

My Title

Joe White^{a,*}

^a*Department of Economics, Miami University, Oxford, OH, USA*

Abstract

1. Introduction

2. Data

This paper employs three main data sources, all of which are maintained by the Center for Medicare and Medicaid Services (CMS). The first is the Nursing Home Compare (NHC) Archive, which contains monthly facility level information on ownership, chain affiliation, certification status, and quality measures. The second is Healthcare Cost Report Information System (HCRIS) which contains annual cost reports submitted by every Medicare certified skilled nursing facility (SNF) in the United States. The cost reports contain information about costs, revenue, and other financial data for each reporting facility. Lastly, we use the Payroll Based Journal (PBJ) system, which provides facility level, monthly staffing hours by staff type.

2.1. Sample selection

Our sample is constructed at the facility-year-month level. We create our sample using the following criteria. First, we limit our sample to the 48 contiguous U.S. states from the period Jan 2017 - June 2024. Secondly, the Care Compare Archive data is matched to the HCRIS data on a unique nursing home provider number. Because HCRIS only contains information on SNFs, our sample focuses on SNFs instead of all nursing facilities in general.

2.2. Changes in ownership

To our knowledge there are no datasets that have information about ownership changes for all homes within the United States. Therefore, we manually assemble this dataset. We collect ownership data from the CMS Care Compare Archive Data from the years 2017 through 2024. We choose this period as it is the most recent data published as well as standardized ownership filing. We identify possible ownership changes during this period, compare results to filings

*Corresponding author

in the Medicare Cost Reports, and assemble a dataset of matching flagged ownership changes. CMS ownership files provide monthly ownership shares for each owner associated with a facility. As a result of the advanced corporate ownership structure detailed previously, it is difficult to discern ownership changes while looking at every ownership level. Therefore, when flagging ownership changes, we look at the highest level of owner type in time period t . The order, highest to lowest, is defined as indirect owners, direct owners, and then individuals or organizations with a partnership interest. There are three main criteria we use when defining a change in ownership. First, we define a change in ownership (CHOW) when a majority of ownership shares turn over between month $t - 1$ and month t . Intuitively, this captures transactions in which control of the facility changes hands, including cases where a new operator acquires most or all shares. Secondly, when ownership percentages are partially or completely missing, we approximate shares by normalizing any reported percentages and, where necessary, assigning equal weights across remaining owners. To reduce false positives from purely legal restructuring, we do not flag a CHOW if at least 80 percent of the ownership stake remains within the same surname family (e.g., a change from Smith Holdings LLC to Smith Family Partners LP). Lastly, we validate ownership changes reported in the NHC ownership files against change of ownership indicators and dates from HCRIS. We retain facilities that either (i) have no ownership change in either source or (ii) have exactly one ownership change in each source with the CHOW date within six months of one another. For these facilities, we define the CHOW month as the first month in which the NHC data indicate a change. All dates are converted to calendar months, so changes occurring at the beginning or end of a month are treated identically.

2.3. Outcome variables

Our main outcomes are nursing hours per patient day (HPPD), constructed from the CMS PBJ data. This data includes daily hours worked by each staff type in each facility, together with the daily census from the Minimum Data Set (MDS). For each facility i , calendar month t , and staff type s , we first aggregate daily hours to the monthly level and then compute

$$\text{HPPD}_{i,t}^{(s)} = \frac{H_{i,t}^{(s)}}{R_{i,t}}, \quad (1)$$

where $H_{i,t}^{(s)}$ is the total hours worked by staff type s during month t and $R_{i,t}$ is the total resident days in that facility-month. Total HPPD is defined as the sum of RN, LPN, and CNA HPPD.

We construct two additional measures of quality. First, we define a monthly coverage ratio

$$C_{i,t} = \frac{\text{DaysReported}_{i,t}}{\text{DaysInMonth}_t},$$

where $\text{DaysReported}_{i,t}$ is the number of distinct days in month t with complete PBJ records for facility i . This measure allows us to identify incomplete data

in our sample and describe our data completeness. Second, we track gaps in the reporting for each facility. For each observation we compute *gap*, defined as the number of calendar months since the previous PBJ observation for that facility. Finally, we exclude a small number of observations with implausible staffing levels: months in which RN and LPN HPPD are both zero, total HPPD is below 1.5 or above 12, or CNA HPPD exceeds 5.25. These observations are outliers that likely reflect reporting errors rather than genuine variation in staffing. These exclusion criteria align with the criteria outlined in the CMS NHC Technical Users Guide.

3. Methods

$$Y_{it}^k = \beta_{\text{post}}^k \text{Post}_{it} + X_{it}'\gamma^k + \alpha_i + \lambda_t + \varepsilon_{it}^k, \quad (2)$$

$$Y_{it} = \sum_{\tau=-24}^{24} \beta_{\tau} \mathbf{1}\{\text{event_time}_{it} = \tau\} \cdot \text{Treat}_i + X_{it}'\gamma + \alpha_i + \lambda_t + \varepsilon_{it}. \quad (3)$$

4. Results

Table 1 reports the baseline two-way fixed effects estimates of the effect of ownership change on hours per patient day. Across specifications, ownership changes lead to meaningful reductions in nurse hours per patient day. RN staffing declines by approximately 0.03 hours per patient day (roughly 9 percent), CNA staffing falls by 0.05 hours (about 3 percent), and total HPPD decreases by 0.08 hours (roughly 3 percent). LPN staffing shows no statistically significant change. The estimates are nearly identical when excluding anticipatory months, consistent with the event study results.

Table 1: Two-Way Fixed Effects Estimates of *post* on Staffing Outcomes (Baseline)

	Outcomes			
	RN	LPN	CNA	Total
Panel A				
With anticipation	−0.031*** (0.005)	0.001 (0.006)	−0.052*** (0.009)	−0.082*** (0.012)
Without anticipation	−0.032*** (0.005)	0.000 (0.006)	−0.054*** (0.009)	−0.086*** (0.013)
Panel B				
With anticipation	−0.086*** (0.016)	0.004 (0.009)	−0.026*** (0.005)	−0.025*** (0.004)
Without anticipation	−0.092*** (0.016)	0.003 (0.009)	−0.028*** (0.005)	−0.026*** (0.004)

Notes: Each cell reports the coefficient on *post* with two-way clustered standard errors (by facility and month) in parentheses. Panel A reports levels (HPPD); Panel B reports logs (HPPD). Samples: *With anticipation* ($N_{\text{levels}} = 632, 231$; $N_{\text{logs}} = 632, 231$). *Without anticipation* ($N_{\text{levels}} = 628, 107$; $N_{\text{logs}} = 628, 107$).

All specifications include facility and month fixed effects and covariates: *government*, *non-profit*, *chain*, *beds*, *occupancy rate*, *percent Medicare*, *percent Medicaid*, and state case-mix quartile indicators.

Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2 examines whether these effects differ before and after the COVID-19 pandemic. The magnitude of the decline is roughly halved in the post pandemic period for RN, CNA, and total HPPD, suggesting that pandemic induced shifts in labor markets and staffing substantially reduced the impact of ownership transitions.

Table 2: TWFE Estimates of *post*: Pre- vs Post-pandemic Periods (Without anticipation)

	Outcomes			
	RN	LPN	CNA	Total
Panel A				
Without anticipation (Pre-pandemic)	−0.029*** (0.006)	−0.016* (0.008)	−0.063*** (0.013)	−0.108*** (0.019)
Without anticipation (Pandemic)	−0.018** (0.007)	−0.000 (0.008)	−0.041*** (0.015)	−0.058*** (0.020)
Panel B				
Without anticipation (Pre-pandemic)	−0.068*** (0.024)	−0.009 (0.012)	−0.032*** (0.006)	−0.033*** (0.006)
Without anticipation (Pandemic)	−0.063*** (0.021)	0.002 (0.011)	−0.020** (0.008)	−0.019*** (0.006)

Notes: Each cell reports the coefficient on *post* with two-way clustered standard errors (by facility and month) in parentheses. Panel A reports levels (HPPD); Panel B reports logs (HPPD). Sample sizes: Row 1 ($N_{\text{levels}} = 255,005$; $N_{\text{logs}} = 255,005$), Row 2 ($N_{\text{levels}} = 359,413$; $N_{\text{logs}} = 359,413$).

All specifications include facility and month fixed effects and covariates: *government*, *non-profit*, *chain*, *beds*, *occupancy rate*, *percent Medicare*, *percent Medicaid*, and state case-mix quartile indicators.

Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Pre-pandemic 2017/01–2019/12; Pandemic 2020/04–2024/06.

Table 3 compares effects for facilities that began the study period as part of a chain and those that did not. Staffing reductions are noticeably larger among non chain facilities. CNA HPPD falls by 0.034 hours in chain facilities and by 0.079 hours in non chains, and total HPPD declines by 0.062 versus 0.113 hours, respectively. On a log scale, these differences translate into substantially larger percentage declines for non-chain facilities. LPN staffing again shows minimal movement.

Table 3: TWFE Estimates of *post*: Chain vs Non-chain Facilities (2017Q1 Baseline, Without anticipation)

	Outcomes			
	RN	LPN	CNA	Total
Panel A				
Without anticipation (Chain 2017Q1)	-0.031*** (0.007)	0.003 (0.008)	-0.034*** (0.011)	-0.062*** (0.015)
Without anticipation (Non-chain 2017Q1)	-0.037*** (0.009)	0.003 (0.011)	-0.079*** (0.019)	-0.113*** (0.025)
Panel B				
Without anticipation (Chain 2017Q1)	-0.099*** (0.021)	0.008 (0.013)	-0.019*** (0.006)	-0.020*** (0.005)
Without anticipation (Non-chain 2017Q1)	-0.085*** (0.030)	0.013 (0.017)	-0.037*** (0.010)	-0.032*** (0.008)

Notes: Each cell reports the coefficient on *post* with two-way clustered standard errors (by facility and month) in parentheses. Panel A reports levels (HPPD); Panel B reports logs (HPPD). Sample sizes: Row 1 ($N_{\text{levels}} = 322,100$; $N_{\text{logs}} = 322,100$), Row 2 ($N_{\text{levels}} = 243,036$; $N_{\text{logs}} = 243,036$).

All specifications include facility and month fixed effects and covariates: *government*, *non-profit*, *chain*, *beds*, *occupancy rate*, *percent Medicare*, *percent Medicaid*, and state case-mix quartile indicators.

Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Baseline chain classification determined by facility status in 2017Q1.

Overall, ownership changes result in sizable and robust reductions in staffing, with heterogeneity across periods and ownership structures. The patterns strongly suggest that ownership transitions worsen staffing conditions for residents, particularly in facilities that were not part of a chain and in the pre pandemic era.

5. Discussion