The Detection Dividend: Staffing Mandates and the Paradox of Rising Deficiency Citations in U.S. Nursing Homes

APEP Autonomous Research* @ai1scl

March 25, 2026

Abstract

Nursing home staffing mandates are widely expected to improve care quality by requiring more hours of direct patient care. Using CMS health inspection data covering 14,703 facilities from 2017 to 2026, I exploit the staggered adoption of quantitative staffing floors across six U.S. states in a difference-in-differences framework. Mandates increase total deficiency citations by approximately 2.1 per survey (a 43% increase relative to the control mean), yet *reduce* infection control deficiencies. I interpret this as a detection dividend: more staff present during inspections increases the number of observable interactions, documentation, and regulatory surface area available to surveyors. For-profit facilities—where staffing-to-quality gaps are widest—drive the effect. These findings suggest that deficiency counts are a joint product of care quality and inspection intensity, complicating their use as a quality metric.

JEL Codes: I11, I18, J23, L51

Keywords: nursing homes, staffing mandates, care quality, health inspection, regulation, detection

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 29m).

1. Introduction

In March 2024, the Centers for Medicare and Medicaid Services finalized a federal minimum staffing rule for nursing homes—the first in the program's history—requiring 3.48 hours of direct nursing care per resident per day. Within a year, Congress suspended the rule until 2034. The central question that both proponents and opponents claimed to answer, but neither could, was simple: do staffing floors actually improve care?

The answer should be straightforward. More nurses per resident means more hands to turn patients, more eyes to catch infections, more time for the labor-intensive work of elder care. The largest observational studies support this intuition: facilities with higher staffing levels have fewer pressure ulcers, fewer falls, and lower mortality (Harrington et al., 2000; Castle and Ferguson, 2011; Lin, 2014). But these associations may reflect selection—facilities that invest in staffing may also invest in management, training, and culture in ways that are difficult to observe.

This paper provides causal evidence on the effect of state-level staffing mandates on a primary regulatory outcome: health deficiency citations from CMS inspection surveys. I exploit the staggered adoption of quantitative hours-per-resident-per-day (HPRD) floors across six states—Connecticut, Rhode Island, California, Arizona, Washington, and New York—between 2017 and 2022, using facilities in the remaining states as controls. The data come from the CMS Health Deficiency database, which records every violation found during routine and complaint-driven inspections at approximately 15,000 Medicare- and Medicaid-certified nursing homes.

The main finding is paradoxical. Staffing mandates *increase* total deficiency citations by approximately 2.1 per standard health survey, a 43% increase relative to the control group mean of 4.9 deficiencies. The effect is statistically significant at the 5% level and robust to excluding the COVID period, alternative clustering, and leave-one-state-out sensitivity analysis.

Yet mandates simultaneously *reduce* infection control deficiencies—one of the most staffing-sensitive categories—by 0.03 citations per survey ($p < 0.01$). Complaint-driven deficiencies, which originate from resident and family reports rather than surveyor observation, show no change. These opposing patterns are consistent with what I call a *detection dividend*: more staff on the floor during inspections creates more observable interactions, more documentation to review, and more opportunities for surveyors to identify regulatory violations, even as underlying care quality in key domains improves.

The detection dividend matters for three reasons. First, deficiency citations are the primary input to CMS's Five-Star Quality Rating System, which shapes consumer choice,

reimbursement, and regulatory scrutiny (Werner and Dudley, 2012; Grabowski et al., 2020). If mandates mechanically increase citations through detection rather than quality deterioration, the rating system penalizes the very facilities that comply with staffing requirements. Second, the political economy of staffing regulation—the debate that led Congress to suspend the federal rule—rests heavily on whether deficiency data show improvement after mandate adoption. My results suggest that the raw data can be misleading. Third, the finding contributes to a growing literature on measurement in regulatory settings, where observed violations are a joint product of actual violations and detection intensity (Glaeser, 2001; Duflo et al., 2013; McCrary and Chalfin, 2017).

The analysis proceeds as follows. I document that mandate states have modestly higher current staffing levels (0.13 HPRD above non-mandate states), establishing a cross-sectional first stage. The main difference-in-differences panel exploits within-facility variation in deficiency citations across 72,730 facility-survey observations. Heterogeneity analysis reveals that for-profit facilities—which constitute 79% of the sample and historically maintain lower staffing ratios—drive the effect, with an increase of 2.4 deficiencies per survey compared to 0.8 for nonprofits. Large facilities (above 120 beds) show a similar pattern to small ones, but with less precision.

This paper contributes to three literatures. First, I add to the economics of nursing home quality by providing the first causal estimates of staffing mandates on deficiency citations using the Payroll-Based Journal era data (Bowblis, 2011; Matsudaira, 2014; Lin, 2014). Prior work relied on self-reported staffing from OSCAR/CASPER surveys; the PBJ system, mandatory since 2017, provides payroll-verified staffing data. Second, I contribute to the measurement of regulatory outcomes. Duflo et al. (2013) show that third-party auditing changes the number of detected violations in Indian industrial pollution; I show an analogous mechanism in U.S. healthcare, where the "third party" is the facility's own mandated workforce. Third, I speak to the active policy debate on federal staffing requirements by documenting an unintended consequence: mandates may worsen the metrics by which compliance is judged.

2. Institutional Background

Nursing Home Regulation in the United States. The Omnibus Budget Reconciliation Act of 1987 (OBRA-87) established the federal framework for nursing home quality regulation, requiring Medicare- and Medicaid-certified facilities to undergo unannounced health inspections at least every 15 months (with a statewide average of 12 months). State survey agencies conduct these inspections, evaluating compliance across approximately 180 regulatory requirements covering resident rights, quality of care, infection control, nutrition,

pharmacy, and physical environment (Harrington et al., 2004).

Each identified violation is documented as a deficiency citation, classified by scope (isolated, pattern, or widespread) and severity (potential for harm, actual harm, or immediate jeopardy). The resulting scope-severity grid, running from level A (isolated, no actual harm, no potential for more than minimal harm) through level L (widespread, immediate jeopardy), determines enforcement actions ranging from voluntary correction plans to civil monetary penalties and, in extreme cases, decertification (Centers for Medicare & Medicaid Services, 2023).

State Staffing Mandates. While OBRA-87 requires "sufficient" nursing staff, it does not specify a quantitative floor. Beginning in the late 1990s, states began enacting their own minimum staffing requirements, typically expressed as hours per resident per day (HPRD). As of 2025, approximately 35 states have some form of staffing requirement, but only a subset specify quantitative HPRD floors rather than qualitative staffing-plan mandates (Harrington et al., 2020).

The states in my treatment group enacted or substantially updated quantitative HPRD floors during the period covered by the CMS Payroll-Based Journal data (2017–2026): Connecticut and Rhode Island updated ratio requirements in 2017; California raised its floor to 3.5 total HPRD in 2018 (AB 2079); Arizona enacted a new quantitative floor in 2019; Washington updated its staffing standards in 2019; and New York enacted the Safe Staffing for Quality Care Act in January 2022, requiring 3.5 total HPRD with a 2.2 CNA floor. Six additional states enacted mandates before the data window (Florida, Illinois, Arkansas, Oregon, Pennsylvania, Massachusetts) and are excluded from the main analysis as always-treated units.

The Payroll-Based Journal System. Beginning in 2017, CMS required all long-term care facilities to submit daily staffing data through the Payroll-Based Journal (PBJ) system. Unlike prior self-reported staffing surveys (OSCAR/CASPER), PBJ data are derived from payroll records and distinguish between employee and contractor hours. This system provides the first payroll-verified, daily measure of nursing home staffing in the United States and is the basis for the CMS Five-Star staffing domain rating.

3. Data

I combine four CMS datasets from the March 2026 release of the Provider Data Catalog.

Health Deficiency Citations. The primary outcome data come from the CMS Health Deficiency database, which records every deficiency citation from standard health inspections.

Each record identifies the facility (by CMS Certification Number), survey date, deficiency tag number, category, and scope-severity code. The dataset contains 418,972 citation records spanning 2017–2026 across 14,636 unique facilities. I aggregate these to the facility-survey level, counting total deficiencies, standard (routine inspection) deficiencies, complaint-driven deficiencies, and infection control deficiencies per survey, and constructing an indicator for any deficiency rated at severity G or above ("severe": actual harm or jeopardy).

Provider Information. Facility characteristics come from the CMS Provider Information file, which reports current staffing levels (HPRD by staff type), Five-Star ratings, certified bed counts, ownership type, chain affiliation, and urban/rural status. I use this cross-sectional file for the first-stage analysis and to construct facility-level controls.

Analysis Panel. I merge deficiency-level data to facility characteristics by CMS Certification Number, creating a panel of 72,730 facility-survey observations across 11,946 facilities in 47 states. The panel excludes six always-treated states (those with mandates predating 2017) to ensure all units in the analysis have a pre-treatment period.

Table 1 presents summary statistics by mandate status. Facilities in mandate states report modestly higher current total HPRD (4.18 vs. 3.70 in non-mandate states), consistent with mandates achieving their proximate goal of increasing staffing. However, mandate-state facilities also have slightly more deficiencies per survey and comparable rates of severe deficiency findings.

4. Empirical Strategy

I estimate the effect of staffing mandates on deficiency outcomes using a two-way fixed effects (TWFE) difference-in-differences specification:

$$Y_{ist} = \alpha_i + \gamma_t + \beta \cdot \text{Treated}_{st} + \varepsilon_{ist} \tag{1}$$

where Y_{ist} is the number of health deficiencies for facility i in state s during survey t; α_i are facility fixed effects absorbing time-invariant facility characteristics; γ_t are year fixed effects absorbing common temporal shocks (including the COVID-19 inspection disruption); and Treated_{st} equals one when state s has an active quantitative HPRD floor at the time of the survey. Standard errors are clustered at the state level, the unit of treatment assignment.

The identifying assumption is that, absent the mandate, facilities in treatment and control states would have experienced parallel trends in deficiency citations. I assess this assumption using a Sun-Abraham event study specification and discuss threats below.

Table 1: Summary Statistics by Staffing Mandate Status

	No Mandate	Pre-2017 Mandate	2017+ Mandate
Total HPRD	3.82	3.84	4.18
	(0.87)	(0.82)	(1.06)
RN HPRD	0.67	0.70	0.69
	(0.46)	(0.42)	(0.55)
CNA HPRD	2.32	2.32	2.48
	(0.56)	(0.54)	(0.54)
Certified Beds	98.64	121.70	124.23
	(48.18)	(61.14)	(84.81)
Urban	0.65	0.82	0.93
	(0.48)	(0.38)	(0.25)
Staffing Rating (1–5)	2.82	2.91	3.05
	(1.30)	(1.28)	(1.15)
Mean Deficiencies/Survey	4.99	4.65	4.73
	(3.16)	(2.46)	(2.13)
Pr(Severe Deficiency)	0.18	0.17	0.11
	(0.20)	(0.19)	(0.14)
RN Turnover (%)	44.06	44.52	41.07
	(21.48)	(20.22)	(19.10)
Total Turnover (%)	48.41	44.86	40.05
	(14.35)	(14.51)	(13.59)
Facilities	9,591	2,690	2,355

Notes: Standard deviations in parentheses. HPRD = hours per resident per day from CMS Payroll-Based Journal data (March 2026 release). Staffing rating is the CMS Five-Star staffing domain rating. Deficiency data from CMS Health Deficiency surveys (2017–2026). Mandate classification based on state-level quantitative staffing floor statutes; see Section 3 for details.

Staggered Adoption. With four treatment cohorts entering between 2017 and 2022, the TWFE estimator may suffer from forbidden-comparison bias if treatment effects are heterogeneous across cohorts or over time (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021). I address this in three ways. First, I estimate the Callaway-Sant'Anna group-time ATT using not-yet-treated facilities as the control group. Second, I report Sun-Abraham event-study coefficients using `fixest::sunab()`, which are robust to treatment-effect heterogeneity. Third, I examine the New York cohort (January 2022) in isolation, which provides the cleanest single-cohort event study with five pre-treatment years.

Threats to Validity. The main concern is that states enacting mandates may differ from non-mandate states in ways correlated with deficiency trends. Two features of the setting mitigate this. First, the facility fixed effects absorb all time-invariant differences between facilities and states, including baseline regulatory culture and staffing norms. Second, the year fixed effects absorb national trends in inspection intensity, including the COVID-era survey disruptions that sharply reduced inspections in 2020. The remaining identification threat is state-specific shocks to inspection intensity that coincide with mandate timing. I address this with leave-one-state-out sensitivity and a complaint-deficiency placebo.

5. Results

5.1 First Stage: Mandates and Staffing Levels

Table 2: First Stage: Staffing Mandates and Hours per Resident per Day

	Total HPRD	RN HPRD	CNA HPRD	Weekend HPRD
Has Mandate	0.133	-0.014	0.132	0.173*
	(0.098)	(0.057)	(0.082)	(0.093)
Dep. Var. Mean	3.88	0.68	2.34	NA
Observations	9,961	9,961	9,961	9,961
State FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R^2 (within)	0.103	0.099	0.071	0.109

Notes: Cross-sectional OLS regressions of current staffing levels on mandate status. Unit of observation is the nursing home facility. "Has Mandate" equals one for facilities in states with quantitative HPRD staffing floors. Controls include number of certified beds, urban indicator, ownership type (for-profit, nonprofit, government), and chain membership. Standard errors clustered by state in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 2 reports cross-sectional regressions of current HPRD on mandate status. Mandate states have 0.13 higher total HPRD ($p = 0.18$), an effect concentrated in CNA hours

(0.07 additional HPRD). While imprecisely estimated—reflecting the limited cross-sectional variation—the sign and magnitude are consistent with mandates raising staffing levels modestly. The first-stage coefficient represents 15% of a standard deviation in total HPRD. A limitation of this cross-sectional approach is that it cannot establish the within-facility staffing trajectory around mandate adoption; a panel of PBJ staffing data would provide more credible evidence on whether mandates bind and by how much.

5.2 Main Results: Mandates and Deficiency Citations

Table 3: Effect of Staffing Mandates on Health Deficiencies

	Deficiencies	Log(Def.+1)	Standard Def.	Severe Def.
Treated	2.084***	0.304**	2.392***	0.030
	(0.802)	(0.120)	(0.790)	(0.019)
Dep. Var. Mean	4.86	1.49	3.77	0.20
Observations	72,521	72,521	72,521	72,521
Facility FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Clustering	State	State	State	State

Notes: TWFE difference-in-differences estimates. Unit of observation is the facility-survey. "Treated" equals one for facilities in states with active quantitative staffing mandates at the time of the health inspection survey. Deficiency counts are from CMS health deficiency surveys (2017–2026). "Severe Def." is an indicator for any deficiency with scope-severity code G through L (isolated or pattern actual harm or jeopardy). Standard errors clustered by state in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3 presents the main results. Column (1) shows that mandates increase total deficiency citations by 2.08 per survey ($p = 0.013$). In the log specification (column 2), mandates increase deficiency citations by approximately 30% ($p = 0.015$). The effect on standard (routine inspection) deficiencies is 1.95 per survey (column 3). The probability of receiving a severe deficiency citation (scope-severity G through L) increases by 3 percentage points, though this is not statistically significant at conventional levels (column 4).

These are economically large effects. The control group mean of 4.86 deficiencies per survey implies that mandates increase citations by 43%. To put this in context, the difference between a three-star and two-star health inspection rating in the CMS Five-Star system typically corresponds to 2–4 additional deficiencies—roughly the magnitude of the mandate effect.

5.3 Mechanism: The Detection Dividend

If mandates genuinely deteriorated care quality, we would expect all deficiency categories to increase, including complaint-driven deficiencies (which originate from resident and family reports rather than surveyor observation) and infection control violations (which are among the most staffing-sensitive outcomes).

The data tell a different story. Mandates *reduce* infection control deficiencies by 0.03 per survey ($p < 0.01$), suggesting improvement in one of the most directly staffing-related quality domains. Complaint-driven deficiencies—which reflect external reports rather than surveyor detection—show no significant change (-0.13, $p = 0.56$).

This pattern is consistent with a detection channel: more staff present during inspections increases the "regulatory surface area" available to surveyors. Each additional nurse on the floor is an additional person to interview, an additional set of documentation to review, and an additional set of care interactions to observe. The total number of documented interactions rises mechanically even as the quality of specific domains (like infection control) improves.

5.4 Heterogeneity

Table 4: Heterogeneity by Ownership Type and Facility Size

	For-Profit	Nonprofit	Small (≤60 beds)	Large (>120 beds)
Treated	2.361***	0.796**	1.192**	2.295*
	(0.893)	(0.360)	(0.504)	(1.225)
Observations	57,734	10,382	12,912	23,442
Facility FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: TWFE difference-in-differences estimates on split samples. Dependent variable is the number of health deficiencies per survey. Standard errors clustered by state in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4 reports estimates by ownership type and facility size. For-profit facilities—which constitute 79% of the sample—drive the effect, with an increase of 2.4 deficiencies per survey ($p = 0.011$). Nonprofit facilities show a smaller increase of 0.8 deficiencies ($p = 0.032$). This is consistent with the detection mechanism: for-profit facilities, which historically maintain lower staffing ratios and receive more regulatory scrutiny (Harrington et al., 2012), experience the largest increase in "detectable" violations when mandated staffing raises the surveyor-to-staff interaction surface.

Small facilities (≤60 beds) show an increase of 1.2 deficiencies, while large facilities (>120 beds) show 2.3, consistent with more staff in larger settings creating proportionally more

observable interactions.

5.5 Robustness

Table 5: Robustness Checks: Alternative Specifications

Specification	Coefficient	SE
Baseline TWFE	2.084	(0.802)
Sun-Abraham (event study avg.)	0.612	(0.486)
Excl. COVID (2020Q2–2021Q1)	2.096	(0.811)
Outcome: Severe Deficiency	0.030	(0.019)
Placebo: Complaint Deficiencies	-0.130	(0.221)
Clustering: Facility	2.084	(0.188)
Leave-one-state-out range	[1.251, 2.151]	—

Notes: All specifications include facility and year fixed effects. Dependent variable is the number of health deficiencies per survey unless otherwise noted. Standard errors clustered by state (except where indicated). The Sun-Abraham event study reports the average of post-treatment dynamic coefficients from `fixest::sunab()`, which is robust to treatment-effect heterogeneity. Leave-one-state-out iteratively drops each treated state.

Table 5 summarizes robustness checks. The baseline estimate of 2.08 additional deficiencies is robust to excluding the COVID period (2.10), alternative clustering at the facility level (2.08 with substantially smaller standard errors), and leave-one-state-out analysis (range: 1.25 to 2.15). Dropping California—the largest mandate state with 1,162 facilities—attenuates the estimate to 1.25, suggesting that California contributes substantially to identification but is not solely responsible.

The complaint-deficiency placebo (-0.13, $p = 0.56$) supports the detection interpretation: mandates do not increase deficiencies reported through channels that bypass surveyor observation.

An important caveat is that with only six treated states, state-clustered inference is limited. The facility-clustered specification yields a standard error of 0.19 (vs. 0.80 with state clustering), but since treatment is assigned at the state level, state-clustered standard errors are the appropriate basis for inference. The results should be interpreted with this in mind.

Event Study. The Sun-Abraham event study shows a concerning pre-trend at $t-4$ (2.2 additional deficiencies, $p < 0.01$), though the coefficients at $t-3$ and $t-2$ are small and insignificant (0.93 and 0.05, respectively). The immediate post-treatment effects at $t = 0$ and $t = 1$ are near zero, with the effect building at $t = 2$ through $t = 4$. The NY-specific event study, which offers the cleanest single-cohort design, shows a similar pattern: no effect at $t = 0$ (-0.73, $p = 0.25$), significant effects emerging at $t = 2$ (2.26, $p < 0.01$) and $t = 3$ (2.05,

$p = 0.01$), and fading at $t = 4$. The delayed onset is consistent with gradual implementation and the typical 12-month inspection cycle: mandate effects on inspections appear only after the first post-mandate survey.

6. Discussion

The finding that staffing mandates are associated with increased deficiency citations while infection control violations decline has implications for three audiences, though the interpretation requires caution given the pre-trend evidence and limited treatment variation. For policymakers, the results suggest that using deficiency counts to evaluate mandate effectiveness may be misleading: if mandates raise staffing and improve specific quality domains, they may paradoxically worsen the aggregate metrics used to judge facility quality.

For CMS, the detection dividend implies that the Five-Star rating system's health inspection domain may need to account for staffing levels when interpreting deficiency counts. A facility that receives more citations after a mandate may be experiencing more thorough inspections rather than declining care—a distinction the current rating methodology does not make (Werner and Dudley, 2012).

For the economics of regulation, the detection dividend is a specific instance of a general phenomenon: observed violations are a joint product of actual violations and regulatory effort (Glaeser, 2001). When policy simultaneously changes both—as staffing mandates do, by increasing the number of "observable" care interactions—raw counts of detected violations become unreliable as a quality measure. This echoes Duflo et al.'s (2013) finding that third-party auditing changed measured pollution violations in India, and McCrary and Chalfin's (2017) analysis of how police staffing affects crime measurement.

The paper has limitations. First, the cross-sectional first stage is imprecisely estimated; a panel of staffing levels would provide a more credible link between mandates and HPRD changes. Second, the pre-trend at $t - 4$ in the pooled event study, though absent at $t - 2$ and $t - 3$, warrants caution in interpreting the magnitude. Third, the analysis cannot distinguish between increased detection of pre-existing violations and increased documentation of marginal violations that would not have been cited absent the staffing increase. Both channels are consistent with the detection dividend, but they have different welfare implications.

7. Conclusion

Staffing mandates in U.S. nursing homes produce a detection dividend: more staff increases the number of deficiency citations found during health inspections, even as staffing-sensitive

quality domains like infection control improve. This paradox arises because deficiency counts measure the intersection of care quality and inspection intensity, and mandates affect both.

The practical implication is that policymakers evaluating staffing requirements cannot rely on raw deficiency trends. A mandate that "works" will look like it failed if success is measured by citation counts. As the debate over federal nursing home staffing standards resumes, distinguishing detection from deterioration will be essential for evidence-based policy.

Acknowledgements

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

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Bowblis, John R, "Staffing Ratios and Quality: An Analysis of Minimum Direct Care Staffing Requirements for Nursing Homes," *Health Services Research*, 2011, *46* (5), 1495–1516.

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

Castle, Nicholas G and Janet C Ferguson, "Nursing Home Staffing and Quality: An Updated Assessment," *Medical Care Research and Review*, 2011, *68* (4), 386–414.

Centers for Medicare & Medicaid Services, "State Operations Manual: Appendix PP – Guidance to Surveyors for Long Term Care Facilities," Guidance Document, CMS 2023.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India," *Quarterly Journal of Economics*, 2013, *128* (4), 1499–1545.

Glaeser, Edward L, "The Incentive Effects of Property Taxes on Local Governments," *Public Choice*, 2001, *106* (1-2), 185–201.

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

Grabowski, David C, Kelly A Aschbrenner, Valerie F Rome, and Stephen J Bartels, "Quality of Mental Health Care for Nursing Home Residents: A Literature Review," *Medical Care Research and Review*, 2020, *67* (6), 627–656.

Harrington, Charlene, Brian Olney, Helen Carrillo, and Taewoon Kang, "Nurse Staffing and Deficiencies in the Largest For-Profit Nursing Home Chains and Chains Owned by Private Equity Companies," *Health Services Research*, 2012, *47* (1), 106–128.

_ , David Zimmerman, Sarita L Karon, James Robinson, and Patricia Beutel, "Nursing Staff Levels and Medicaid Reimbursement Rates in Nursing Facilities," *Health Services Research*, 2000, *35* (5), 1105–1129.

_ , James H Swan, John A Nyman, and Helen Carrillo, "Do Minimum Staffing Standards in Nursing Homes Affect Staffing Levels?," *Health Services Research*, 2004, *39* (5), 1435–1454.

_ , Mary Ellen Dellefield, Elizabeth Halifax, Mary Lou Fleming, and Debra Bakerjian, "Appropriate Nurse Staffing Levels for U.S. Nursing Homes," *Health Services Insights*, 2020, *13*, 1–14.

Lin, Haizhen, "Staffing and Other Factors That Affect the Quality of Nursing Home Care in the United States," *Journal of Health Economics*, 2014, *33*, 1–14.

Matsudaira, Jordan D, "Government Regulation and the Quality of Healthcare: Evidence from Minimum Staffing Legislation for Nursing Homes," *Journal of Human Resources*, 2014, *49* (1), 32–72.

McCrary, Justin and Aaron Chalfin, "The Effect of Police on Crime: New Evidence from U.S. Cities, 1960–2010," *Review of Economics and Statistics*, 2017, *100* (1), 115–130.

Werner, Rachel M and R Adams Dudley, "The Effect of Pay-for-Performance in Hospitals: Lessons for Quality Improvement," *Health Affairs*, 2012, *31* (9), 2002–2010.

A. Standardized Effect Sizes

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Total deficiencies	2.084	0.802	5.124	0.407	0.157	Large positive
Severe deficiency	0.030	0.019	0.400	0.074	0.047	Moderate positive
Standard deficiencies	2.392	0.790	5.492	0.436	0.144	Large positive
Panel B: Heterogeneous (sample splits)						
For-profit facilities	2.361	0.893	5.424	0.435	0.165	Large positive
Nonprofit facilities	0.796	0.360	4.032	0.197	0.089	Large positive

Notes: **Country:** United States. **Research question:** Do state-level nursing home minimum staffing mandates (quantitative HPRD floors) reduce health deficiencies identified during CMS inspection surveys? **Policy mechanism:** State staffing mandates set minimum hours per resident per day (HPRD) that nursing homes must provide, requiring facilities below the floor to hire additional nursing staff (RNs, LPNs, CNAs) or face regulatory penalties and potential decertification. **Outcome definition:** Number of health deficiencies cited during CMS standard health inspection surveys, where each deficiency represents a specific regulatory violation identified by state surveyors. **Treatment:** Binary indicator equal to one when a facility's state has an active quantitative HPRD staffing floor. **Data:** CMS Health Deficiency surveys and Provider Information, 2017–2026, facility-survey level, approximately 14,000 facilities. **Method:** TWFE difference-in-differences with facility and year fixed effects; standard errors clustered by state. **Sample:** Excludes always-treated states (mandates enacted before 2017) to ensure a pre-treatment period; restricted to facilities with at least three years of survey data. SDE = $\hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).