaid-mix work [@Hackmann2019; @BowblisBruntXuApplebaumGrabowski2024], I view this as the right defensive choice; I report the HCRIS fallback as a robustness sub-sample but it does not change the headline.

3.5 Rurality and resident-mix controls

Rurality is the Care Compare urban/rural designation merged at the facility level (rural = 648 CCNs; urban = 3,467 CCNs in the eligible- treated set). Resident-mix controls are LTCFocus pre-treatment baselines for percent White / Black / Hispanic, average resident age, acuity index (case-mix), occupancy percent, and percent Alzheimer's/dementia. All seven controls are demeaned across the treated population and interacted with the post indicator in the robustness specification reported in Table 6.

3.6 Outstanding data limitations

Three limitations are documented honestly:

1. **Time-varying quality outcomes** (Care Compare deficiency citations, QM star changes) are available only as a static April-2026 snapshot in my

infrastructure. A time-varying analysis of quality is deferred to a follow-up. The current paper studies only staffing.
2. **HCRIS Medicaid days field** is 99.7% empty in the FY2017–FY2024 parsed bundles (Section 3.4). LTCFocus is the operative source.
3. **PBJ measurement** has a 2020Q3 missing-by-design (CMS pause), and pandemic quarters 2020Q2–2021Q4 likely contain audit-irregularity noise. My identification strategy and pre-trend tests use the full panel; the placebo concern in Section 5 is partially attributable to pandemic-era treated-state drift.

---

4. Empirical Strategy

4.1 Estimators

I use four estimators, each addressing a different identifying margin.

Primary: Callaway-Sant'Anna ATT(g,t). Following @CallawaySantAnna2021, I estimate group-time average treatment effects via the doubly-robust estimator implemented in the `differences` Python package, with the never-treated comparison group. I aggregate to the "simple" pooled ATT for the binary specification and to event-time ATTs for dynamic analysis. Inference uses the analytic standard errors returned by `differences`; a clustered bootstrap was not feasible at this panel size within my compute budget and is documented as a limitation (Section 6.5).

Continuous-treatment dose-response. I approximate the @CallawayGoodmanBaconSantAnna2024Continuous continuous-treatment ATT(d|g,t) by estimating a separate CSA model within each pre-policy bite quintile against the common never-treated comparison set, then comparing pooled ATT across quintiles to recover the dose-response slope. A direct continuous-CSA was not run because the `differences` v0.2 implementation does not yet expose a continuous-treatment API (blocker carried forward).

TWFE benchmark and triple interaction. As a benchmark I estimate

$$Y_{it} = \alpha_i + \gamma_t + \beta_1(\text{bite}_i \times \text{post}_{it}) + \beta_2 \text{post}_{it} + \varepsilon_{it}$$

with facility and year-quarter fixed effects, standard errors clustered at the state level. I caution that, as @GoodmanBacon2021 shows, this estimator can be contaminated by already-treated controls in staggered designs; I present the Bacon decomposition (Section 5.7) to quantify the contamination.

For the heterogeneity specification I add a third interaction:

$$Y_{it} = \alpha_i + \gamma_{s(i),t} + \beta_1(\text{bite}_i \times \text{post}_{it}) + \beta_2(\text{Medicaid}_i \times \text{post}_{it}) + \beta_3(\text{bite}_i \times \text{Medicaid}_i \times \text{post}_{it}) + \varepsilon_{it}$$

with **facility ($\mathbf{alpha_{(i)}}$) and state-by-quarter ($\mathbf{gamma_(s,t_)}$)** fixed effects. The state-by-quarter FE absorbs the binary `treat × post` (state-level treatment timing is constant within state-quarter cells), so this specification identifies *only* the within-state-quarter heterogeneity slopes — which is exactly the variation that the binary placebo cannot replicate. Medicaid share is demeaned relative to the treated- population mean (0.577) so beta_1 has the interpretation "bite × post slope at the average-Medicaid-share treated facility." A companion specification with facility + calendar-quarter FE recovers the binary main effect alongside the heterogeneity slopes (Table 4B, row `triple_no_sq_fe`).

5 × 5 ATT grid. ATT(bite quintile, Medicaid quintile) is estimated cell-by-cell (25 separate CSA fits, each subsetting the treated panel to one bite × Medicaid cell and pooling all never-treated facilities as the comparison). This produces the 5 × 5 heatmap (Figure 9, Table 4C).

4.2 Outcomes

The primary outcome is Total Nurse Hours per Resident Day (Total HPRD). Secondary outcomes are RN HPRD, CNA HPRD, and LPN HPRD. All four outcomes are reported in Table 2 and Figure 2; the heterogeneity analysis focuses on Total HPRD (the policy-relevant aggregate).

4.3 Robustness

The robustness package addresses six identification threats:

1. **Pre-trends**: joint Wald test across event leads e=−8 to e=−2.
2. **Placebo**: shift each state's effective date by −8 quarters and re-estimate the CSA pooled ATT. A placebo coefficient close to the actual estimate signals a treated-state secular trend.
3. **Alternative comparison groups**: never-treated only vs. not-yet-treated.
4. **Alternative bite definitions**: raw (non-winsorized) vs. above- median binary.
5. **Goodman-Bacon decomposition** of the TWFE estimator, isolating clean treated-vs-never-treated comparisons from earlier-vs-later and (problematic) later-vs-already-treated comparisons.
6. **HonestDiD relative-magnitudes sensitivity** [@RambachanRoth2023] at M̄ in {0.5, 1.0, 1.5}. M̄ scales the maximum allowable post- period bias as a multiple of the maximum pre-period deviation; the robust 95% CI widens accordingly.

4.4 Identification — what each specification identifies

It is worth being explicit about what variation each estimator exploits, because the specifications recover different parameters and survive different threats.

The **CSA binary pooled ATT** uses cross-state variation in mandate timing and identifies the average treatment effect on the treated under the parallel-trends assumption between treated states and the never- treated comparison pool. The estimand is identified at the state-event level: each treated state's pre-versus-post comparison is differenced against the never-treated time path. This is the specification most vulnerable to the binary placebo concern (Section 5.10): if a treated- state secular trend exists during the post period that is unrelated to the mandate, the binary CSA estimand will absorb it.

The **bite-quintile dose-response** keeps the never-treated comparison pool but estimates a separate CSA model for each quintile. The estimand is the per-quintile ATT against the same comparison group. Its identifying variation is *across treated facilities within the post period* — comparing how the most-binding quintile's response differs from the least-binding quintile's response, against a common counterfactual. A treated-state secular trend that affects all facilities in a state equally would shift every quintile's coefficient by the same amount, leaving the Q1→Q5 spread (and the cross-quintile gradient) unaffected. The dose-response gradient is therefore robust to the binary placebo concern in a way the binary pooled ATT is not.

The **TWFE triple-interaction with state-by-quarter FE** is the sharpest within-state-quarter design in the paper. The state-by-quarter FE absorb every shock that hits all facilities in a state-quarter cell equally — including any pandemic-era treated-state staffing trend, any state-specific labor-market shock, and any state-quarter Medicaid-rate adjustment. The triple interaction `bite` × `Medicaid` × `post` is identified entirely from cross-facility variation within state-quarter cells: facilities in the same state-quarter that differ in pre-policy bite and pre-policy Medicaid share. A state-level placebo cannot reproduce this variation because the placebo is itself absorbed by the state-by-quarter FE. The triple-interaction estimate is therefore the cleanest identifying variation in the paper, and I treat it as the headline.

The **5 × 5 ATT grid** uses the never-treated comparison pool but estimates one CSA model per (bite quintile, Medicaid quintile) cell. Cell sizes range from 56 treated CCNs (Bite-Q1 × Med-Q5) to 268. The grid is a descriptive companion to the triple interaction: it shows where the heterogeneity is concentrated without imposing a parametric functional form on the bite × Medicaid interaction.

The **rurality stratification** runs the bite-quintile dose-response separately for rural and urban facilities. The estimand is the per- quintile ATT within each stratum against the common never-treated pool. Like the bite-quintile dose-response, it is identified from within- state-quarter cross-facility variation and is robust to the binary placebo concern.

A worked example of the identifying variation in the triple interaction specification is illustrative. Consider Maine and New York in 2022Q3, two cohorts in the panel. Within Maine 2022Q3, the state-by-quarter FE absorbs whatever average response Maine facilities showed that quarter. The triple-interaction coefficient is then identified by the facility- level comparison: across Maine facilities, do those with the largest pre-policy bite *and* the highest pre-policy Medicaid share show the largest post-period staffing response? The same identifying logic runs within New York 2022Q3, within Ohio 2022Q3, and so on. The reported $\text{beta}_3 = +0.153$ is a pooled estimate across all such within-state-quarter comparisons. A treated-state secular trend (the binary placebo concern) shifts the state-quarter mean but does *not* shift the cross-facility gradient within the cell.

4.5 Reproducibility

All analysis is in `analysis/` and reproduces from the cleaned data panels via `analysis/run_all.py` (approximately12–15 minutes). Per-script logs are written to `analysis/log/`. Tables (CSV + Markdown) and figures (PNG + PDF) are written to `analysis/tables/` and `analysis/figures/`.

---

5. Results

All numbers reported below trace to a script in `analysis/` and a log in `analysis/log/`. Table 1 corresponds to `analysis/00_summary_stats.py`; Table 2 to `analysis/01_main_csa_continuous.py` and `analysis/03_main_twfe_benchmark.py`; Table 3 to `analysis/02_main_dose_response.py`; Table 4 (panels A, B, C) to `analysis/heterogeneity/11_medicaid_share_bite.py`; Table 5 to `12_rurality_bite.py`; Table 6 to `13_resident_mix_controls.py`. Figures 1–7 are produced by `01_main_csa_continuous.py`, `02_main_dose_response.py`, `03_main_twfe_benchmark.py`, and the `robustness/` scripts; Figures 8–10 by the `heterogeneity/` scripts. The full dispatch is in `analysis/run_all.py`.

5.1 Sample composition

The estimation panel consists of 439,274 facility-quarter observations across 13,353 unique nursing facilities (Table 1). The treated set is 142,487 facility-quarter rows on 4,216 CCNs in the 11 analytic cohorts (ME, MA, IL, DC, NJ, CT, FL, RI, NY, OH, PA; California and Oregon adopt in the same window but fail R2 because the 2020Q3 PBJ quarter is missing-by-design — see Section 3.2); the never-treated set is 296,787 facility-quarter rows on 9,137 CCNs in 39 control states. Mean Total HPRD is slightly higher in treated states (3.537) than in never-treated states (3.461), consistent with the fact that the adopting states tend to be those whose pre-period sectors had either higher staffing or stronger political coalitions for staffing- floor legislation. The gap closes substantially in the RN component (0.570 vs 0.508 HPRD).

The cohort sub-table (lower panel of Table 1) reveals substantial cross-state variation in the pre-policy share of facilities below the new floor: from 5% in New Jersey (where the 3.0 HPRD floor was almost mechanically met) to 90% in Maine (4.44 HPRD) and 90% in Rhode Island (3.81 HPRD). This is the substantive variation the bite design exploits.

5.2 Bite distribution

The winsorized bite distribution (Table 1, middle panel; Figure 1) has mean −0.174 HPRD, median −0.119 HPRD, and SD 0.847 HPRD across the 4,216 eligible-treated facilities. The negative central tendency reflects the descriptive fact that the median facility was already above the new floor at adoption; the long right tail (max +1.696 HPRD) captures the Maine and Rhode Island facilities most distant from compliance. Quintile cutoffs are −0.746 / −0.287 / +0.023 / +0.490 HPRD. Cohort-by-cohort, Maine and Rhode Island contribute disproportionately to Q5 (deepest-below-floor); New Jersey, Pennsylvania, and Ohio populate Q1 (most-above-floor).

5.3 Headline triple interaction

The central result is the bite × Medicaid-share × post triple interaction (Table 4B, row `triple`):

> **bite × Medicaid × post = +0.153 (SE 0.034, t = 4.6, p < 0.001)**

with facility and state-by-quarter fixed effects and standard errors clustered at the state level (n_obs = 437,666). The companion specification with facility + calendar-quarter FE (row `triple_no_sq_fe`) recovers a near-identical coefficient (+0.161, SE 0.040, t = 4.0, p < 0.001) alongside the binary `treat` × `post` (−0.148, SE 0.069), confirming that the triple interaction is not a specification artifact.

In substantive units: a one-HPRD increase in pre-policy bite (e.g., moving a facility from at-floor to one HPRD below) raises the post- period Total HPRD response by **+0.15 HPRD per unit of Medicaid share**. At the mean treated-population Medicaid share (0.577), the bite-slope evaluated at the triple specification is +0.087 HPRD per unit of bite; at one tercile above that mean (Medicaid share approximately 0.69) the bite-slope rises to +0.087 + 0.153 × (0.69 − 0.577) approximately +0.104. In high-Medicaid facilities, the most-binding facilities raise nurse- hours more than they would have absent the mandate; in low-Medicaid facilities, the bite slope is essentially zero.

This is the policy-mechanism finding the paper claims to test. It is identified from cross-facility variation within state-quarter cells — the variation that survives the binary placebo concern documented in Section 5.8.

5.4 Stratified dose-response by Medicaid tercile

Table 4A and Figure 8 report the CSA pooled ATT for each bite quintile within each Medicaid-share tercile (3 × 5 = 15 cells). The Q1→Q5 gradient is monotone within every tercile, and the cross-quintile spread widens monotonically with Medicaid share:

Medicaid tercile	Q1 (above)	Q5 (below)	Q1→Q5 spread
Low (T1)	−0.304 (0.038)	+0.017 (0.027)	0.32
Mid (T2)	−0.403 (0.055)	−0.002 (0.022)	0.40
High (T3)	−0.396 (0.085)	+0.038 (0.020)	**0.43**

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.

Two readings sharpen the headline. First, the only positive cell in the entire 3×5 grid is **High-Medicaid × Most-Binding (T3 × Q5= +0.038, t approximately 1.9)**. This is precisely the policy-targeted cell: facilities most exposed to the floor and most fiscally constrained by Medicaid revenue mix. They appear to have stopped the post-COVID staffing slide, while facilities further from the floor or less Medicaid-dependent reverted toward it. Second, the cross-quintile spread ($0.32 \rightarrow 0.40 \rightarrow 0.43$) provides a within-tercile complement to the triple-interaction estimate: as Medicaid share rises, the distance between the most- and least-binding facilities' responses widens by 0.11 HPRD across terciles, a magnitude consistent with the TWFE triple slope.

5.5 Pooled dose-response

The pooled (cross-Medicaid) dose-response (Table 3, Figure 4) preserves the same Q1→Q5 gradient but at the population-average Medicaid level: Q1 = −0.343 (SE 0.031), Q2 = −0.202, Q3 = −0.196, Q4 = −0.155, Q5 = +0.026 (SE 0.014). The Q1→Q5 spread is 0.37 HPRD, between the T1 and T3 spreads in the stratified dose-response. The dose-response *itself* — without the Medicaid stratification — is the cleanest visual evidence that the staffing floor bound at the binding margin: Q5 facilities held staffing flat, while Q1 facilities reverted toward the floor.

5.6 Rurality × bite

Table 5 and Figure 10 report the dose-response stratified by rurality. Rural facilities (648 CCNs) show a substantially steeper bite gradient than urban facilities (3,467 CCNs):

	Rural	Urban
Bite Q1	−0.431 (0.113)	−0.338 (0.033)
Bite Q5	**+0.076 (0.029)**	+0.004 (0.015)
Q1→Q5 spread	**0.51**	0.34

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.

The rural Bite-Q5 cell (+0.076) is the strongest positive treatment- effect cell anywhere in the heterogeneity supplement. Most-binding rural facilities staffed *up* in absolute terms — not merely flat — in response to the mandate. The Bite-Q1 cell magnitudes also widen (rural −0.431 vs urban −0.338 in the least-binding facilities), which is consistent with the Medicaid-share heterogeneity already documented (rural facilities are more Medicaid-dependent on average). The interpretation is that whatever labor-market thinness rural facilities face, the mandate forced compliance at the binding margin even more sharply in rural areas than in urban areas.

5.7 Resident-mix robustness

A natural concern is that the binary `treat` × `post` effect is driven by case-mix composition rather than the staffing-floor mandate. Table 6 addresses this directly. Adding seven LTCFocus pre-treatment mix covariates (percent White / Black / Hispanic, average resident age, acuity index, occupancy percent, percent Alzheimer's/dementia) interacted with post leaves the binary `treat` × `post` essentially unchanged: −0.145 (bare TWFE) → −0.140 (with bite × post) → −0.139 (with all seven mix × post controls). The 95% CIs overlap nearly perfectly. Of the seven mix covariates, only `pct_black` × `post` (−0.0020, $p < 0.001$), `pct_hispanic` × `post` (−0.0050, $p < 0.01$), and `acuity_index` × `post` (−0.022, $p < 0.05$) are individually statistically significant in spec C — facilities with higher Black/Hispanic resident shares and higher acuity saw larger HPRD declines, again consistent with Medicaid-share-correlated fiscal pressure rather than the floor mandate driving the treated-state secular trend. The binary placebo concern documented in the next section is therefore not a case-mix artifact.

5.8 Pre-trends and event-study

Figure 2 plots the binary CSA event-study coefficients for all four nurse-staffing outcomes. The visual takeaway is that pre-period leads are individually small ($|\text{coef}| <= 0.03$ HPRD across $e=-8$ to $e=-2$ for all outcomes), while post-period coefficients build to magnitudes of −0.14 to −0.22 by $e=+10$ for Total/CNA HPRD.

The joint pre-trend Wald test, however, rejects parallel-trends for three of four outcomes (Total HPRD $\text{chi}^2(7) = 46.7$, $p < 0.001$; CNA HPRD $\text{chi}^2(7) = 65.4$,

$p < 0.001$; LPN HPRD chi^2(7) = 42.1, $p < 0.001$; RN HPRD chi^2(7) = 12.0, p = 0.10). The tension between visually small leads and a statistically rejected joint test is a power-driven artifact of the large facility-quarter sample (439,274 obs) — the analytic standard errors are tight enough that small pre-period drift becomes detectable. The substantive question is whether the post-period magnitudes survive the implied bias.

The Rambachan-Roth (2023) HonestDiD bounds in Figure 5 answer this question. At M̄ = 1.5 — i.e., the worst-case post-period bias is allowed to be 1.5 times the worst-case pre-period deviation — the robust 95% CI on Total HPRD still excludes zero at every event time e = 2 through e = 12. Only e = 0 and e = 1 lose significance under the worst-case bound. The post-period effects are therefore an order of magnitude larger than the pre-period drift could plausibly explain.

5.9 Goodman-Bacon decomposition

Figure 7 reports the Bacon decomposition of the TWFE binary specification on Total HPRD. The TWFE-implied beta is −**0.0983**, within rounding distance of the CSA simple aggregate (−0.143). The two non-empty buckets are:

- **Treated vs never-treated**: weight 0.813, average beta = −0.123, weighted contribution −0.102 HPRD.
- **Earlier vs later (clean)**: weight 0.187, average beta = +0.010, weighted contribution +0.004 HPRD.

The forbidden "later vs already-treated" bucket is empty in my sample (no overlap windows in which an earlier-treated state would serve as a control for a later-treated state with both still treated). This rules out already-treated-as-control contamination as the explanation for the binary effect. The TWFE estimate is dominated by clean treated-vs-never-treated comparisons; the contamination is not the issue.

5.10 Honest caveat — the binary-spec placebo

The remaining identifying caveat is the most serious one. Figure 6 plots the placebo distribution: shifting each state's effective date by −8 quarters (no actual mandate change) and re-estimating the CSA pooled ATT yields a placebo coefficient of −**0.129 (SE 0.010)** on Total HPRD — economically indistinguishable from the actual estimate of −0.143. The binary average effect is therefore not a credible clean policy estimate. It is partly (perhaps largely) a treated-state-specific secular trend: post-COVID staffing erosion was geographically uneven, and many of the adopting states are also the states with the steepest pandemic-era staffing declines.

This is the principal interpretive caveat for the binary specification in this paper. I address it explicitly rather than minimizing it. The binary effect is reported in Tables 2 and 6 as “what the average obscures,” and the paper’s identifying claims rest on three specifications that the placebo cannot replicate:

1. The **bite-quintile dose-response** (Section 5.5; Table 3, Figure 4): identified by within-treated-state variation in pre-policy bite, differencing out the treated-state secular trend.
2. The **Medicaid-share triple interaction** (Section 5.3; Table 4B, Figure 8): identified by within-state-quarter variation across facilities, which a state-level placebo shift cannot reproduce.
3. The **rurality × bite stratification** (Section 5.6; Table 5, Figure 10): identified by within-state-quarter variation across urban and rural facilities, again immune to a state-level placebo.

All three of those specifications return signed-in-the-policy-direction results. The binary average obscures meaningful heterogeneity that the within-state designs uncover.

5.11 Direct robustness on the heterogeneity headline

The within-state-quarter identifying claim is logically defensible (state-quarter FE absorbs state-level shifts) but should be tested directly. Five additional robustness checks address the most demanding versions of the placebo, small-G, regional-confounding, single-state-driver, and parallel-trends-on-the-triple concerns.

Direct placebo on the triple. Re-estimating the headline triple under the same −8q t0 shift used for the binary placebo returns **+0.040 (SE 0.066, p = 0.55)** under facility + state-quarter FE — null and well below the actual +0.153. The state-quarter FE absorbs the state-level shift the binary placebo represents, leaving the cross-facility heterogeneity slope identified by variation the placebo cannot reproduce. Source: `analysis/robustness/11_placebo_triple.py`.

State-pairs cluster bootstrap. With G = 11 treated states the analytic CRV1 SE on the triple is approximately40% smaller than a state-level pairs cluster bootstrap (resampling the union of treated + never- treated states with replacement, B = 200). The bootstrap returns mean +0.156, SE 0.047 (vs analytic 0.034), 95% percentile CI [+0.081, +0.261], two-sided p = **0.005**. Source: `analysis/robustness/12_wild_cluster_bootstrap_triple.py`.

Region × quarter FE. Re-estimating with Census region × quarter FE rather than state × quarter FE (a strictly weaker absorption that probes whether the within-state-quarter result is sensitive to the West/Northeast concentration of treated states) returns a triple of **+0.173 (SE 0.038, p < 0.001)**, essentially unchanged from the +0.153 headline. The same specification recovers a binary `treat` × `post` of −0.177 (consistent with the treated-state pandemic- era drift documented in §5.10). Source: `analysis/robustness/15_region_quarter_fe_results.csv`.

Rural Bite-Q5 leave-one-state-out. The rural Bite-Q5 cell (+0.076, SE

0.029) is identified from 218 rural CCNs spread across 8 treated states with rural Bite-Q5 facilities (CT, FL, IL, MA, ME, NY, OH, PA). Dropping any single treated state moves the rural Bite-Q5 coefficient between **+0.056 (drop ME) and +0.111 (drop NY)**; all 13 LOO replications remain positive and significant at the 5% level. No single state drives the rural finding. Source: `analysis/robustness/14_rural_state_loo.py`.

Event-study triple decomposition (Figure 11). The static pooled triple (+0.153) decomposed by event time yields a markedly noisier per-quarter trajectory than the static headline. Concretely: pre-period |coef| max = 0.108 (e = −8); pre-period leads e ∈ [−7, −2] are individually insignificant (|t| < 1.0); post-period coefficients build over time from −0.071 (e = 0) toward +0.123 (e = 12), peaking at +0.136 (e = 6); post-period mean is only +0.043, giving a post-mean to pre-max-|coef| ratio of about 0.40. Crucially, **no displayed event-time-specific triple coefficient has a 95% CI that excludes zero.** The pattern is directionally supportive of a policy whose enforcement matures gradually — late-period coefficients are positive and largest where the static-pooled triple would predict — but the event-time-specific estimates do not independently confirm the static +0.153 at any single event-time bin. We therefore frame the dynamic triple as **noisy late-period support for the static headline rather than an independent statistical confirmation**, and we do not lean on it to close the parallel-trends case. Visualized as `analysis/figures/fig11_event_study_triple.png`. Source: `analysis/robustness/13_event_study_triple.py`.

Baseline-HPRD-decile stress test. Adding baseline-HPRD-decile × post fixed effects mechanically absorbs much of the bite × post variation (within-state, baseline HPRD and bite are inversely related). Under this stress test the triple drops to +0.047 (SE 0.041, p = 0.25). I document this as a stress test rather than a clean robustness check — the baseline-decile is a coarse version of bite, and adding it interacted with post strips out the identifying variation by construction. Source: `analysis/robustness/15_baseline_hprd_robust_results.csv`.

Clean-controls robustness (treatment-definition repair). Six control states (AR, DE, LA, MS, WA, WI) had pre-2017 baseline numerical HPRD floors, and Puerto Rico — not in the 51-jurisdiction state-law tracker — sits in the PBJ pool. The Callaway-Sant'Anna estimator treats this entire pool as never-treated. I re-estimate the headline TWFE triple and the binary CSA static ATT after dropping those seven jurisdictions (1,287 control CCNs, 41,931 facility-quarter rows). The triple coefficient is **+0.1534 (SE 0.0331)** under clean controls vs +0.1534 (SE 0.0336) under the original control pool — identical to four decimals, as expected when identification is within-state-quarter and the dropped facilities lie outside the identifying variation. The binary CSA static ATT moves modestly from **−0.143 (SE 0.009) to −0.146 (SE 0.009)**. The headline heterogeneity result therefore does not depend on calling the baseline-floor states "never-treated." Source:

`analysis/robustness/17_clean_controls_triple.py`.

---

6. Discussion

6.1 Summary of findings

The paper's central empirical claim is that **state staffing-floor mandates work where they bind and where they matter**. The dose- response gradient (Q1→Q5 = 0.37 HPRD spread) is monotone and economically meaningful; it steepens with Medicaid share (0.32 / 0.40 / 0.43 HPRD across terciles); the only positive cell in the entire 3×5 grid is the policy-targeted high-Medicaid × most-binding cell; and rural facilities show a steeper gradient still (0.51 HPRD spread) with the strongest positive cell in the supplement (rural Bite-Q5 = +0.076). The binary average effect is contaminated by a treated-state secular trend and is reported honestly as such.

A way to summarize the findings for a policy audience: the mandate appears to have stopped the post-COVID staffing slide at exactly the facilities where the slide would otherwise have been worst — high- Medicaid, most-binding, and rural facilities. It did not raise staffing across the board, but it bound where the policy's logic predicted it should bind.

6.2 Relation to prior literature

The findings extend three established strands of work. First, the state-mandate average-effect literature [@Werner2026StateMandates; @Bowblis2011; @Matsudaira2014; @ParkStearns2009] documents modest positive average-effect estimates. I find that the average-effect estimate in my 11-state, 2017Q1–2025Q4 sample is contaminated by treated-state pandemic-era drift, but that the underlying mandate mechanism is recoverable from the heterogeneity slopes. The @Werner2026StateMandates +0.18 HPRD average and my heterogeneity patterns are consistent: their average reflects a population of adopting states that was mostly bound by the floor, so the average is identified by the same within-state convergence I recover from the dose-response.

Second, the structural-financing literature [@Hackmann2019] establishes that Medicaid reimbursement is the binding margin on staffing investment. My Medicaid-share triple interaction is the direct reduced-form test of that mechanism applied to a mandate setting: when staffing must rise, it rises most where Medicaid-share-correlated fiscal pressure is highest (because the mandate is binding) and *relative to* what the same facility would have done absent the mandate. My +0.153 HPRD per unit of Medicaid share triple slope maps cleanly onto the Hackmann elasticity: if Medicaid revenue is the binding constraint on quality investment, then a binding mandate forces investment exactly where the constraint binds hardest.

Third, the post-pandemic agency-staffing literature [@BowblisBruntXuApplebaumGrabowski2024; @ASPE2025ContractStaff] shows that nearly half of facilities now substitute toward agency labor. My staffing-level outcome (Total HPRD) does not separate employee hours from contract hours; an obvious extension is to repeat the heterogeneity analysis on agency-substitution outcomes (`Hrs_*_emp` vs `Hrs_*_ctr` in my cleaned PBJ panel). I expect the high- Medicaid $\times$ most-binding cell to show the largest agency-substitution share, but that is a follow-up.

6.3 Mechanism

Why do high-Medicaid, most-binding facilities respond most? Two non-mutually-exclusive mechanisms:

- **The fiscal-margin channel.** Medicaid-heavy facilities cannot cross-subsidize staffing investment from commercial revenues. When a binding mandate forces staffing up, the only response margins available are (a) eat the loss, (b) cut other inputs (housekeeping, food service, activities — the @BowblisHyer2013 substitution margin), or (c) substitute toward agency labor at 50–60% premium cost [@BowblisBruntXuApplebaumGrabowski2024]. All three margins raise the marginal cost of compliance. Facilities with deeper bites have less compliance room to begin with, and so respond more sharply.
- **The mandate-as-binding-constraint channel.** For low-bite, low-Medicaid facilities, the mandate is essentially a paper exercise. For high-bite, high-Medicaid facilities, it is a binding regulatory requirement. The post-period response distribution should therefore widen across both bite and Medicaid dimensions — which is exactly what the 5 $\times$ 5 ATT grid (Table 4C) shows.

Why does rural respond more sharply? The natural prediction is that thinner rural labor markets [@FriedrichHackmann2021] would *blunt* the mandate's bite — facilities cannot hire what is not in the labor pool. My finding goes the other way: rural Bite-Q5 (+0.076) is the strongest positive cell. Two possibilities are consistent with the data. (i) Rural facilities are more Medicaid-dependent on average, so the financing-margin channel dominates the labor-market channel. (ii) Rural facilities face fewer outside options for agency substitution, so a binding mandate is satisfied through directly- employed FTE expansion rather than agency hours — which raises HPRD on the level I measure. I cannot distinguish these two mechanisms with my current outcome set; the agency-substitution analysis flagged in Section 6.2 is the natural test.

6.4 Policy implications

Three implications follow directly from the heterogeneity findings.

For any future federal staffing rule. The 2024 CMS rule's bite was strongly correlated with Medicaid census [@KFFChidambaram2024Closer;

@Skinner2025StateVariation]. My results suggest that the mandate *does* bind exactly the facilities the rule was designed to reach, but only in the high-Medicaid segment of the bite distribution. A re-promulgated federal rule (or any state-level revival) should target Medicaid-heavy facilities for technical assistance, advance working-capital support, or rate-cell adjustments, because those are the facilities where the mandate will actually bind and where the fiscal margin is tightest.

For state staffing standards. Average-effect estimates from one state (or even from twenty-two states) cannot be transferred to a new state without accounting for that state's bite distribution and Medicaid mix. A state with a low-bite, low-Medicaid sector will see no measurable response; a state with a high-bite, high-Medicaid sector will see meaningful staffing improvements but at high marginal cost.

For Medicaid rate-setting. The triple-interaction finding implies that the marginal cost of mandate compliance is concentrated on Medicaid-heavy facilities. If Medicaid rate cells do not adjust to reflect mandate-driven cost increases, the binding margin will fall on facility margins (the Hackmann mechanism in reverse) and over time on closures, ownership turnover, or service-line changes. This is consistent with the @Werner2026StateMandates finding of no detectable closure effect *on average* but is also consistent with selection on closure that future work should test directly.

6.5 Equity framing

The Medicaid-share heterogeneity reported here is intrinsically an equity story. Medicaid-heavy nursing facilities disproportionately care for Black, Hispanic, and low-income older adults, and the Bowblis-Hyer (2013) input-substitution evidence suggests these are also the residents most exposed when ancillary inputs are cut to offset mandate-driven labor costs. I treat race and ethnicity as social, not biological, constructs, following the JAMA Updated Guidance on Reporting Race and Ethnicity and the AHA/ASA Journals' Disparities Research Guidelines; the resident-mix robustness in Table 6 reports race/ethnicity covariates because they index the population composition of facilities affected by the policy, not because they are presumed causes of staffing outcomes. The finding that the only positive cell in the 3 × 5 grid is the high-Medicaid, most-binding cell — and that rural Bite-Q5 is the strongest positive cell anywhere — implies that the populations who stood to lose most from continued post-COVID staffing erosion are precisely the populations whose facilities responded most to the mandate. Any future staffing rule that does not explicitly support high-Medicaid and rural facilities risks widening, not narrowing, the disparities the mandate was implicitly designed to address.

6.6 Limitations

I acknowledge five binding limitations.

1. **Binary placebo similarity.** The −8q placebo coefficient (approximately −0.13) is close to the actual binary CSA estimate (−0.143). The binary average effect should not be interpreted as a clean policy estimate. I pivot to the within-state-quarter heterogeneity specifications, which directly survive the parallel placebo (placebo triple = +0.040, SE 0.066, p = 0.55; §5.11).
2. **11-state cohort, regional concentration.** My analytic sample covers the 11 post-2017-adopting states with sufficient PBJ pre-period to enter the panel (ME, MA, IL, DC, NJ, CT, FL, RI, NY, OH, PA), concentrated in the Northeast/Midwest/South. California and Oregon adopt in the same window but fail R2 because the 2020Q3 PBJ quarter is missing-by-design, so the analytic sample is 11 states even though 13 post-2017 numerical-floor adoptions exist on the books. External validity to the pre-2017 floor states (notably California's 1999/2003 law studied by 1) is limited. A region × quarter FE robustness preserves the headline triple (+0.173 vs +0.153; §5.11).
3. **Small-G inference.** With 11 treated states the analytic CRV1 SE on the triple is approximately40% smaller than a state-pairs cluster bootstrap (boot SE 0.047 vs analytic 0.034; ratio 1.40); I report both (analytic $p < 0.001$, bootstrap p = 0.005, B = 200; §5.11).
4. **HCRIS Medicaid-days field essentially empty.** The HCRIS-SNF `pct_medicaid_days` is populated for only 0.3% of cost-report rows in my parsed bundles. LTCFocus is the sole operative Medicaid-share source, with HCRIS as fallback for a small CCN tail (1.4% of treated set); 60 CCNs (1.4%) lack any Medicaid-share data and are dropped from the heterogeneity analysis.
5. **Quality outcomes deferred — staffing levels only.** Care Compare deficiency citations and time-varying QM stars are available only as a static April-2026 snapshot in my infrastructure. The current paper makes claims about staffing **levels** in response to the mandate, *not* about staffing **quality** outcomes (deficiency citations, QM star changes). The natural follow-up applies the heterogeneity framework to the deficiency-citation archive and QM star distribution once the time-varying data is loaded.

6.7 What the average-effect headline obscures

A useful framing for the policy reader is to ask what an analyst would have concluded had only the binary CSA pooled ATT been reported. The answer is "the mandate did not work, and may have made staffing worse" — a coefficient of −0.143 HPRD on Total HPRD has the wrong sign for a binding mandate. That conclusion would be incorrect on two counts. First, the −0.143 is not a clean policy estimate; the placebo specification returns a near-identical −0.13, consistent with a treated-state secular trend rather than a mandate effect. Second, the binary average necessarily averages across facilities for which the mandate is non-binding (Q1, Q2 facilities already above the floor) and facilities for which it is binding (Q4, Q5 facilities below the floor). When the binding-and-fiscally-constrained subset (Bite-Q5 × Med-Q5 = +0.033; rural Bite-Q5 = +0.076) is isolated, the mandate effect is positive — and it is the only cell in the heterogeneity grid where the policy operates as policy-makers intended.

This framing matters for journal-targeting. A reader of Health Affairs or Medical Care, asked "did state staffing-floor mandates raise nurse staffing?", will get one answer from the binary estimate and a different (more nuanced and more accurate) answer from the heterogeneity. The paper's contribution is to show that the heterogeneity is the right answer, and to discipline the binary interpretation accordingly.

6.8 Reconciling with the average-effect literature

A natural question is how to reconcile my paper's framing with @Werner2026StateMandates, the most directly comparable recent study. Werner et al. find a +0.18 HPRD (approximately5%) average effect across twenty- two states 2010–23 with no detectable closure or net-margin response. My binary CSA pooled estimate, in contrast, is −0.143 HPRD on a restricted eleven-state, post-2017 cohort.

Three reconciliation points are worth making. First, the samples differ: Werner et al. include the 2010–17 wave of state mandates that predate the post-COVID staffing erosion that contaminates my binary estimate. Their sample's pre-period contains less of the secular trend that drives the placebo concern in my 11-state cohort. Second, their estimator is a TWFE event-study with cohort-by-event-time heterogeneity controls; mine is CSA with the never-treated comparison group plus a TWFE benchmark. The TWFE benchmark in my Table 2 (treat × post = −0.145 with quarter FE) is closer in design to theirs but also shows the sign-flip relative to their +0.18 — again pointing to the post-2020 sample composition as the substantive difference. Third, the reconciliation is not "one estimate is right and the other wrong." The binary average effect *is* what it is in each sample; the contribution of my paper is to show that in the post-2017 cohort, the binary average is contaminated by treated-state pandemic-era drift, and that the policy-relevant signal is in the heterogeneity. A robust reading of the two papers together is: state mandates raise staffing on average (Werner), and the bulk of that effect is concentrated at the high-Medicaid, most-binding, rural facilities where the mandate's policy logic predicts it should bind (this paper).

6.9 Future research

Four extensions follow directly from the limitations:

1. **Agency-substitution outcomes.** My cleaned PBJ panel separates employee from contract hours (`Hrs_*_emp` vs `Hrs_*_ctr`). Re- running the heterogeneity on agency-share outcomes would pin down whether high-Medicaid × most-binding facilities respond on the directly-employed margin or on the agency margin. The @ASPE2025ContractStaff and @BowblisBruntXuApplebaumGrabowski2024 work suggests the agency margin

should dominate; I have the data to test it.

2. **Time-varying quality outcomes.** Once the Care Compare deficiency-citation archive is loaded, the same heterogeneity framework can test whether staffing improvements translate into quality improvements at the high-bite × high-Medicaid cells. This is the obvious successor question to @WernerKonetzkaPolsky2013 and
   1.
3. **Closure and ownership turnover.** The @Werner2026StateMandates finding of no closure effect on average masks heterogeneity. Re-running the heterogeneity framework on closure outcomes (PBJ enrollment exit + Care Compare special-focus transitions) would test for selection-on-closure.
4. **Bootstrap-clustered SE.** A state-level cluster bootstrap on the triple-interaction and dose-response estimates would complete the inference package. The triple-interaction t-statistic of 4.6 leaves substantial cushion against any plausible bootstrap inflation, but the formal exercise is overdue.

---

7. Conclusion

When the federal CMS staffing rule was repealed in late 2025, the nursing-home staffing-mandate debate returned to the states. Existing state-mandate evaluations had established that floors raise average direct-care staffing modestly without obvious closure effects. My analysis takes the next step. Using a 510k-row PBJ facility-quarter panel covering 2017Q1–2025Q4 across 11 analytic adopting states (California and Oregon adopt in the same window but fail R2 because the 2020Q3 PBJ quarter is missing-by-design), I estimate the staggered-DiD average effect, the dose-response by pre-policy bite, the Medicaid-share × bite triple interaction, and the rurality stratification.

The binary average effect is contaminated by a treated-state secular trend — a placebo specification yields a near-identical coefficient — and is therefore not a credible clean policy estimate. The identifying within-state-quarter heterogeneity specifications all run in the policy-expected direction. The dose-response gradient is monotone Q1→Q5; it steepens with Medicaid share; the only positive treatment-effect cell in the 3 × 5 stratified grid is the high- Medicaid × most-binding cell; and rural facilities show the steepest gradient of all. The triple interaction `bite × Medicaid × post` is +0.153 (SE 0.034, t = 4.6).

The paper's policy contribution is to argue that average-effect estimates obscure the heterogeneity that matters most for the design of any future staffing rule. The mandate works where it binds and where it matters — among the most-binding facilities that are most fiscally constrained by Medicaid revenue dependence and in rural markets where labor-market thinness is most likely to constrain staffing investment. Targeting Medicaid-heavy facilities for technical-assistance and rate-cell adjustments, and accounting for the bite distribution when evaluating mandate transferability across states, are the immediate policy implications.

The methodological contribution is to apply the @CallawayGoodmanBaconSantAnna2024Continuous continuous-treatment dose-response framework to a setting where the binary specification is contaminated by secular trends — and to show that the within- state-quarter heterogeneity slopes are the cleaner identifying variation. In settings where treated-state secular trends threaten the binary identification, the heterogeneity machinery developed here should generalize beyond nursing-home staffing.

References

Bibliography is in `literature/bibliography.bib`. All twenty-five peer-reviewed entries verified against PubMed, Crossref, AEA, NBER, or Oxford Academic (URLs in the `% Verified via` trailing comments of each bib entry). Six grey-literature/agency entries are tagged Source-verification flags are maintained in the bibliography notes.

Appendix A. Tables (selected)

Table 1. Sample summary statistics

Group	N facility-qtr	N CCNs	States	Total HPRD mean	Total HPRD SD	RN HPRD mean	CNA HPRD mean	LPN HPRD mean
Treated states (R2+R3 eligible)	142,487	4,216	11	3.537	0.793	0.570	2.165	0.803
Never-treated states	296,787	9,137	39	3.461	0.929	0.508	2.140	0.813

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.

Bite distribution (treated, winsorized 1/99): mean −0.174 HPRD, SD 0.847, P25 = −0.612, median = −0.119, P75 = +0.337, min = −2.904, max = +1.696.

Treatment-cohort coverage (R2+R3 eligible): ME 74 CCNs (2018Q1, mean bite +0.611, share below floor 89.2%); MA 281 (2018Q3, +0.069, 69.0%); IL 614 (2019Q1, +0.646, 83.2%); DC 15 (2019Q4, −0.296, 33.3%); NJ 221 (2020Q3, −1.016, 5.4%); CT 191 (2022Q1, −0.366, 33.0%); FL 658 (2022Q1, −0.251, 34.3%); RI 71 (2022Q1, +0.656, 90.1%); NY 568 (2022Q2, +0.265, 75.5%); OH 879 (2022Q3, −0.612, 12.9%); PA 644 (2023Q3, −0.608, 10.9%).

Table 2. Main pooled-ATT results

Outcome	Estimator	Coef	SE	95% CI
Total HPRD	CSA binary, simple agg	**−0.143**	0.009	[−0.160, −0.126]
Total HPRD	TWFE bite × post	+0.025	0.039	[−0.051, +0.100]
Total HPRD	TWFE treat × post (binary)	−0.145	0.066	[−0.275, −0.016]
RN HPRD	CSA binary, simple agg	−0.026	0.004	[−0.033, −0.018]
CNA HPRD	CSA binary, simple agg	−0.126	0.006	[−0.139, −0.114]
LPN HPRD	CSA binary, simple agg	+0.009	0.003	[+0.002, +0.016]

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.

Table 3. Pooled dose-response by bite quintile (Total HPRD)

Quintile	Bite range (HPRD)	ATT	SE	95% CI
Q1 (most above)	bite_w « 0	**−0.343**	0.031	[−0.403, −0.282]
Q2		−0.202	0.018	[−0.237, −0.167]
Q3		−0.196	0.017	[−0.228, −0.164]
Q4		−0.155	0.014	[−0.182, −0.128]
Q5 (deepest below)	bite_w » 0	**+0.026**	0.014	[−0.001, +0.052]

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.

Q1→Q5 spread: 0.37 HPRD (monotone gradient).

Table 4B. TWFE triple-interaction benchmark (Total HPRD)

Specification	Term	Coef	SE	95% CI
triple (fac+state-q FE)	bite × post	+0.087	0.039	[+0.011, +0.163]
triple (fac+state-q FE)	medicaid × post	−0.136	0.065	[−0.263, −0.009]
triple (fac+state-q FE)	**bite × med × post**	**+0.153**	**0.034**	**[+0.088, +0.219]**
triple_no_sq_fe (fac+quarter FE)	treat × post	−0.148	0.069	[−0.283, −0.013]
triple_no_sq_fe	bite × post	+0.040	0.036	[−0.030, +0.110]
triple_no_sq_fe	medicaid × post	−0.033	0.055	[−0.141, +0.075]
triple_no_sq_fe	bite × med × post	+0.161	0.040	[+0.083, +0.240]

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.

Table 5. Rurality × bite

Bite quintile	Rural ATT (n=648 CCNs)	Urban ATT (n=3,467 CCNs)
Q1	−0.431 (0.113)	−0.338 (0.033)
Q2	−0.256 (0.033)	−0.192 (0.020)
Q3	−0.091 (0.035)	−0.215 (0.018)
Q4	−0.087 (0.038)	−0.170 (0.014)
Q5	**+0.076 (0.029)**	+0.004 (0.015)
Q1→Q5 spread	**0.51**	0.34

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.

Table 6. Resident-mix robustness (binary `treat` × `post`)

Specification	Controls added	treat × post coef	SE
A bare	facility + quarter FE	−0.145	0.066
B + bite × post	bite × post added	−0.140	0.068
C + 7 mix × post	+ percent White/Black/Hispanic, age, acuity, occupancy, percent Alz/dem × post	**−0.139**	**0.060**

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.

The headline binary `treat` × `post` is robust to LTCFocus resident-mix controls; the binary placebo concern is therefore not a case-mix artifact.

Appendix B. Selected pre-trend and HonestDiD diagnostics

Joint pre-trend Wald chi^2(7) tests (binary CSA event-study, leads e=−8 to e=−2):

Outcome	chi^2(7)	p
Total HPRD	46.7	<0.001
RN HPRD	12.0	0.10
CNA HPRD	65.4	<0.001
LPN HPRD	42.1	<0.001

Notes: This table summarizes the quantities listed in the rows and columns. It is intended to clarify the sample, comparison, and main empirical objects used in the surrounding text.

HonestDiD (Rambachan-Roth 2023) post-period coefficient summary, Total HPRD, $\bar{M}$ = 1.5:

At $\bar{M}$ = 1.5, the robust 95% CI excludes zero at event times e = 2 through e = 12. The post-period coefficients themselves are e=2: −0.063; e=4: −0.174; e=8: −0.156; e=12: −0.211. Only e = 0 and e = 1 lose significance under the worst-case bound. Detailed CI bounds at $\bar{M}$ in {0.5, 1.0, 1.5} are in `analysis/robustness/10_honestdid_bounds.csv`.

Goodman-Bacon decomposition (Total HPRD):

Comparison	Weight	Avg beta	Weighted contribution
Treated vs never-treated	0.813	−0.123	−0.102
Earlier vs later (clean)	0.187	+0.010	+0.004
Later vs already-treated	0.000	—	0.000

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.

TWFE-implied beta (Bacon weighted sum) = −**0.098**, close to the CSA simple aggregate (−0.143). No "later-vs-already-treated" contamination.

Appendix C. Reproducibility

All analysis is in `analysis/` and reproduces from cleaned data via `analysis/run_all.py` (approximately12–15 minutes). Key entry points:

- `analysis/01_main_csa_continuous.py` — CSA binary + event-study.
- `analysis/02_main_dose_response.py` — bite-quintile dose-response.
- `analysis/03_main_twfe_benchmark.py` — TWFE benchmark.
- `analysis/robustness/05_pretrends_eventstudy.py` through `analysis/robustness/10_honestdid_bounds.py` — six robustness specifications.
- `analysis/heterogeneity/11_medicaid_share_bite.py`, `12_rurality_bite.py`, `13_resident_mix_controls.py` — heterogeneity analyses.

Per-script logs in `analysis/log/`; tables in `analysis/tables/`; figures (PNG + PDF) in `analysis/figures/`. Cleaned data panels in `data/clean/`; raw inputs (treated as immutable) in `data/raw/`. Bite construction is canonical (Section 3.3); state-floor tracker is under `data/raw/state_staffing_floor_tracker_2017_2025.csv` with 100% primary-source URL coverage.