NBER WORKING PAPER SERIES

# GOVERNMENT MONITORING OF HEALTH CARE QUALITY: EVIDENCE FROM THE NURSING HOME SECTOR

Yiqun Chen
Marcus Dillender

Working Paper 34037
http://www.nber.org/papers/w34037

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2025, Revised August 2025

We are grateful to Sandy Black, Colleen Carey, Dave Chan, Amy Finkelstein, Ashvin Gandhi, Jacob Goldin, Lee Lockwood, David Molitor, Steve Rivkin, Adam Sacarny, Dan Sacks, and many seminar and conference participants for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Government Monitoring of Health Care Quality: Evidence from the Nursing Home Sector
Yiqun Chen and Marcus Dillender
NBER Working Paper No. 34037
July 2025, Revised August 2025
JEL No. D22, I11, I18, L21, L23, L51, L88

**ABSTRACT**

In contracting out, monitoring is an important policy tool to extract information on firm quality and incentivize quality provision. This paper examines a central quality inspection of nursing homes, a sector with significant welfare implications but widespread public concerns about the quality of care. Using data on nursing homes across the US, we find that nursing homes exhibit strategic responses to the inspection. Nursing homes increase the quantity and quality of labor inputs, reduce admissions, increase temporary discharges, and improve patient care in response to the inspection. However, nearly all responses drop immediately once the inspection is completed. While inspection ratings are unlikely to reflect nursing homes' absolute quality given the strategic responses, using a quasi-experimental research design we find that inspection ratings predict nursing homes' relative quality. Finally, we examine the effects of quality deficiency citations issued by the inspection on incentivizing nursing homes to improve quality of care, finding mixed impacts.

Yiqun Chen
University of Illinois Chicago
Department of Economics
and NBER
yqchen@uic.edu

Marcus Dillender
Vanderbilt University
Department of Economics
and NBER
marcus.o.dillender@vanderbilt.edu

# 1 Introduction

In contracting with private firms to provide social services (e.g., health care and education), governments often face firms' misaligned incentives to invest in the quality of services (Hart, Shleifer, and Vishny 1997). Monitoring is an important regulatory tool to extract quality information and identify and penalize quality deficiencies to incentivize quality provision. Yet firms may respond strategically by temporarily enhancing quality when they are inspected and dropping efforts when they are not. It is unclear whether government monitoring can successfully capture and incentivize quality or merely introduces wasteful regulatory costs. Understanding firms' responses to monitoring and the effectiveness of monitoring in revealing and fostering quality is highly relevant given the government's reliance on monitoring for many important quality regulations in the US.[1]

This article studies government monitoring in the health care sector. Nursing homes, primarily operated by private entities but largely financed by the government (through Medicare and Medicaid), are an industry with large economic size but widespread public concerns about the quality of care.[2] Over 1.2 million patients resided in nursing homes in 2024 (Chidambaram and Burns 2024). The annual spending on nursing home care totaled \$193 billion in 2022, equivalent to one-seventh of the national expenditures on hospital care (CMS 2024). Yet quality shortfalls in nursing homes have been a persistent public concern (e.g., Rau 2018; Jacobs 2024). To monitor and incentivize care quality, CMS mandates annual health inspections of nursing homes, which are onsite evaluations of nursing home quality of care. A health inspection often lasts a few days (U.S. Senate Special Committee on Aging 2023). Health inspections are a central quality regulation in the US nursing home sector. They also play a significant role in the allocation of patients and resources across facilities: the inspection results determine nursing homes' eligibility to receive Medicare and Medicaid reimbursements and form the core of the five-star rating for nursing homes, a CMS quality rating that impacts patients' and their families' choice of facilities, health care professionals' referral decisions, and regulators' funding decisions (e.g., Thomas 2014; Werner, Konetzka, and Polsky 2016; 12 U.S.C. §1715w).[3]

We begin by examining nursing homes' responses to health inspections. Using comprehensive payroll-based staffing data for the near-universe of nursing homes in the US and detailed patient data from the

1. Across a range of important areas, including health care, education, workplace safety, and food safety, agencies such as the Centers for Medicare & Medicaid Services (CMS), federal and state education agencies, the Occupational Safety and Health Administration (OSHA), and the Food and Drug Administration (FDA) consistently monitor facilities' quality compliance.

2. Medicare and Medicaid are the primary payers for three-quarters of nursing home residents, yet only six percent of nursing homes are government-owned (Chidambaram and Burns 2024).

3. The five-star rating consists of ratings from three components: health inspections, staffing, and quality measures. The five-star rating is mainly based on the health inspection rating (a composite of ratings from regular health inspections, which we study, and complaint investigations), adjusted only when staffing and quality measures receive a very high or very low rating. In addition, if the health inspection rating is one star, the nursing home's overall rating is capped at two stars (see details in CMS 2018a).

Minimum Data Set (MDS), we compare daily outcomes within cells of nursing homes and time categories (year, month, and day of the week). Under the plausible assumption that conditional on our baseline controls (nursing-home-by-time-category indicators), daily outcomes would not deviate systematically around arbitrary days in the absence of inspections, the empirical design allows us to study nursing homes' responses to health inspections.

We find that nursing homes exhibit significant changes in multiple core dimensions in response to health inspections. First, we find clear evidence of increased quantities of labor inputs. Nursing homes significantly increase the work hours of nurses and nurse assistants, the main nursing home staff, in response to health inspections. On the day of peak response, the work hours of nurses and nurse assistants rise by 10.2 and 6.6 hours on average, respectively. These effect sizes are 26 percent and 38 percent of the differences in nurse and nurse assistant hours between weekdays and weekends, respectively.[4] We also find similar increases in the work hours of every other leading type of staff observed, including physical therapists, occupational therapists, social workers, activities professionals, administrators, physical therapy assistants, occupational therapy assistants, dietitians, feeding assistants, and speech/language pathologists. Second, we find a modest but clear rise in the quality of labor inputs as indicated by skill and credentials in response to health inspections. The share of nursing staff hours provided by nurses (higher-skilled) as opposed to nurse assistants (lower-skilled) and the share of nurse hours provided by registered nurses (RNs, higher credential) as opposed to licensed practical nurses (LPNs, lower credential) both increase in response to the inspection.[5] Third, we find that nursing homes exhibit declined admissions in response to health inspections. On the peak day, admissions decline by nine percent. Fourth, we find a significant increase in temporary discharges, i.e., the number of patients who are discharged but return to a nursing home within two weeks, in response to health inspections.[6] On the peak day, temporary discharges increase by four percent. Decomposition shows that the increased temporary discharges are nearly all driven by patients returning to their original nursing homes (rather than to other facilities) and that the original nursing homes largely hold these patients in hospitals before readmitting them. Both declined admissions and increased temporary discharges can raise the staff-to-patient ratio. Finally, we find suggestive evidence that nursing homes improve the quality of resident care in response to health inspections as measured by residents' influenza and pneumococcal

4. Due to variation in inspection length, the share of nursing homes with an ongoing inspection lasting the entire day increases as the day approaches the final day of each inspection, and the magnitude of the estimated responses varies accordingly. The final day of each inspection may not last the entire day, potentially leading to a smaller response on the final day than on the day before.

5. Nursing staff includes nurses and nurse assistants; nurses include RNs and LPNs. RNs typically complete two to six years of professional education, while LPNs generally undergo only about one year of training and provide care under RNs' direction; nurse assistants mostly acquire only a few months' training and work under the supervision of RNs and LPNs (Bureau of Labor Statistics 2024a; 2024b; 2024c).

6. We choose two weeks as the cutoff because it is likely beyond the inspection period but still close to the discharge day. Results are qualitatively similar with a longer cutoff of three weeks or a shorter cutoff of one week.

vaccination rates. We also find slight declines in the use of chemical restraints (antipsychotics), though the estimates are less precise.[7]

However, nearly all responses described above drop immediately once the inspection is completed. The work hours of all types of staff decline to the typical levels (i.e., those on days relatively distant from the inspection), or even slightly below the typical levels, on the day right after the inspection. Similarly, admissions and temporary discharges immediately return to the typical rates and the quality of labor inputs declines to even below the typical level once the inspection is completed. For vaccination rates, since the increases persist as long as vaccinated patients remain in the facility, we do not observe an immediate return to the typical levels postinspection. However, the increases diminish within two weeks as earlier patients are gradually replaced by new admissions, consistent with decreased vaccination efforts postinspection.

We show empirical evidence supporting the validity of our estimation. First, we present dynamic effects and find that, except for the short period around the inspection, our examined outcomes are well balanced across days (conditional on the baseline controls: nursing-home-by-time-category indicators). Second, we address concerns about concurrent events. We show that inspection dates are spread nearly evenly throughout the year (except for declines around holidays when inspectors may be less likely to work). We also show that our estimates are robust to excluding inspections conducted in any week of the year (including the exclusion of holidays) and that inspection timing has little correlation with various nursing home characteristics.

Next, we examine correlates of nursing homes' responses to health inspections. We find that staff hour increases are larger among for-profit nursing homes, potentially reflecting financial incentives since inspection ratings can affect demand. The increases are also larger among nursing homes with lower ratings in prior cycles, suggesting stronger response incentives with lower existing ratings. In addition, staff hour increases are greater among nursing homes in more competitive areas, consistent with demand-side competition enhancing provider incentives to invest in quality (or quality ratings), and are greater among nursing homes in areas with higher staff availability, indicating that the capacity to respond plays a role. For admissions and temporary discharges, we find that admission declines are more pronounced in competitive areas and temporary discharges rise more notably among hospital-affiliated nursing homes; the latter is

7. The CMS contact for the health inspection data notes that inspectors are required to report inspection end dates but not start dates, and although start dates can be reported, there are reliability issues: despite reported start dates, inspections can begin later due to rescheduling or other issues that arise. The lack of reliable data on inspection start dates limits the ability to rigorously evaluate whether the strategic responses emerge before an inspection begins. It is also worth noting that observing some responses emerging only when the inspection starts may not exclude preinspection responses. E.g., when on-duty staff schedules are set in advance (a frequent practice in health care settings), an increase in staff hours during inspections could imply preinspection responses. Also, interestingly, we find that the day of peak response for strategic changes that may be more efficient for nursing homes to implement before inspections (e.g., increasing temporary discharges and vaccinations) is earlier than that for improvements that could be more valuable when observed by inspectors (e.g., increased staffing levels), indicating potential preinspection responses. While inspections are mandated as unannounced (CMS 2023), anecdotes suggest that inspection dates may not be strictly hidden from nursing homes (e.g., Thomas 2014; Rau 2018).

consistent with hospital affiliation facilitating transfers between nursing homes and hospitals. Together, empirical evidence suggests that both incentives and the capacity to respond may drive the gaming behavior.

The strategic responses suggest that the quality rating from health inspections is unlikely to reflect nursing homes' *absolute* quality compared to their typical levels, a finding highly relevant to a range of important nursing home regulations that rely on facilities' absolute quality assessments, including enforcing compliance with quality standards and certification that regulates nursing homes' eligibility to admit Medicare and Medicaid patients. Next, we ask whether the rating is predictive of nursing homes' *relative* quality. We measure quality by patient survival, a well-defined and unambiguously welfare-relevant outcome. To account for non-random sorting of patients into nursing homes, we leverage an instrumental variables (IV) design. Specifically, building on McClellan, McNeil, and Newhouse (1994), we classify nursing homes into two groups based on their quality ratings from health inspections and use the differential distance from a patient's zip code of residence to the nearest high-quality-rated (above-median) nursing home versus the nearest low-quality-rated (below-median) nursing home as an instrument for admission to high-quality-rated nursing homes. Using the administrative Medicare claims data, our sample includes 5.6 million cases in over 15,000 nursing homes across the nation. We find that the differential distance strongly predicts whether a patient is admitted to a high- versus low-quality-rated nursing home. Conditional on our baseline controls for the IV design (county-by-year indicators), patient predicted 90-day mortality generated using a rich set of demographics and comorbidities is well balanced across values of our instrument. In contrast, patient actual 90-day mortality rises significantly with the instrument. The IV estimate suggests that staying at high-quality-rated nursing homes is associated with a decline of 0.8 percentage points in 90-day mortality, a 5.2 percent reduction relative to the mean. Interestingly, the relationship is more pronounced among nursing homes with a smaller staffing response to health inspections, suggesting that quality ratings are more effective at predicting quality when strategic responses are weaker. The mortality decline is observed over longer horizons and persists for at least one year after admission. We further estimate each nursing home's effect on patient mortality using a control function approach that follows Einav, Finkelstein, and Mahoney (2025), finding that 10.3 percent of the variance in nursing homes' mortality effects is explained by quality ratings. Together, despite gaming, inspection ratings on average are predictive of nursing homes' relative quality.

Finally, we examine the role of health inspections in incentivizing nursing home quality of care improvements, through their key component designed to induce quality investments: quality deficiency citations. The citations are issued to facilities found in violation of quality standards by health inspections and can lead to fines. The potentially tougher reputation of regulators and the salience of penalties resulting from citations may incentivize nursing homes to improve care to avoid future citations. On the other hand, policies may not always achieve their intended effects. We use an event-study design that compares outcome changes at the cited nursing homes following a citation with changes at non-cited facilities. Examining the effects of 14 key citations with measurable outcomes, we find mixed results. Some citations lead to improvements in the cited quality dimensions: the effects emerge in the citation quarter or the following quarter and persist over time. Yet for citations that appear to have less precise evaluation criteria and lower financial penalties, we find limited effects. Supporting the validity of our estimation, we show that the treatment and control groups do not exhibit diverging trends precitation in most instances and that in the few instances where pre-trends seem to exist, they are unlikely to qualitatively affect our findings. We also present a series of robustness checks, including correcting for potential bias in staggered treatment timing using various estimators (De Chaisemartin and d'Haultfoeuille 2020; Callaway and Sant'Anna 2021; Sun and Abraham 2021; Borusyak, Jaravel, and Spiess 2024) and excluding days around health inspections to limit the influence of inspections, all of which support the validity of our estimation. Lastly, we examine whether citations have spillover effects on non-cited quality dimensions and nearby facilities, finding limited evidence.

This paper contributes to several strands of literature. First, this paper contributes to a large literature studying the production of health care quality. This body of research has examined the role of factors including health care providers' skill (e.g., Currie and MacLeod 2017; Chan, Gentzkow, and Yu 2022), the efficient organization of the physician workforce (e.g., Epstein, Ketcham, and Nicholson 2010; Chen 2021; Branco et al. 2023), providers' financial incentives (e.g., Cutler 1995; Celhay et al. 2019; Gupta 2021), and competition (e.g., Bloom et al. 2015) in shaping health care quality. Our paper contributes to this literature by examining the role of inspections, an important regulatory tool, in capturing and incentivizing health care quality. Our clear visual evidence of providers' strong responses to inspections reveals important insights about the possible overrating of providers' quality performance and underdetection of quality deficiencies during inspections. We also find that, in the face of incomplete regulations, inspections may lead to unintended consequences, e.g., declined admissions and increased temporary discharges in our setting, that may not be socially efficient. Accounting for providers' strategic responses can lead to better public policy design. However, we also find that inspection ratings, though likely inflated, predict quality ranking, and quality deficiency citations can incentivize quality improvements when properly designed.

Second, our research relates to the broad empirical literature on government monitoring of firms, spanning product quality, workplace safety, tax enforcement, and environmental compliance (e.g., Dranove et al. 2003; Jin and Leslie 2003; Almunia and Lopez-Rodriguez 2018; Johnson 2020; Zou 2021; Shi 2024). Contributing to this literature, we show evidence that firms' gaming of government monitoring does occur.[8]

8. Relatedly, Zou (2021) finds that intermittent monitoring of air quality leads local areas to temporarily suppress polluting

This finding has important implications for regulations that rely on inspection-based assessments of firms' absolute quality, including certification and quality standard enforcement. Our finding that firms' capacity to respond may matter for strategic behavior aligns with the organizational economics studies that suggest firms' compliance with regulation is a function of their capability to respond to regulatory incentives (e.g., Boudreau 2024).

Third, this paper contributes to a growing economics literature on nursing homes, which has shown that nursing homes are responsive to financial incentives (e.g., Hackmann 2019; Gandhi 2020; Gupta et al. 2024; Hackmann, Pohl, and Ziebarth 2024), vary widely in their quality of care (e.g., Hackmann, Rojas, and Ziebarth 2022; Olenski and Sacher 2024; Einav, Finkelstein, and Mahoney 2025), and are capable of boosting the quality of attributes subject to scrutiny but may shirk on the unreported ones (e.g., Lu 2012). We contribute to the literature by examining the role of government inspections, an understudied but central quality regulation in the nursing home sector, in capturing and fostering nursing home quality of care.

The remainder of this article is organized as follows. Section 2 provides background on the nursing home industry and health inspections. Section 3 introduces our data and presents descriptive statistics. Section 4 examines nursing homes' strategic responses to health inspections. We also report heterogeneity in the responses to health inspections by nursing home and market characteristics. Section 5 explores whether inspection rating is predictive of nursing homes' relative quality. Section 6 examines the impacts of quality deficiency citations on nursing homes' quality of care. Section 7 concludes.

## 2 Institutional Background

### 2.1 Nursing Homes

Nursing home residents include both short-stay patients receiving post-acute rehabilitative therapy and long-stay patients requiring assistance with daily living activities. In the US, nursing homes are mostly operated by private entities but largely financed by the government through Medicare and Medicaid: only six percent of nursing homes are government-owned, yet Medicare and Medicaid are the primary payers for three-quarters of nursing home residents (Chidambaram and Burns 2024).[9] Most short-stay nursing home patients receive coverage from Medicare, while long-stay patients largely rely on Medicaid.[10]

activities on monitored days, and local governments' gaming reactions on monitored days appear to be an important mechanism. Relative to Zou (2021), we examine firms' gaming.

9. Nursing homes are often referred to as skilled nursing facilities (SNF), which are nursing homes certified by CMS to treat Medicare patients and account for more than 95 percent of nursing homes nationwide (National Center for Health Statistics 2024).

10. Medicare covers a maximum of 100 days of nursing home care for each benefit period. Patients can transition to Medicaid for stays exceeding Medicare's maximum coverage if they meet Medicaid's eligibility criteria.

The US nursing home industry is large. The annual patient volume ranged from 1.1 to 1.4 million over the past decade. In 2022, total spending reached $193 billion, equivalent to one-seventh of national hospital expenditures and four percent of total national health spending (Chidambaram and Burns 2024; CMS 2024). As baby boomers continue to age, spending on nursing home care and the number of nursing home residents are expected to grow sizably (Congressional Budget Office 2013).

However, quality shortfalls in nursing homes have been a source of tremendous concern. Patients frequently suffer adverse events during nursing home stays, with a sizable share of those events preventable.[11] Researchers and regulators typically find insufficient staffing levels and substandard staff quality in nursing homes (e.g., Geng, Stevenson, and Grabowski 2019; Government Accountability Office 2021), which an overwhelming consensus suggests as the sources of and potential remedies for a wide range of adverse outcomes among nursing home residents, including mortality (Friedrich and Hackmann 2021; Ruffini 2024).

## 2.2 Health Inspection

To evaluate and incentivize quality of care, CMS requires nursing homes participating in Medicare and Medicaid to undergo annual health inspections. CMS is charged with overseeing the inspection, and state agencies, typically state health departments, conduct the inspection on behalf of CMS. An inspection generally lasts a few days (U.S. Senate Special Committee on Aging 2023), during which a team of inspectors employed by states (often nurses) travels to the nursing home to conduct an onsite evaluation of care quality. During health inspections, the inspection team examines various aspects of nursing home care, such as staffing levels, staff quality, and resident care (e.g., influenza and pneumococcal vaccinations, the use of physical and chemical restraints, and assistance with daily living activities). When inspectors find nursing homes in violation of any quality standards, they issue quality deficiency citations for each violation observed, which are reported publicly and can result in fines.

As shown in Appendix Figure A.1, inspection dates are spread nearly evenly throughout the year (except for declines around holidays when inspectors may be less likely to work). Though health inspections are conducted on an annual basis, CMS does not require them to be conducted exactly every 365 days for individual nursing homes, as long as an inspection is within 15 months of the previous one and the state average interval between adjacent inspections for individual nursing homes does not exceed 12 months (CMS 2023). Panel A of Appendix Figure A.2 plots a histogram of the number of days between adjacent inspections, showing an approximately normal distribution with a wide range from below 300 days to above

11. For example, according to Office of Inspector General (2014), one-third of Medicare nursing home residents experienced an adverse event that results in harm (e.g., infections, falls, and pressure ulcers) during their nursing home stays, and 59 percent of those events are clearly or likely preventable.

500 days (standard deviation: 70 days).[12]

There exist strong financial incentives for nursing homes to perform well in health inspections. First, the performance rating from health inspections forms the core of the five-star rating for nursing homes,[13] which can significantly affect nursing home demand through its influence on patients' and their families' choices, health care professionals' referral decisions, and insurers' selection of preferred networks (e.g., Thomas 2014; Werner, Konetzka, and Polsky 2016; Government Accountability Office 2016). Second, nursing homes will receive quality deficiency citations, which are reported publicly, and consequently, financial penalties if they are found in violation of quality standards during health inspections. Both the probability of receiving citations and the amount of the resulting financial penalties are high. For example, in 2019, approximately 80 percent of nursing homes received at least one citation from health inspections,[14] and the average amount of fines is $ 66,367 per day conditional on receiving penalties (CMS 2025). Third, nursing homes found to be providing severely substandard care will be barred from receiving Medicare and Medicaid reimbursements, which are the primary source of revenues for nursing homes.

While continuous monitoring can yield more representative results, concerns over cost burdens limit its feasibility. To balance between the goal of obtaining representative inspection results and the costs of continuous monitoring, CMS requires inspection dates to be unannounced and unpredictable, to prevent nursing homes from anticipating when inspections will be conducted (CMS 2023). Yet concerns remain that nursing homes may expect health inspections and engage in "window-dressing." For example, as described in two *New York Times* articles (Thomas 2014; Rau 2018),

> *"Homes often know when an inspection will occur, and many of them have learned to add workers in the period leading up to it."*
>
> *"The homes sometimes anticipated when an inspection would happen and could staff up before it."*

# 3 Data

## 3.1 Data Sources

**Nursing Home Health Inspection, Rating, and Citation Data.** The data on nursing home health inspections are obtained from the Provider Data Catalog for Nursing Homes including rehabilitation facilities provided

12. Appendix Figures A.1 and A.2 are obtained using the data described in Section 3 below. There appears to be slight bunching at approximately 350 days in Panel A of Appendix Figure A.2, which may be due to state agencies trying to meet the requirement that the state average interval between adjacent health inspections for individual nursing homes should not exceed 12 months. The average interval between adjacent health inspections is 390 days, somewhat longer than the requirement of not exceeding 12 months.

13. Details are described in footnote 3.

14. This statistic is obtained using the data described in Section 3 below.

by CMS. The data include information on the identifier of the nursing home inspected, the end date of the inspection, and its rating. The data do not directly provide inspection start dates. The CMS contact for the data described that inspectors are required to report inspection end dates but not start dates, and while start dates can be reported, there are reliability issues: despite reported start dates, delays can occur due to rescheduling or other issues that arise. Inspection end dates are unlikely to be misreported as the reporting is mandated.[15] Thus, our analyses use inspection end dates rather than start dates. The data also report quality deficiency citations issued to nursing homes by health inspections, e.g., violating the requirement to "use a registered nurse at least 8 hours a day, 7 days a week" or the requirement to "hire a qualified dietitian." We use the data from 2014 to 2019, ending before the onset of the COVID pandemic in the U.S.[16]

**Staffing Data.** Our staffing data come from the Payroll Based Journal (PBJ), an administrative dataset that contains the near-universe of nursing homes in the US. The data report work hours for essentially every type of staff at nursing homes at the facility-by-day level. The data are submitted to CMS typically automatically through facilities' payroll platforms. CMS conducts audits on the data, and nursing homes found to submit false information are subject to penalties (i.e., receiving the lowest star rating for its staffing performance in the five-star rating (CMS 2018b)). Since the data are only available from 2017, in analyses using the PBJ data, we restrict attention to the years 2017-2019.

**Resident Data.** We use the MDS and the 100 percent Medicare claims data from 2014 to 2019. The MDS reports federally mandated, standardized assessments of the universe of residents in Medicare or Medicaid certified nursing homes, which account for more than 95 percent of nursing homes in the US (National Center for Health Statistics 2024). The data include information on patient health status, care received during the nursing home stay, admission and discharge dates, and discharge destinations. The MDS has been widely used in economic studies of nursing homes (e.g., Grabowski, Gruber, and Angelelli 2008; Hackmann 2019; Einav, Finkelstein, and Mahoney 2025; Gupta et al. 2024).

The 100 percent Medicare claims data are administrative claim records for the universe of traditional Medicare enrollees. The data include information on patient demographics, health history, and health care use. The data also report patient zip code of residence. Vital statistics documenting patient death dates are linked to Medicare claims, allowing us to measure patient mortality.

**Nursing Home and Area Characteristics.** We obtain nursing home characteristics from the "Long-Term Care: Facts on Care in the US" (LTCFocus) database for heterogeneity analyses and robustness checks controlling for nursing home characteristics. The LTCFocus provides data derived from administrative

15. The CMS contact for the health inspection data, BetterCare@cms.hhs.gov, shared the information.

16. The only exception is that when examining the effects of quality deficiency citations, we extend the data period through 2021 to identify nursing homes receiving citations beyond the analysis period of 2017-2019 (details provided below in Section 6).

datasets including the Certification and Survey Provider Enhanced Reporting (CASPER) database, MDS, and Medicare claims. We use the 2014-2019 LTCFocus for information on nursing homes' hospital affiliation, for-profit status, and number of beds. We also use nursing homes' county-level Herfindahl-Hirschman Index from the LTCFocus, as a proxy for nursing home competition in the local market.

We supplement the LTCFocus with the Area Health Resource File (AHRF). Specifically, we use county-level populations and numbers of RNs, LPNs, and nurse assistants from the 2018 AHRF (the earliest county-level AHRF file during our study period with all these variables) to calculate the numbers of RNs, LPNs, and nurse assistants per capita in each county.

### 3.2 Descriptive Statistics

Table 1 summarizes the rich set of outcomes in our analyses. We report the quantity of labor inputs as measured by the daily work hours of each type of staff in the PBJ data, the quality of labor inputs as measured by the share of daily nursing staff hours provided by nurses and the share of daily nurse hours provided by RNs, the number of admissions, the number of temporary discharges (cases who are discharged but return to a nursing home within two weeks), and patient care that includes influenza and pneumococcal vaccinations, physical restraint use, and antipsychotic use. Nursing staff consists of nurse assistants and nurses, with the latter being higher-skilled. Nurses further include RNs and LPNs, with RNs receiving more advanced training.[17] We calculate the share of daily nursing staff hours provided by nurses and the share of daily nurse hours provided by RNs as proxies for the quality of labor inputs. To be consistent with our regression sample, Table 1 limits observations to those between -100 and 100 days of the last day of each health inspection, with observations between -14 and 14 days excluded to restrict the influence of inspections.

Table 1 shows that our sample includes a large number of nursing homes, over 15,000 across the nation, and that there is a fair spread in each outcome. Nursing staff are the primary workforce in nursing homes based on hours worked, with an average of 322 hours per facility per day compared to an average below 15 hours for every other type of staff. About 40 percent of nursing staff hours are provided by nurses, and 41 percent of nurse hours are provided by RNs. The numbers of admissions and temporary discharges per nursing home per day are, respectively, 0.79 and 0.23. Approximately one-third and two-thirds of cases have received influenza and pneumococcal vaccinations, respectively, and 17 percent of cases received antipsychotics in the last seven days.[18]

17. As described in footnote 5, RNs generally complete two to six years of professional education; LPNs typically receive only about one year of training and provide care under RNs' direction; nurse assistants help patients with daily living activities (e.g., feeding, dressing, and bathing) and work under the supervision of RNs and LPNs (Bureau of Labor Statistics 2024a; 2024b; 2024c).

18. The sample size varies across panels in Table 1 due to two main factors: (i) Panels A and B include only data between 2017 and 2019 because the PBJ data is only available since 2017, while Panels C-E include data between 2014 and 2019; (ii) the unit of

# 4 Responses to Health Inspections

In this section, we investigate nursing homes' responses to health inspections. We find that nursing homes show strong responses to health inspections, in multiple core dimensions. First, we find that nursing homes increase the quantity of labor inputs: they increase the work hours of every main type of staff in response to health inspections. Second, we find that nursing homes show modest but clear increases in the quality of labor inputs as measured by nursing staff skill and credentials in response to health inspections. Third, nursing homes reduce admissions with health inspections. Fourth, nursing homes appear to temporarily discharge patients to hospitals and admit them back postinspection. Fifth, we find improved patient care in some dimensions (e.g., increased influenza and pneumococcal vaccinations) with health inspections. However, nearly all responses drop immediately once the inspection is completed. Next, we explore potential mechanisms behind the responses. We find heterogeneous responses across nursing homes that suggest both incentives and the capacity to respond may drive nursing homes' gaming behavior. Finally, we present a series of robustness checks.

## 4.1 Empirical Framework

**Estimation Specification.** Our empirical specification takes the following form:

$$y_{it} = \sum_{k=-30}^{k=30} \beta_k d_{it}^k + \mathbf{T}_{it}\theta + \epsilon_{it}, \tag{1}$$

where $i$ indicates nursing homes, $t$ indicates calendar days, and $y_{it}$ represents the outcome of interest. We use $d_{it}^k$ to denote event day indicators that take the value of one if day $t$ is $k$ days away from the last day of an inspection. Our analysis includes observations ranging from -100 to 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. In Section 4.4, we extend the reporting window from 30 days before and after the inspection to a longer period (75 days before and after) and show that the results are stable. The coefficients of interest are $\beta_k$, representing differences in outcomes between a day that is $k$ days away from the last day of an inspection and days in the reference group that are relatively distant from the inspection and thus unlikely to be affected by it.

observations is at the facility-by-day level in Panels A-D but at the patient assessment level in Panel E. The sample size is smaller in Panel B than in Panel A and varies slightly between outcomes in Panel B because a small number of observations have zero hours in the denominator when we construct the share variables. The sample size also varies slightly across outcomes in Panel E due to missing values. The number of observations in Panels A-D is smaller than the number of nursing homes multiplied by 201 (given that we include days [-100, 100]) and the number of years studied. This difference arises for two main reasons: (i) due to entries and exits, we do not observe each nursing home in each year; (ii) as described in footnote 12, the average length between adjacent health inspections is 390 days, i.e., the average number of health inspections is below one per nursing home per year.

The vector $\mathbf{T}_{it}$ contains our baseline controls: nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators. We condition on $\mathbf{T}_{it}$ to allow for systematic differences across nursing homes and time categories. For instance, staffing levels may differ between weekdays and weekends and across nursing homes. Controlling for $\mathbf{T}_{it}$ captures these potential systematic differences. Finally, $\epsilon_{it}$ is the error term. We cluster standard errors by nursing home.

**Analysis Sample.** We restrict the sample to observations between -100 and 100 days from the last day of each health inspection. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days. As shown in Table 1, we obtain a final sample of about seven million observations for outcomes pertaining to the quantity and quality of labor inputs, 17 million for admissions and temporary discharges, and 34 million for patient care outcomes. Our sample includes over 15,000 unique nursing homes across the nation.

**Identification.** Our identification assumption is that, conditional on $\mathbf{T}_{it}$, daily outcomes would not deviate systematically around arbitrary days in the absence of health inspections. We show evidence supporting the validity of the identification assumption. First, our results below show that except for the period close to inspections, the examined outcomes are well balanced across days (conditioning on $\mathbf{T}_{it}$).

Second, we examine concerns about concurrent events. Appendix Figure A.1 shows that except for declines around holidays when inspectors are less likely to work, inspection dates are spread nearly evenly throughout the year. We also present below in Section 4.4 the robustness of our estimates to excluding inspections conducted in any week of the year and to excluding holidays from our analysis sample. Additionally, we show that there is little correlation between inspection timing and various nursing home characteristics.

## 4.2 Main Results

**Quantity of Labor Inputs.** Figure 1 reports estimates of the effect of health inspections on the work hours of nursing staff, the main staffing inputs in nursing homes, using Equation (1). Panel A shows the results for nurses' work hours. Day 0 on the $x$-axis represents the last day of the inspection. We find increased nurse work hours from around day -10 to day 0. The magnitude of the estimate declines from day -1 to day -10, potentially because the variation in inspection length reduces the share of nursing homes with an ongoing inspection from day -1 to day -10. The magnitude of the estimate also declines from day -1 to day 0, which may be due to some inspections not lasting the entire day on the final day. On the day of peak response—i.e., day -1, which is likely the day with the highest share of nursing homes with an ongoing health inspection lasting the entire day—nurse hours rise by 10.2 hours, or 8.0 percent relative to the sample mean.[19] This effect size is large, equivalent to 26.4 percent of the difference in nurse hours between weekdays and weekends. However, nurse hours decline immediately once the inspection is completed, to even slightly below the typical level (i.e., that during the reference period, days [-100, -30) and (30, 100], which are relatively distant from the inspection and are unlikely to be affected by it as suggested by the economically negligible coefficients for days not near the inspection in Figure 1).

Panel B of Figure 1 shows the results for nurse assistant work hours. We find a similar pattern that nurse assistant hours rise in response to health inspections. On the peak day, nurse assistant hours increase by 6.6 hours, a 3.4 percent increase relative to the mean, or 38.0 percent of the difference between weekday and weekend nurse assistant hours. However, the increase drops immediately on the day after the inspection. Panels C and D of Figure 1 report results for RNs and LPNs (two subtypes of nurses) separately, showing patterns similar to those in Panels A and B of Figure 1. Consistent with the validity of our empirical design, Figure 1 shows that except for the period close to the inspection, the examined outcomes are balanced across days (conditional on $\mathbf{T}_{it}$). Though some estimates for days distant from the inspection are statistically significant given the large sample size, their magnitudes are negligible.

Figure 2 and Appendix Figure A.3 show the results for all other types of staff observed in the PBJ data. Figure 2 presents the results for the main types of staff, defined as those whose average daily work hours are at least one hour (except for nursing staff which are shown in Figure 1). Figure A.3 presents the results for less common types of staff, i.e., those whose daily average work hours are below one hour. Figure 2 shows increased work hours in response to health inspections for every main type of staff examined. The work hour increases on the peak day range from two percent to as large as about 40 percent relative to sample means. Appendix Figure A.3 shows that even for some less common types of staff whose daily work hours are below one hour (e.g., pharmacists and therapeutic recreation specialists), their work hours rise in response to health inspections. However, all work hour increases disappear immediately once the inspection is completed.

To summarize and mitigate concerns about multiple hypothesis testing, we calculate total staff hours, i.e., the sum of work hours for all staff in Figures 1 and 2 and Appendix Figure A.3. We show in Appendix Figure A.4 that, on the day of peak response (day -1), total staff hours increase by 23.6 hours, a 5.9 percent increase relative to the mean or one-fifth of the difference between weekdays and weekends. However, total staff hours decline on the day right after the inspection, to even slightly below the typical level.

In Appendix Figure A.5, we examine sources of increased staff hours by decomposing hours into those provided by employees and those provided by contract workers.[20] We find that for staff types primarily

19. To restrict the influence of health inspections, we estimate all sample means in this section using only days relatively distant from inspections: days [-100, -14) and (14, 100].

20. In addition to total hours for each type of staff, the PBJ data report hours worked by employees and contract workers separately.

composed of employees—e.g., nursing staff and administrators, who have significantly more employee hours than contract worker hours—the increased work hours associated with health inspections are almost entirely driven by employees. Nursing homes seem to have these workers work extra shifts in response to health inspections rather than hiring temporary contract workers. For staff types with a large share of contract workers (e.g., dietitians and physical therapists), the increased work hours arise from employees and contract workers in comparable magnitudes.

**Quality of Labor Inputs.** We next examine changes in the quality of labor inputs in response to health inspections. Nursing staff, the primary workforce in nursing homes as shown by their much higher work hours relative to other staff in Table 1, include nurses who are skilled and nurse assistants who are less skilled. Nurses further include RNs who hold a relatively higher credential and LPNs who hold a relatively lower credential. Panels A and B of Figure 3 examine the share of nursing staff hours provided by nurses and the share of nurse hours provided by RNs, respectively, finding modest but clear increases in both outcomes with health inspections. However, the increases disappear immediately after the inspection. While higher-skilled nursing staff can improve patient outcomes, they are costlier to hire: the median hourly pay for RNs, LPNs, and nurse assistants is, respectively, \$41.4, \$28.7, and \$18.4 (Bureau of Labor Statistics 2024a; 2024b; 2024c). Wage costs may discourage nursing homes from using higher-skilled staff, yet facilities appear willing to do so in response to health inspections, which may enhance inspection ratings directly through higher average staff qualifications and indirectly through better patient care.

A question is whether nursing homes overreport the quantity and quality of staff in the PBJ data. In principle, this possibility may exist. In practice, this concern is lessened for our empirical identification, which uses variation in staffing relative to inspection timing: overreporting would have to be correlated with inspections to confound our estimates. Facilities likely have limited incentives to overreport in response to inspections given that: (i) the staffing data are not submitted to the PBJ until after the quarter, i.e., inspectors do not observe the reported staffing during the inspection; and (ii) since inspections are conducted onsite which allows inspectors to directly observe staffing, over-reporting may entail greater costs than benefits given the high risk of detection. In addition, given (ii), if anything, overreporting may be less likely during inspections.

**Admissions.** Panel C of Figure 3 assesses nursing home admissions in response to health inspections. On the day of peak response, the number of admissions declines by nine percent relative to the mean. The decrease may be due to nursing homes reducing admissions to increase the staff-to-patient ratio, an important inspection area. It may also be because handling the inspection reduces nursing homes' capacity to admit new patients. Similar to the quantity and quality of labor inputs, admissions return to the typical level immediately after the inspection.

In Appendix Figure A.6, we decompose admissions by patient characteristics. We find similar declines in admissions for patients with high versus low predicted 90-day mortality, indicating that nursing homes do not seem to selectively lower admissions for risky or healthy patients, at least in terms of observable risks. This may be because the average number of admissions per day is low (0.79 in the absence of health inspections as shown in Table 1), leaving limited scope for selective admissions based on patient health risks. Appendix Figure A.6 also shows that admissions of less advantaged patients (e.g., disabled and Medicaid-eligible) do not seem to be disproportionately affected by health inspections.

**Temporary Discharges.** Panel D of Figure 3 displays changes in the number of temporary discharges—cases who were discharged but returned to a nursing home within two weeks—in response to health inspections. We use two weeks as the cutoff to obtain a period likely beyond the inspection but still close to the discharge. Results are qualitatively similar with a longer cutoff of three weeks or a shorter cutoff of one week (Appendix Figure A.7). Panel D of Figure 3 shows a marked increase in temporary discharges in response to health inspections: on the day of peak response, the number of temporary discharges increases by 4.3 percent relative to the mean. As it may be more efficient to discharge patients before the inspection, the day of peak response for temporary discharges is earlier than that for the quantity and quality of labor inputs and admissions. Temporary discharges can lower patient census and increase the staff-to-patient ratio, even if staffing inputs remain constant.

In Panel E of Figure 3, we decompose temporary discharges by patients' return destination: original versus other nursing homes. We find that the increased temporary discharges associated with health inspections stem almost entirely from patients returning to their original nursing homes. In Panel F of Figure 3, we further decompose patients who were temporarily discharged and returned to their original nursing homes by discharge destinations, finding that the increase in these cases with health inspections is nearly all driven by patients discharged to hospitals. Taken together, it seems that in response to health inspections, nursing homes transfer patients to hospitals and admit them back after the inspection, which may lead to unnecessary use of hospital resources.[21]

Appendix Figure A.8 decomposes temporary discharges by patient characteristics. We do not find that riskier patients experience a higher increase in temporary discharges due to health inspections. If anything, the increase appears to be more pronounced among younger, non-disabled, and non-black patients. Unnecessary hospital transfers may pose higher risks for fragile patients, or hospitals may be less willing to

21. Interestingly, despite increased temporary discharges, Panel C of Figure 3 shows no systematic rise in admissions postinspection. The temporary discharge increase, though clear, is modest in magnitude, and readmissions occur over a wide two-week period following discharge, limiting the possibility of admission spikes on a specific day postinspection.

accept fragile patients for temporary stays, both of which may limit the possibility of risky patients being disproportionately exposed to inspection-driven temporary discharges. Additionally, while we are unable to fully disentangle whether the increase in temporary discharges is driven by (i) over-discharges due to health inspections or (ii) under-discharges in the absence of inspections, the finding that riskier patients do not experience a larger increase is more consistent with (i).

**Patient Care.** Panels G–J of Figure 3 assess changes in patient care in response to health inspections.[22] We measure patient care in the following key dimensions observable in the MDS: whether the patient received (i) influenza vaccinations at the nursing home, (ii) pneumococcal vaccinations, (iii) antipsychotics (chemical restraints), and (iv) physical restraints.[23]

Panels G–H of Figure 3 show that both the share of patients receiving influenza vaccinations and the share of patients receiving pneumococcal vaccinations increase in response to health inspections. As with temporary discharges, the day of peak response for vaccinations appears to be earlier than that for the quantity and quality of labor inputs and admissions. Vaccinating patients shortly before inspections can present a high vaccination rate during inspections, while allowing nursing homes to allocate more staff to daily care during inspections to further enhance ratings.[24] Different from that for labor inputs, admissions, and temporary discharges, the changes in influenza and pneumococcal vaccination rates do not fully disappear right after the inspection. This may be because the increases persist as long as vaccinated patients stay in the nursing home. However, the increases fade over time postinspection as earlier patients are gradually replaced by new admissions, suggesting declined vaccination efforts after inspections.

In Panel I of Figure 3, we examine antipsychotic use (chemical restraints). Although the results are less precise, we observe a slight decline in antipsychotic use in response to health inspections. Panel J of Figure 3 shows limited changes in physical restraint use in response to health inspections. The low share of patients reported to be physically restrained (one percent) may limit the possibility of detecting significant changes.

22. The empirical specification is analogous to Equation (1) applied to the quantity and quality of labor inputs, admissions, and temporary discharges, except that an observation is at the case level rather than the facility level.

23. While these care outcomes are self-reported by nursing homes, our identification uses variation in these outcomes relative to inspections, and there could be limited incentives for nursing homes to change reporting in response to inspections since the MDS data are unavailable to inspectors at the time of inspection due to delays in data processing. For these patient care outcomes, we include those reported at regular assessments in the MDS and exclude those from admission and discharge assessments. The MDS reports federally mandated, standardized assessments of patients at admission, discharge, and regular intervals since admission (e.g., 30 days and 90 days). We measure influenza vaccinations at the current nursing home rather than at any facility because the data report the former. For antipsychotics, we measure their use in the prior seven days rather than at more granular levels (e.g., the current day) because the latter is not available in the data. To capture responses to health inspections, we define antipsychotic use in the prior seven days on event day $k$ as our outcome for antipsychotic use on day $k - 7$. This approach may shift the observed timing of changes in antipsychotic use due to health inspections earlier, whereas without it, those changes would be observed later.

24. To the extent that the assessment reports whether the patient has received the vaccinations, vaccinations may occur prior to the index assessment date, and responses to health inspections may take place even earlier than those observed in Panels G–H of Figure 3.

**Preinspection Responses.** An interesting question is whether the strategic responses described above emerge before an inspection begins. By regulation, inspections are unannounced (CMS 2023), yet anecdotes suggest that inspection dates may not be strictly hidden from nursing homes (see details in Section 2.2). The lack of reliable data on inspection start dates as described in Section 3.1 restricts the ability to rigorously evaluate whether the strategic responses emerge before an inspection begins. It could also be worth noting that even if some responses (e.g., staffing levels) emerge only when the inspection begins, it does not exclude preinspection responses. E.g., if on-duty staff schedules are set in advance (a frequent practice in health care settings), an increase in staff hours during inspections implies preinspection responses. Interestingly, as described above, the day of peak response for strategic changes that may be more efficient for nursing homes to implement before inspections (e.g., increasing temporary discharges and vaccinations) is earlier than that for improvements that could be more valuable when observed by inspectors (e.g., increased staffing levels), indicating potential preinspection responses.

**Summary.** Taken together, nursing homes show strategic responses to health inspections, in multiple core dimensions: the quantity of labor inputs, the quality of labor inputs, admissions, temporary discharges, and patient care in some dimensions. However, nearly all responses drop immediately once the inspection is completed. These results show that health inspections can overrate nursing home quality of care and underdetect quality deficiencies compared to their typical levels in the absence of inspections. It is also worth noting that our results do not necessarily imply that health inspections are inefficient: while reduced admissions and increased temporary discharges may be socially inefficient, the increased quantity and quality of labor inputs and improved patient care around the inspection may benefit patients exposed to them.

## 4.3 Correlates of Nursing Homes' Responses to Health Inspections

In this section, we explore correlates of nursing homes' responses to health inspections to shed light on mechanisms behind the strategic reactions. We first estimate individual nursing homes' responses to health inspections. We then correlate each nursing home's response with a range of observable nursing home and area characteristics. Appendix A.1 provides details of this analysis. For the quantity of labor inputs, rather than using various individual measures, we summarize the work hours of all types of staff.

Figure 4 reports the findings. The figure shows that staff hour increases are larger among for-profit nursing homes, potentially driven by stronger financial incentives since inspection ratings can influence demand. We also find that the increases are more pronounced among nursing homes with lower ratings in prior cycles, suggesting stronger response incentives with poorer existing ratings, and that the increases are larger in more competitive areas, consistent with competitive pressure increasing provider incentives to invest in quality, or quality ratings. Additionally, staff hour increases are greater in areas with higher staff availability (the number of RNs per capita). This pattern suggests that the capacity to respond may also play a role in the strategic responses to inspections.

For the other outcomes examined, although there are fewer significant patterns, we find that the quality of labor inputs rises more among for-profit nursing homes and that admission declines are more pronounced among for-profit facilities and facilities subject to greater competition. Interestingly, temporary discharges rise more notably among hospital-affiliated nursing homes, potentially reflecting these facilities' capability to leverage hospital ties to facilitate temporary transfers.

Taken together, these findings suggest that both incentives and the capacity to respond may play a role in nursing homes' gaming responses to health inspections.

### 4.4 Robustness Checks

Appendix Figures A.9-A.13 report robustness checks. Appendix Figure A.9 shows that our estimates are robust to excluding inspections conducted in any week of the year. Specifically, for each outcome, we first run separate regressions that exclude inspections conducted in each week of the year using Equation (1). We then report the maximum, minimum, and mean of the estimates from all regressions, along with our main estimate that includes all inspections for comparison. In Appendix Figure A.10, we show that our estimates are stable when we exclude holidays from the analysis sample. Appendix Figure A.11 extends the non-omitted period, from 30 to 75 days before and after the inspection, and shows that the results are virtually unchanged. Appendix Figures A.12-A.13 show that there is little correlation between inspection timing and various nursing home characteristics, in line with non-systematic variation in inspection timing.

## 5 Inspection Rating and Nursing Home Relative Quality

The strategic responses suggest that quality ratings from health inspections are unlikely to reflect nursing homes' *absolute* quality compared to their typical levels, a finding highly relevant given the reliance of important nursing home regulations on absolute quality assessments, including quality standard enforcement and certification determining nursing homes' eligibility to admit Medicare and Medicaid patients. Next, we ask whether the rating is predictive of nursing homes' *relative* quality. On the one hand, facilities with lower quality may exhibit stronger strategic responses to health inspections to improve ratings, potentially leading to an insignificant or even negative relationship between rating and quality. On the other hand, despite greater incentives, low-quality facilities may lack the capacity to perform as well as their high-quality counterparts during inspections, and thus a positive relationship remains between rating and quality. In this section, we start by classifying nursing homes into two groups based on their quality ratings from health inspections and use the differential distance from a patient's zip code of residence to the nearest high-quality-rated nursing home (above the median level within states) versus the nearest low-quality-rated nursing home (below the median level within states) as an instrument for admission to high-quality-rated nursing homes.[25] We measure nursing home quality by patient mortality, an unambiguous measure of patient welfare. Within our sample of Medicare beneficiaries, we find that admission to high-quality-rated nursing homes predicts a 0.8 percentage point decline in 90-day mortality, equivalent to a 5.2 percent reduction compared to the mean. Next, we estimate individual nursing homes' effects on patient 90-day mortality using a control function approach following Einav, Finkelstein, and Mahoney (2025), finding that 10.3 percent of the variance in nursing home mortality effects is explained by quality ratings from health inspections.

## 5.1 Empirical Framework

**Estimation Specification.** Our empirical specification is a two-stage least squares (2SLS) model that takes the following form:

$$y_c = \pi_1 \text{high}_c + \mathbf{L}_c \lambda_1 + \mathbf{X}_c \beta_1 + \varepsilon_{1,c}, \tag{2}$$

$$\text{high}_c = \pi_2 Z_c + \mathbf{L}_c \lambda_2 + \mathbf{X}_c \beta_2 + \varepsilon_{2,c}, \tag{3}$$

where $c$ denotes a case, $y_c$ is the outcome of interest (90-day mortality), and $\text{high}_c$ indicates whether case $c$ is admitted to a high-quality-rated nursing home. We use $Z_c$ to denote the instrument: the differential distance from a patient's home zip code to the nearest high-quality-rated nursing home versus the nearest low-quality-rated nursing home. The vector $\mathbf{L}_c$ encodes interactions between indicators for the county and indicators for the year of the patient's admission, i.e., $\mathbf{L}_c$ includes county-by-year indicators that account for potential differences across county-year cells. The parameter of interest is $\pi_1$, representing a local average treatment effect (LATE), i.e., the average effect among cases induced by the differential distance into being admitted to a high-quality-rated nursing home.

As robustness checks, our specification also includes a vector of patient covariates $\mathbf{X}_c$, including indicators for five-year age bins, gender, black race, disability status, Medicaid coverage in the prior year (a proxy

25. In health inspection ratings, higher scores represent lower quality performance. We define nursing homes with an average score during our study period below the state median score as high-quality-rated and those with an average score above the state median score as low-quality-rated. We divide high- and low-quality-rated nursing homes within states rather than broader geographic units, as health inspections are conducted by state-level agencies, making cross-state comparisons less meaningful.

for income), and 27 chronic conditions reported in the Chronic Conditions file of Medicare claims at the start of the year of the nursing home admission.[26] For each patient covariate with missing values, we add an indicator for missing values and replace the missing values with zero. Finally, $\varepsilon_{1,c}$ and $\varepsilon_{2,c}$ are error terms. We cluster standard errors by patient zip code of residence. We also show in Section 5.3 that standard errors are stable when clustered at a more aggregate level.

**Analysis sample.** We restrict the sample to patients' first nursing home stay during our study period 2014-2019 (we further ensure that these observations have no prior stay in at least one year before our study period). Since nursing home residents often transfer between nursing homes and other facilities (e.g., hospitals), restricting attention to patients' first nursing home stay allows us to focus on a relatively homogeneous sample and reduce the influence of other facilities. We also restrict the sample to patients between 66 and 100 years old. (We exclude patients aged 65, to allow for a look-back period.) Appendix Table A.1 summarizes the characteristics of our sample. Column 1 includes all analyzed patients. Columns 2 and 3 provide summary statistics separately for patients admitted to nursing homes with high and low quality ratings. Along multiple dimensions, patients admitted to high- and low-quality-rated nursing homes are systematically different, suggesting that OLS comparison can be biased by patient selection.

**Identification Assumptions.** Below we discuss empirical evidence supporting the validity of the identification assumptions for our IV design: relevance, conditional independence, exclusion restriction, and monotonicity.

*Relevance.* Panel A of Figure 5 presents the first stage of our IV model, conditional on the baseline controls $\mathbf{L}_c$.[27] The figure shows that the patient's probability of being admitted to high-quality-rated nursing homes declines with the instrument: differential distance to high- versus low-quality-rated nursing homes. The first-stage relationship is highly significant, with an $F$-statistic above 10,000. The strong first-stage effect is consistent with earlier studies documenting a significant role of differential distance in affecting patients' choice of health care providers, including the choice of nursing homes (e.g., McClellan, McNeil, and Newhouse 1994; Card, Fenizia, and Silver 2023; Gupta et al. 2024; Einav, Finkelstein, and Mahoney 2025).

*Conditional Independence.* For our instrument to be valid, the differential distance must be uncorrelated

26. We measure Medicaid coverage in the prior year rather than the current year because nursing home stays may affect Medicaid coverage: while less than one-fifth of nursing home residents are covered by Medicaid at the time of admission, about two-thirds of them are covered by Medicaid at discharge (see, e.g., Hackmann 2019). The 27 chronic conditions include acute myocardial infarction, Alzheimer's disease, Alzheimer's disease and related disorders or senile dementia, atrial fibrillation, cataract, chronic kidney disease, chronic obstructive pulmonary disease, heart failure, diabetes, glaucoma, hip/pelvic fracture, ischemic heart disease, depression, osteoporosis, rheumatoid arthritis/osteoarthritis, stroke/transient ischemic attack, breast cancer, colorectal cancer, prostate cancer, lung cancer, endometrial cancer, anemia, asthma, hyperlipidemia, benign prostatic hyperplasia, hypertension, and acquired hypothyroidism (Research Data Assistance Center (ResDAC) 2024).

27. The share of patients admitted to high-quality-rated nursing homes is higher than 50 percent, reflecting that high-quality-rated nursing homes on average have larger patient volumes than low-quality-rated nursing homes.

with patient potential outcomes, conditional on the baseline controls $\mathbf{L}_c$. We show several pieces of evidence supporting conditional independence. First, Panel B of Figure 5 shows patient predicted 90-day mortality, a composite index of patient pre-determined characteristics $\mathbf{X}_c$ that are highly predictive of 90-day mortality (joint $F$-statistic: 3,592), is well balanced across the values of our instrument conditional on $\mathbf{L}_c$.[28]

Second, we show in the empirical results below the robustness of our IV estimate to including various different combinations of patient controls. Specifically, we divide observable patient characteristics into six groups and estimate separate IV models that control for each of the $2^6 = 64$ different combinations of patient covariates. As presented in Section 5.3 below, our IV estimate is stable despite any combination of patient covariates. Following the reasoning of Altonji, Elder, and Taber (2005), this evidence indicates limited selection bias due to either observed or unobserved patient characteristics that predict patient outcomes.

Third, we further apply two methods to account for possible selection on patient unobservable characteristics: (i) we bound our estimates by allowing for selection on patient unobservables using an approach by Oster (2019) (details in Appendix A.2); and (ii) following Altonji and Mansfield (2018), we control for average characteristics of other patients living in the index patient's zip code of residence, which can absorb selection beyond the observable characteristics of the index patient. Section 5.3 below shows that our estimate is stable. Together, empirical evidence suggests that our IV estimate is unlikely to be driven by selection in either observable or unobservable patient characteristics.

*Exclusion Restriction.* Rather than the effect of rating on patient mortality, our study question is whether rating *predicts* mortality outcomes, after accounting for patient selection bias. While there can exist other differences between high- and low-quality-rated nursing homes correlated with rating that influence patient mortality, e.g., facility management quality, they can be viewed as mechanisms driving the relationship between nursing home quality rating and patient mortality.

*Monotonicity.* In the presence of heterogeneous treatment effects, monotonicity must be assumed to interpret the IV estimate as a LATE, i.e., the average effect among cases on the margin of being admitted to high- versus low-quality-rated nursing homes. In our setting, monotonicity requires that cases admitted to high-quality-rated nursing homes when the differential distance is high would also be admitted to high-quality-rated nursing homes when the differential distance is low, and vice versa.

Following the literature, we examine a testable implication of the monotonicity assumption: the first-stage estimate should be consistent across subsamples defined by patient characteristics (e.g., Dahl, Kostøl, and Mogstad 2014; Bhuller et al. 2020). Appendix Table A.2 shows the test. Specifically, we divide the sample by patient characteristics including age, gender, race, Medicaid coverage in the prior year, disability status, and

28. Predicted 90-day mortality is generated from a linear regression of actual 90-day mortality on $\mathbf{X}_c$.

predicted 90-day mortality and examine the first-stage effect separately for each subsample. Appendix Table A.2 shows that for all subsamples, the first-stage estimates are negative and highly significant, supporting the monotonicity assumption.

### 5.2 Results

Panel B of Figure 5 shows the reduced-form estimates: the relationship between patient 90-day mortality and the instrument, conditional on $\mathbf{L}_c$. Patient 90-day mortality increases significantly with the instrument, whereas predicted 90-day mortality is well balanced across the instrument.

Table 2 reports the reduced-form and IV estimates for 90-day mortality, both of which are stable when we add various patient controls. The IV estimate with the full set of controls (Panel B, Column 4) shows that, relative to admission to low-quality-rated nursing homes, admission to high-quality-rated nursing homes on average predicts a 0.8 percentage point decline in 90-day mortality, equivalent to a 5.2 percent decrease relative to the mean. This mortality reduction is notable, approximately equal to three-quarters of the returns to Medicare eligibility, which lowers hospitalized patients' 28-day mortality by 1.1 percentage points (Card, Dobkin, and Maestas 2009).[29]

Figure 6 presents IV estimates for mortality measured at different horizons, ranging from 30 days to one year since admission in monthly intervals. The estimations follow the model in Equations (2) and (3). The figure shows a consistent pattern of mortality declines for follow-up periods up to one year: the estimates increase in magnitude over the first 180 days and are stable thereafter. The persistence of mortality declines over longer horizons indicates that high-quality-rated nursing homes prevent rather than merely temporarily delay the deaths of fragile patients.

In Appendix Figure A.14, we examine heterogeneity in the relationship between nursing home quality ratings and patient 90-day mortality by nursing homes' responses to health inspections in total staff hours (a highly salient response as shown in Section 4). We use an IV model analogous to that in Equations (2) and (3).[30] Among facilities with staff hour responses in the bottom three quartiles, admission to a high-

29. In Appendix Table A.3, we characterize compliers following the approach developed by Abadie (2003) (details provided in Appendix A.3). The table shows that relative to the average patient, compliers are slightly older (82.3 versus 81.0 years), likelier to be female (61.1 versus 60.6 percentage points), less likely to be black (3.5 versus 9.4 percentage points), and less likely to be disabled (12.5 versus 13.7 percentage points).

30. Specifically, the empirical specification takes the form $y_c = \sum_{g=1}^{G} \mathbf{1}(\text{Group}_c = g) \left[\pi_{1g}\text{high}_c + \omega_{1g}\right] + \mathbf{L}_c\lambda_1 + \mathbf{X}_c\beta_1 + \varepsilon_{1,c}$, where $\mathbf{1}(\text{Group}_c = g)$ is an indicator for whether the nursing homes' responses in total staff hours to health inspections is in the top quartile or the bottom three quartiles. As a natural extension of our main IV model, we instrument for the interactions between $\left\{\mathbf{1}(\text{Group}_c = g)\right\}_{g=1}^{G}$ and $\text{high}_c$ by interacting $\left\{\mathbf{1}(\text{Group}_c = g)\right\}_{g=1}^{G}$ with $Z_c$, where $Z_c$ is the differential distance from a patient's zip code of residence to the nearest high-quality-rated nursing home versus the nearest low-quality-rated nursing home. As described in Appendix A.1, to limit estimation noise, in estimating individual nursing homes' responses we restrict the sample to nursing homes with at least three health inspections during our study period. Thus, the sample for this heterogeneity analysis is smaller than that for the main IV estimation, and the weighted average of the IV estimates in Appendix Figure A.14 is below the main IV

quality-rated nursing home predicts a statistically significant decline in patient 90-day mortality. In contrast, the analogous relationship is smaller in magnitude and statistically insignificant among facilities in the top quartile of responses. These results suggest that inspection ratings are more effective at predicting quality when facilities exhibit weaker strategic responses to inspections.

### 5.3 Robustness

Appendix Figure A.15 shows the robustness of our IV estimate to the inclusion of different combinations of patient controls. We first divide patient covariates into six groups: (i) five-year age-bin indicators; (ii) gender; (iii) black race; (iv) Medicaid coverage in the prior year; (v) disability status; and (vi) indicators for 27 chronic conditions described in Section 5.1. Next, we run separate IV regressions controlling for each of the $2^6 = 64$ different combinations of patient covariates. For each $n = 0, 1, ..., 6$, we report in Appendix Figure A.15 the mean, maximum, and minimum of the IV estimates using all possible combinations with $n$ subsets of patient covariates. For reference, we also plot the 95 percent confidence interval of the IV estimate with only the baseline controls $\mathbf{L}_c$. The IV estimates are not statistically different despite controlling for any combination of patient covariates, despite these covariates being strong predictors of patient 90-day mortality (joint $F$-statistic of 3,592).

In Appendix Table A.4, we apply two methods to further account for possible selection on patient unobservables: (i) we apply an approach by Oster (2019), bounding our estimate to account for possible selection on patient unobservables (details in Appendix A.2); and (ii) we apply the method by Altonji and Mansfield (2018) and control for patient unobservables by controlling for average characteristics of other patients living in the index patient's zip code of residence.[31] We show that our estimates are stable.

Appendix Table A.5 reports additional robustness checks. Column 2 excludes patients whose stays overlap with the two weeks before and the two weeks after the last day of a health inspection, to limit the possibility that the estimated relationship between quality ratings and patient survival is driven by nursing homes' strategic responses during inspections. We show that the estimates are robust. In Column 3, we construct the instrument using the patient's zip code of residence in the prior year, in case the current-year zip code reflects the nursing home's location. The results are stable. Indeed, this concern is likely minimal, since our sample includes only patients' first nursing home stays, restricting the possibility that patients'

estimate reported in Table 2.

31. These characteristics include average demographics and comorbidities that include five-year bin indicators for mean age, the shares of patients who are female, black, disabled, and covered by Medicaid in the prior year, and the shares of patients with each of the 27 chronic conditions described in Section 5.1 obtained using other patients living in the index patient's zip code of residence, as well as average socioeconomic characteristics that include per capita income, the share of individuals with a college degree, and unemployment rates at the zip code level obtained from the 2015-2019 American Community Survey.

residence reflects the nursing home's location. Further, 95 percent of patients in our sample have the same zip code in the current and prior year, equal to the share among Medicare beneficiaries with no nursing home stays. Column 4 shows that our estimates are robust when we cluster standard errors at a more aggregate level: county.

### 5.4 Individual Nursing Homes' Mortality Effects and Quality Ratings

We further estimate individual nursing homes' effects on patient 90-day mortality and correlate them with quality ratings from health inspections. Specifically, we first estimate a nursing home survival value-added model using a control function approach that leverages patients' distance to each nursing home in the choice set as excluded instruments to account for patient sorting into nursing homes, as in Einav, Finkelstein, and Mahoney (2025). We then correlate individual nursing homes' effects on patient 90-day mortality from the control function approach with their quality ratings from health inspections. Appendix A.4 provides details of this analysis.

Figure 7 shows the results. We find a highly statistically significant decline in mortality effects with higher quality ratings, consistently suggesting that inspection ratings on average are predictive of nursing home quality. There is also notable dispersion around the regression line. Using a split-sample approach to account for estimation noise, we find that 10.3 percent of the variance in nursing home mortality effects is explained by quality ratings (details in Appendix A.5).

Our finding of a positive relationship between nursing home quality ratings from health inspections and patient survival is consistent with Einav, Finkelstein, and Mahoney (2025), which finds a positive association between nursing homes' five-star ratings (including inspection-based five-star ratings) and their value added in improving patient health status, although there is sizable dispersion around the regression line. Our finding also relates to Olenski and Sacher (2024), which reports a near-zero correlation between nursing homes' inspection-based five-star ratings and their effects on patient survival. Complementing these studies, we examine the relationship between direct ratings from health inspections and nursing homes' effects on patient survival. Notably, inspection-based five-star ratings include ratings from both health inspections—on which we focus—and complaint investigations. Different from health inspections, which are conducted on nearly all nursing homes with a relatively standardized scope, complaint investigations are triggered by patient allegations and vary in scope depending on the complaints. Both the occurrence and the content of complaint investigations can vary systematically across nursing homes given patient sorting into facilities. This can lead to differences in how strongly health inspection ratings and inspection-based five-star ratings predict nursing home quality. This difference is policy-relevant: it highlights the caveats of including nonstandardized evaluation content in five-star ratings, a key tool used by CMS to guide patients and their families in choosing nursing homes.

## 6 Effects of Quality Deficiency Citations

The results thus far show the role of health inspections in capturing nursing home quality of care. In this section, we examine the role of health inspections in incentivizing nursing home quality improvements. We focus on quality deficiency citations, the key component of health inspections designed to enhance nursing home quality of care. The citations are issued to facilities found in violation of quality standards by health inspections and can lead to financial penalties.[32] The potentially tougher reputation of regulators and the salience of penalties associated with citations may incentivize nursing homes to improve care quality. The citations may also reveal previously unknown care deficiencies, prompting facilities to address them. On the other hand, policies may not always achieve intended effects. We study the effects of 14 key citations with related outcomes measurable in our data. Using an event-study design that compares outcome changes at the cited nursing homes following a citation with changes at nursing homes that never received the citation during our study period or in adjacent years, we find mixed results. A number of citations lead to significant improvements in the related quality outcomes, whereas citations that appear to have less precise evaluation criteria and potentially lower financial penalties have limited effects. We also examine whether citations have spillover effects on non-cited quality dimensions and nearby facilities, finding limited evidence.

### 6.1 Empirical Framework

We examine the effects of 14 key citations with related outcomes measurable in our data. While there are various citations,[33] a large number of citations do not have related outcomes trackable in our data, which include the PBJ, MDS, and Medicare claims—an extensive set of currently available data on nursing homes.[34] Table 3 lists the description of each citation provided by CMS and the related outcomes we evaluate. For brevity, we assign a short name to each citation. Since CMS modified citation descriptions during our study period, we present both versions of descriptions in Table 3. Appendix Figure A.16 shows that a fair share of nursing homes receive each studied citation.

32. Section 2 describes quality deficiency citations in more detail.
33. See https://data.cms.gov/provider-data/dataset/r5ix-sfxw for the list of citations.
34. For example, we are unable to observe outcomes for citations such as "maintain comfortable sound levels" and "provide clean bed and bath linens that are in good condition," and outcomes for citations such as "provide necessary care and services to maintain or improve the highest well-being of each resident" seem vague.

Our empirical specification is an event-study design:

$$y_{it} = \sum_{\tau} \delta_\tau \text{Citation}_{it}^{\tau} + u_i + \eta_t + \mathbf{X}_{it}\psi + \varepsilon_{it}, \tag{4}$$

where $i$ indexes nursing homes, $t$ indexes year-quarters, and $\text{Citation}_{it}^{\tau}$ denotes indicators for $\tau$ quarters relative to the quarter the nursing home received the citation. The vectors $u_i$ and $\eta_t$ are nursing home and year-quarter fixed effects, respectively. In our baseline specification, $\mathbf{X}_{it}$ is a dummy for whether the nursing home has a health inspection in the quarter, to capture potential responses to health inspections shown in Section 4. In Section 6.4 below, we also show the robustness of our estimates to excluding days around health inspections and to adding various additional controls to $\mathbf{X}_{it}$. Finally, $\varepsilon_{it}$ is the error term. We cluster standard errors by nursing home.

The coefficients of interest are $\delta_\tau$'s. While we include all event quarters in the regression, we report $\delta_\tau$'s for $\tau$=[-4, 8], as $\delta_\tau$'s become noisier and less informative when $\tau$ is further from zero due to declined sample sizes. Along with a two-way fixed effects estimation (TWFE), we implement various estimators to correct for potential bias in staggered treatment timing in the presence of treatment effect heterogeneity, including De Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Borusyak, Jaravel, and Spiess (2024).

We obtain citations issued between 2015 and 2021. To limit the possibility that control nursing homes are affected by citations, we focus our analysis on the three-year period from 2017 to 2019 and define control nursing homes as those that never received a citation during 2015-2021. We define the treatment group as nursing homes receiving the citation during 2017-2019 but never in at least the prior two years.[35]

## 6.2 Main Results

Figure 8 reports the estimated $\delta_\tau$'s, along with the outcome means for treated nursing homes during the precitation period and for control nursing homes over the full analysis period.[36] While a direct comparison of outcome means between treatment and control nursing homes is not meaningful for outcomes such as work hours of nurses since, e.g., treatment and control facilities can differ in size, Figure 8 suggests that nursing homes receiving citations appear to perform worse than those that do not in outcomes that may be more comparable across facilities, such as whether having RNs at least eight hours per day (a minimum

35. For nursing homes receiving the same citation more than once during 2017-2019, we define the treatment date as the date of the first citation. We show below in Section 6.4 that excluding nursing homes with repeated citations does not meaningfully affect our estimates.

36. We report the sample means for the treated group in the precitation period to minimize the effects of citations.

staffing requirement) and pneumococcal and influenza vaccination rates.[37]

Figure 8 shows that a number of citations lead to improvements in the examined outcomes. Reassuringly, in most instances, there is no indication of diverging trends between treatment and control facilities prior to the citation. In the instances where there appear to be diverging trends, the pre-trends seem unlikely to qualitatively affect our conclusions. For example, though Panels A and N of Figure 8 suggest pre-trends, the magnitudes are relatively small, and accounting for them may increase the estimated treatment effect but is unlikely to qualitatively change our conclusions. Panel A of Figure 8 shows that a citation for violating the standard requiring RNs to be on duty for at least eight hours per day increases the likelihood that nursing homes comply with the standard by an average of more than five percentage points in the post-period, an over 40 percent decrease in the likelihood of violation relative to the treatment group's precitation mean. The effect persists over time, lasting for at least eight quarters postcitation.[38] We find similar patterns for alternative citations. Panel C of Figure 8 shows that receiving a citation for not hiring a qualified full-time social worker in facilities with more than 120 beds raises these nursing homes' propensity to employ a qualified full-time social worker. Figure 8 also shows that receiving a citation for underperformance in developing and implementing policies and procedures for influenza and pneumococcal vaccinations increases the share of residents receiving the two vaccines, and receiving citations for not adequately updating residents' assessments at least once every three months and not encoding/transmitting assessment data to states within seven days improve nursing homes' performance in the corresponding outcomes. Appendix Table A.6 reports the simple difference-in-differences estimates for the citations in Figure 8.

An important question is whether the improvements associated with citations reflect sustained changes or merely temporary improvements around subsequent health inspections to avoid being cited again. While Section 4 shows strategic responses to health inspections, the share of days in a given quarter impacted by health inspections is low. Taking the average length between adjacent health inspections (390 days) and assuming that the number of days with strategic improvements around a health inspection is 14 days, there are fewer than four percent ($\frac{14}{90} \times \frac{90}{390}$) of days in a quarter affected by strategic responses to health inspections. Further, in Section 6.4 below, we show that our estimates are similar when we exclude days around health inspections (from two weeks before to two weeks after the inspection), suggesting that the improvements reported above are not merely driven by temporary changes around subsequent health inspections.

37. A related question is that, since Figure 8 reports the outcome means for treated and control nursing homes in the analysis period prior to the citation and during our entire analysis period, respectively, whether the outcome mean differences between treated and control nursing homes may be driven by time trends. To examine this possibility, we estimate the outcome mean differences adjusting for quarter-year fixed effects, finding qualitatively similar results.

38. The exact descriptions of citations changed with CMS's modification as shown in Table 3. For simplicity, in the main text we describe citations based on the earlier version.

It is also notable that some citations do not appear to affect the examined outcomes. For example, we find no evidence that citations for not having sufficient nurse staffing (i.e., the nurse citations in Table 3) impact nurse hours. Similarly, we do not find that citations for non-medically necessary use of physical or chemical restraints (i.e., the physical and chemical restraint citations) affect the likelihood that nursing home residents receive these restraints. The citations related to pharmacist tasks (the pharmacist and medication review citations) also do not seem to influence pharmacists' hours.

We next explore differences between citations that lead to improvements and those that have limited effects. An exhaustive examination is difficult due to data availability, but we probe two areas: evaluation criteria and the amount of financial penalties. Based on the CMS descriptions in Table 3, citations that lead to improvements seem to have more precise evaluation criteria. For example, we observe significant improvements in response to the RN citation "use a registered nurse at least 8 hours a day, 7 days a week" but not the nurse citation "have enough nurses to care for every resident in a way that maximizes the resident's well being," despite the examined outcomes of the two citations being highly correlated as RNs are a subtype of nurses. The RN citation seems precisely defined, while the nurse citation seems vague as "enough" could be subject to interpretation. For other citations with clear evaluation criteria, e.g., "encode each resident's assessment data and transmit these data to the State within 7 days of assessment," "hire a qualified full-time social worker in a facility with more than 120 beds," and "develop policies and procedures for influenza and pneumococcal immunizations," we also observe significant improvements in the examined outcomes after the citations. With clear evaluation criteria, current citations may strongly predict future citations if the nursing home does not improve the cited deficiency, potentially providing stronger incentives for nursing homes to invest in improvements. It is also possible that clear evaluation criteria render the examined outcomes less fungible, thereby increasing the likelihood of detecting an effect.[39]

Financial penalties may also be associated with whether a citation leads to improvements. Health inspections categorize the scope of each cited deficiency based on the pervasiveness of the deficiency in the facility—including isolated, a pattern, or widespread—as well as the severity level based on the impact the deficiency has on patients' safety (CMS 2023).[40] A broader scope or higher severity level indicates greater

39. In addition to the examined outcomes, nursing homes may respond to some citations for which we find limited effects by reducing patient volume or altering patient composition. For example, in response to nurse or dietary support staff citations, facilities may lower patient volume to maintain adequate staffing levels without increasing staff hours; for rehabilitative service citations, facilities may limit admissions of patients requiring rehabilitative services. In Section 6.4, we show that the estimated effects of these citations remain limited when we control for patient census and patient characteristics.

40. Specifically, as described in CMS (2023), scope is isolated when "one or a very limited number of residents or employees is/are affected and/or a very limited area or number of locations within the facility are affected," a pattern when "more than a very limited number of residents or employees are affected, and/or the situation has occurred in more than a limited number of locations but the locations are not dispersed throughout the facility," and widespread when "the problems causing the deficiency are pervasive (affect many locations) throughout the facility and/or represent a systemic failure that affected, or has the potential to affect, a large portion

financial penalties. Although the data do not report fines for individual citations, it documents the scope and severity of each citation separately. Appendix Figure A.17 displays the distribution of scopes and severity levels for each examined type of citations. For RN and social worker citations which have a high share of "widespread" scopes, we find significant effects in Figure 8. For pressure ulcer citations which have a high average severity level, Figure 8 also suggests reducing effects. (The positive coefficients for event quarter zero for pressure ulcers in Panel H of Figure 8 may reflect high pressure ulcer rates in the quarter that lead to the citation. The panel suggests that pressure ulcers decline from event quarter one relative to event quarter zero.) Interestingly, the evaluation criteria for pressure ulcer citations seem relatively vague in Table 3. Their potentially high financial penalties as indicated in Panel B of Appendix Figure A.17 may contribute to the effects. Overall, there appears to be a pattern that citations associated with relatively precise evaluation criteria or relatively large financial penalties are more likely to lead to significant improvements, shedding light on potential avenues to enhance the effectiveness of citations.

### 6.3 Spillover Effects on Other Quality Dimensions and Nearby Facilities

Given the significant effects of some citations presented above, we ask in this section whether citations have spillover effects on non-cited quality dimensions and nearby facilities. We begin by examining whether receiving a citation in one dimension affects the nursing home's performance in non-cited dimensions. Such within-facility spillovers may exist and be positive if citations induce general investments in care quality in the facility. The spillovers may also be negative if nursing homes divert resources from the non-cited to the cited dimensions or may not exist if investments in different quality dimensions are independent. Appendix Figure A.18 displays the estimated effects of citations on non-cited outcomes, using the event-study design described in Section 6.1. In each panel, we exclude nursing homes receiving a citation relevant to the examined outcome to restrict confounding effects. Appendix Figure A.18 shows limited within-facility spillover effects from cited to nondirectly cited quality metrics for most citations. Two main exceptions are that citations for underperformance in updating residents' assessments every three months lead to improved performance in transmitting assessment data to states within seven days of assessments, and citations for underperformance in the latter metric lead to small improvements in the former.[41]

Next, we examine whether citations at a nursing home generate spillover effects on nearby facilities.

or all of the residents or employees." For severity, there are four levels: level 1 representing "no actual harm with potential for minimal harm," level 2 representing "no actual harm with a potential for more than minimal harm that is not immediate jeopardy," level 3 representing "actual harm that is not immediate jeopardy," and level 4 representing "immediate jeopardy to resident health or safety."

41. Since Figure 8 shows that results are stable when using different estimators to account for potential bias due to staggered treatment timing, in analyses in this subsection and robustness checks in Section 6.4 below, we present results using the estimator by Callaway and Sant'Anna (2021).

Geographic proximity may facilitate interactions, enabling knowledge spillovers from a cited facility that affect nearby facilities' beliefs about the probability and severity of citations. Not-yet-cited facilities may invest in quality compliance to avoid future citations even though they are not cited. To explore geographical spillovers, we define non-cited nursing homes within five miles of the cited facilities in our main estimation sample in Section 6.2 as the treatment group and the remaining non-cited facilities located at least five miles from any facilities cited between 2015 and 2021 as the control group. Results are qualitatively similar when we use a lower cutoff of three miles or a higher cutoff of ten miles. Nursing homes receiving the citation are excluded from both the treatment and control groups in this exercise. We estimate the effects using the event-study design described in Section 6.1. Appendix Figure A.19 suggests limited spillover effects from cited to nearby non-cited nursing homes. Put differently, citations do not appear to incentivize quality improvements beyond those in the cited facilities.

### 6.4 Robustness Checks

Appendix Figures A.20-A.21 and Appendix Table A.6 report robustness checks. Appendix Figure A.20 presents three robustness checks. First, we exclude days around health inspections (from two weeks before to two weeks after the last day of each inspection) to limit the effect of responses to inspections. We show that our estimates are stable. Second, we show that our estimates are robust to excluding from the control group nursing homes close to the cited facilities (within five miles), to limit potential confounding from geographic spillovers. Third, we show that our estimates are similar when we exclude nursing homes receiving the same citation again during our analysis period. Appendix Figure A.21 shows that our estimates are virtually unchanged when we add various nursing home time-varying characteristics and patient characteristics as controls.[42] Appendix Table A.6 examines whether the estimated significant effects of citations in Figure 8 may be driven by mean reversion, using the synthetic difference-in-differences estimation (Arkhangelsky et al. 2021). The results are stable, suggesting that mean reversion is unlikely to explain the findings.

## 7 Conclusion

In contracting out, monitoring is an important policy tool to extract firms' quality information and incentivize firms to provide quality. Yet high operating costs and fiscal constraints often lead monitoring to be conducted on an intermittent rather than a continuous basis. Firms' potential temporary efforts in response to inspections

42. The nursing home time-varying characteristics include for-profit status, hospital affiliation, total number of beds, and patient census. For patient covariates, we include age, number of chronic health conditions, gender, black race, disability status, and Medicaid coverage in the prior year, collapsed at the nursing-home-by-quarter level (the unit of observation in Equation (4)).

raise the question of whether government monitoring can successfully capture and incentivize quality or mainly lead to wasteful regulatory costs.

In this paper, we examine government monitoring of nursing homes' quality of care. Nursing homes are a sector with significant welfare implications, yet there are widespread concerns about their quality of care. To monitor and incentivize care quality, CMS mandates health inspections of nursing homes. We find that nursing homes show strong strategic responses to health inspections. Nursing homes increase the quantity and quality of labor inputs, reduce admissions, increase temporary discharges, and improve patient care (e.g., increase influenza and pneumococcal vaccinations) in response to the inspection. However, nearly all the examined responses drop immediately once the inspection is completed. We also find heterogeneity in the responses across nursing homes that suggests both incentives and the capacity to respond may drive the gaming behavior. The strategic responses suggest that inspection ratings are unlikely to reflect firms' *absolute* quality compared to their typical levels. On the other hand, we find that inspection ratings on average reveal firms' *relative* quality. We find a positive relationship between nursing homes' quality ratings from health inspections and quality as measured by patient survival. Finally, we find effects of some quality deficiency citations on incentivizing sustained improvements in the cited quality metrics. There appears to be a pattern in which citations with more precise evaluation criteria and higher financial penalties are likelier to incentivize quality improvements, shedding light on potential avenues to enhance the effectiveness of quality deficiency citations.

Our findings highlight the complex interplay between regulatory monitoring and firm behavior. We find that firms' gaming of government monitoring can and does occur. The strategic responses show that inspections can overrate firms' quality performance and underdetect quality deficiencies relative to their typical levels. This finding is highly relevant given the reliance of common forms of regulation on firms' absolute quality assessments, such as certification and enforcing compliance with quality standards. In addition, there may be unintended consequences of inspections (e.g., reduced admissions and increased potentially unnecessary hospital transfers in our setting). Firms' gaming responses should be taken into account when designing regulatory inspection policies. On the other hand, inspection ratings on average reveal firms' relative quality (quality ranking), and quality deficiency citations can provide a means to incentivize firms to improve quality when properly designed. Although a full analysis of the efficiency of regulatory inspections is beyond the scope of this article, our findings provide important inputs to this calculation.

## References

**Abadie, Alberto.** 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–263.

**Almunia, Miguel, and David Lopez-Rodriguez.** 2018. "Under the Radar: The Effects of Monitoring Firms on Tax Compliance." *American Economic Journal: Economic Policy* 10 (1): 1–38.

**Altonji, Joseph, Todd Elder, and Christopher Taber.** 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–184.

**Altonji, Joseph, and Richard Mansfield.** 2018. "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects." *American Economic Review* 108 (10): 2902–2946.

**Arkhangelsky, Dmitry, Susan Athey, David Hirshberg, Guido Imbens, and Stefan Wager.** 2021. "Synthetic Difference-in-differences." *American Economic Review* 111 (12): 4088–4118.

**Bhuller, Manudeep, Gordon Dahl, Katrine Løken, and Magne Mogstad.** 2020. "Incarceration, Recidivism, and Employment." *Journal of Political Economy* 128 (4): 1269–1324.

**Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen.** 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." *Review of Economic Studies* 82 (2): 457–489.

**Borusyak, Kirill, Xavier Jaravel, and Jann Spiess.** 2024. "Revisiting Event-study Designs: Robust and Efficient Estimation." *Review of Economic Studies* 91 (6): 3253–3285.

**Boudreau, Laura.** 2024. "Multinational Enforcement of Labor Law: Experimental Evidence on Strengthening Occupational Safety and Health Committees." *Econometrica* 92 (4): 1269–1308.

**Branco, Danyelle, Bladimir Carrillo, Di Fang, and Wilman Iglesias.** 2023. "Expertise Overlap and Team Productivity: Evidence from the Hospital Industry." Accessed May 31, 2025. https://sbe.org.br/anais/45EBE/microeconomia/45_EBE_paper_95.pdf.

**Bureau of Labor Statistics.** 2024a. "Licensed Practical and Licensed Vocational Nurses." Accessed August 29, 2024. https://www.bls.gov/ooh/healthcare/licensed-practical-and-licensed-vocational-nurses.htm.

———. 2024b. "Nursing Assistants." Accessed August 29, 2024. https://www.bls.gov/oes/2023/may/oes311131.htm.

———. 2024c. "Registered Nurses." Accessed August 29, 2024. https://www.bls.gov/ooh/healthcare/registered-nurses.htm.

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

**Card, David, Carlos Dobkin, and Nicole Maestas.** 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics* 124 (2): 597–636.

**Card, David, Alessandra Fenizia, and David Silver.** 2023. "The Health Impacts of Hospital Delivery Practices." *American Economic Journal: Economic Policy* 15 (2): 42–81.

**Celhay, Pablo, Paul Gertler, Paula Giovagnoli, and Christel Vermeersch.** 2019. "Long-run Effects of Temporary Incentives on Medical Care Productivity." *American Economic Journal: Applied Economics* 11 (3): 92–127.

**Centers for Medicare & Medicaid Services.** 2018a. "Design for Care Compare Nursing Home Five-star Quality Rating System: Technical Users' Guide." Accessed September 16, 2024. https://www.cms.gov/files/document/fivestarusersguidejuly2018pdf.

———. 2018b. "Transition to Payroll-Based Journal (PBJ) Staffing Measures on the Nursing Home Compare Tool on Medicare.gov and the Five Star Quality Rating System." Accessed December 29, 2024. https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/SurveyCertificationGenInfo/Downloads/QSO18-17-NH.pdf.

———. 2023. "Survey and Enforcement Process for Skilled Nursing Facilities and Nursing Facilities." Accessed August 29, 2024. https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/som107c07pdf.pdf.

———. 2024. "NHE Tables." Accessed March 17, 2025. https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/nhe-fact-sheet.

**Centers for Medicare & Medicaid Services.** 2025. "Civil Money Penalty (CMP) Report." Accessed Feburary 6, 2025. https://qcor.cms.gov/nh_wizard.jsp?which=0&report=enf_cmp.jsp.

**Chan, David, Matthew Gentzkow, and Chuan Yu.** 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *Quarterly Journal of Economics* 137 (2): 729–783.

**Chen, Yiqun.** 2021. "Team-specific Human Capital and Team Performance: Evidence from Doctors." *American Economic Review* 111 (12): 3923–3962.

**Chidambaram, Priya, and Alice Burns.** 2024. "A Look at Nursing Facility Characteristics between 2015 and 2024." *KFF.* Accessed December 14, 2024. https://www.kff.org/medicaid/issue-brief/a-look-at-nursing-facility-characteristics/.

**Congressional Budget Office.** 2013. "Rising Demand for Long-term Services and Supports for Elderly People." Accessed January 6, 2025. https://www.cbo.gov/sites/default/files/113th-congress-2013-2014/reports/44363-ltc.pdf.

**Currie, Janet, and W. Bentley MacLeod.** 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics* 35 (1): 1–43.

**Cutler, David.** 1995. "The Incidence of Adverse Medical Outcomes under Prospective Payment." *Econometrica* 63 (1): 29–50.

**Dahl, Gordon, Andreas Ravndal Kostøl, and Magne Mogstad.** 2014. "Family Welfare Cultures." *Quarterly Journal of Economics* 129 (4): 1711–1752.

**De Chaisemartin, Clément, and Xavier d'Haultfoeuille.** 2020. "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Eeview* 110 (9): 2964–2996.

**Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite.** 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *Journal of Political Economy* 111 (3): 555–588.

**Einav, Liran, Amy Finkelstein, and Neale Mahoney.** 2025. "Producing Health: Measuring Value Added of Nursing Homes." *Econometrica* 93 (4): 1225–1264.

**Epstein, Andrew, Jonathan Ketcham, and Sean Nicholson.** 2010. "Specialization and Matching in Professional Services Firms." *RAND Journal of Economics* 41 (4): 811–834.

**Friedrich, Benjamin, and Martin Hackmann.** 2021. "The Returns to Nursing: Evidence from a Parental-Leave Program." *Review of Economic Studies* 88 (5): 2308–2343.

**Gandhi, Ashvin.** 2020. "Picking Your Patients: Selective Admissions in the Nursing Home Industry." Accessed September 9, 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3613950.

**Geng, Fangli, David Stevenson, and David Grabowski.** 2019. "Daily Nursing Home Staffing Levels Highly Variable, Often below CMS Expectations." *Health Affairs* 38 (7): 1095–1100.

**Government Accountability Office.** 2016. "Nursing Homes: Consumers Could Benefit from Improvements to the Nursing Home Compare Website and Five-star Quality Rating System." Accessed August 29, 2024. https://www.gao.gov/products/gao-17-61.

———. 2021. "Additional Reporting on Key Staffing Information and Stronger Payment Incentives Needed for Skilled Nursing Facilities." Accessed January 6, 2025. https://www.gao.gov/assets/gao-21-408.pdf.

**Grabowski, David, Jonathan Gruber, and Joseph Angelelli.** 2008. "Nursing Home Quality as a Common Good." *Review of Economics and Statistics* 90 (4): 754–764.

**Gupta, Atul.** 2021. "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program." *American Economic Review* 111 (4): 1241–1283.

**Gupta, Atul, Sabrina Howell, Constantine Yannelis, and Abhinav Gupta.** 2024. "Owner Incentives and Performance in Healthcare: Private Equity Investment in Nursing Homes." *Review of Financial Studies* 37 (4): 1029–1077.

**Hackmann, Martin.** 2019. "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry." *American Economic Review* 109 (5): 1684–1716.

**Hackmann, Martin, Vincent Pohl, and Nicolas Ziebarth.** 2024. "Patient versus Provider Incentives in Long-term Care." *American Economic Journal: Applied Economics* 16 (3): 178–218.

**Hackmann, Martin, Juan Rojas, and Nicolas Ziebarth.** 2022. "Creative Financing and Public Moral Hazard: Evidence from Medicaid Supplemental Payment." Accessed September 9, 2024. https://drive.google.com/file/d/1SEY5tllP_iFPGAJLSkAB5IsMp58JnxQD/view?pli=1.

**Hart, Oliver, Andrei Shleifer, and Robert W Vishny.** 1997. "The Proper Scope of Government: Theory and An Application to Prisons." *Quarterly Journal of Economics* 112 (4): 1127–1161.

**Jacobs, Andrew.** 2024. "Nursing Home Staffing Shortages and Other Problems Persist, U.S. Report Says." *New York Times.* Accessed August 29, 2024. https://www.nytimes.com/2024/02/29/health/nursing-home-staffing-shortages-pandemic.html.

**Jin, Ginger Zhe, and Phillip Leslie.** 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Quarterly Journal of Economics* 118 (2): 409–451.

**Johnson, Matthew.** 2020. "Regulation by Shaming: Deterrence Effects of Publicizing Violations of Workplace Safety and Health Laws." *American Economic Review* 110 (6): 1866–1904.

**Lu, Susan.** 2012. "Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes." *Journal of Economics & Management Strategy* 21 (3): 673–705.

**McClellan, Mark, Barbara McNeil, and Joseph Newhouse.** 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272 (11): 859–866.

**National Center for Health Statistics.** 2024. "Data from the 2020 National Post-acute and Long-term Care Study." *KFF.* Accessed December 14, 2024. https://www.cdc.gov/nchs/npals/webtables/overview.htm.

**Office of Inspector General.** 2014. "Adverse Events in Skilled Nursing Facilities: National Incidence among Medicare Beneficiaries." Accessed January 6, 2025. https://oig.hhs.gov/reports/all/2014/adverse-events-in-skilled-nursing-facilities-national-incidence-among-medicare-beneficiaries/.

**Olenski, Andrew, and Szymon Sacher.** 2024. "Estimating Nursing Home Quality with Selection." *Review of Economics and Statistics* forthcoming.

**Oster, Emily.** 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

**Rau, Jordan.** 2018. "'It's Almost Like a Ghost Town.' Most Nursing Homes Overstated Staffing for Years." *New York Times.* Accessed August 29, 2024. https://www.nytimes.com/2018/07/07/health/nursing-homes-staffing-medicare.html.

**Research Data Assistance Center.** 2024. "Master Beneficiary Summary File (MBSF): 27 CCW Chronic Conditions Segment." Accessed December 15, 2024. https://resdac.org/cms-data/files/mbsf-27-cc/data-documentation.

**Ruffini, Krista.** 2024. "Worker Earnings, Service Quality, and Firm Profitability: Evidence from Nursing Homes and Minimum Wage Reforms." *Review of Economics and Statistics* 106 (6): 1477–1494.

**Shi, Maggie.** 2024. "Monitoring for Waste: Evidence from Medicare Audits." *Quarterly Journal of Economics* 139 (2): 993–1049.

**Sun, Liyang, and Sarah Abraham.** 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–199.

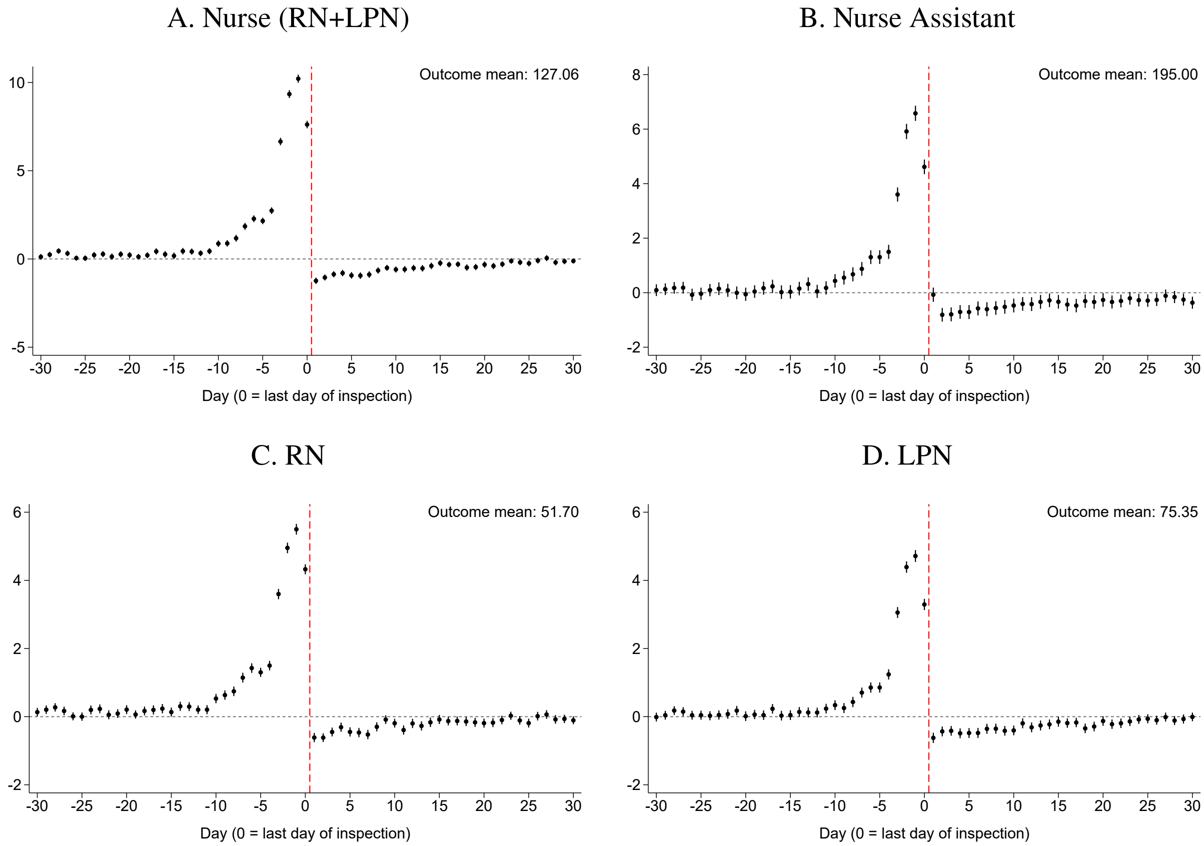
**Thomas, Katie.** 2014. "Medicare Star Ratings Allow Nursing Homes to Game the System." *New York Times.* Accessed September 16, 2024. https://www.nytimes.com/2014/08/25/business/medicare-star-ratings-allow-nursing-homes-to-game-the-system.html.

**U.S. Senate Special Committee on Aging.** 2023. "Uninspected and Negelected: Nursing Home Inspection Agencies are Severely Understaffed, Putting Residents at Risk." Accessed December 1, 2024. https://www.aging.senate.gov/imo/media/doc/UNINSPECTED%20&%20NEGLECTED%20-%20FINAL%20REPORT.pdf.

**Werner, Rachel, R Tamara Konetzka, and Daniel Polsky.** 2016. "Changes in Consumer Demand Following Public Reporting of Summary Quality Ratings: An Evaluation in Nursing Homes." *Health Services Research* 51:1291–1309.

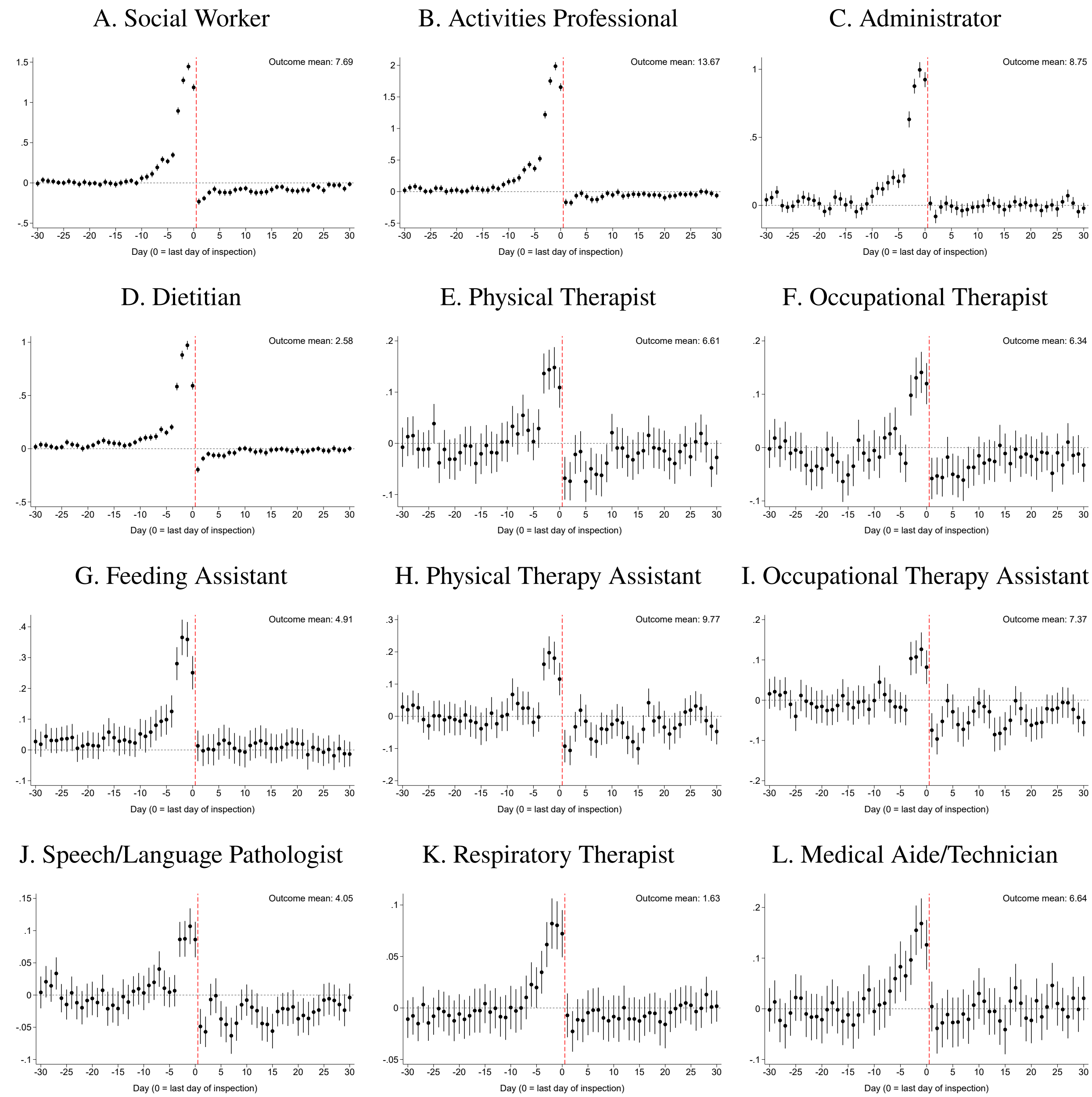
**Zou, Eric Yongchen.** 2021. "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality." *American Economic Review* 111 (7): 2101–2126.

Figure 1: Responses to Health Inspections: Quantity of Labor Inputs (Nursing Staff)



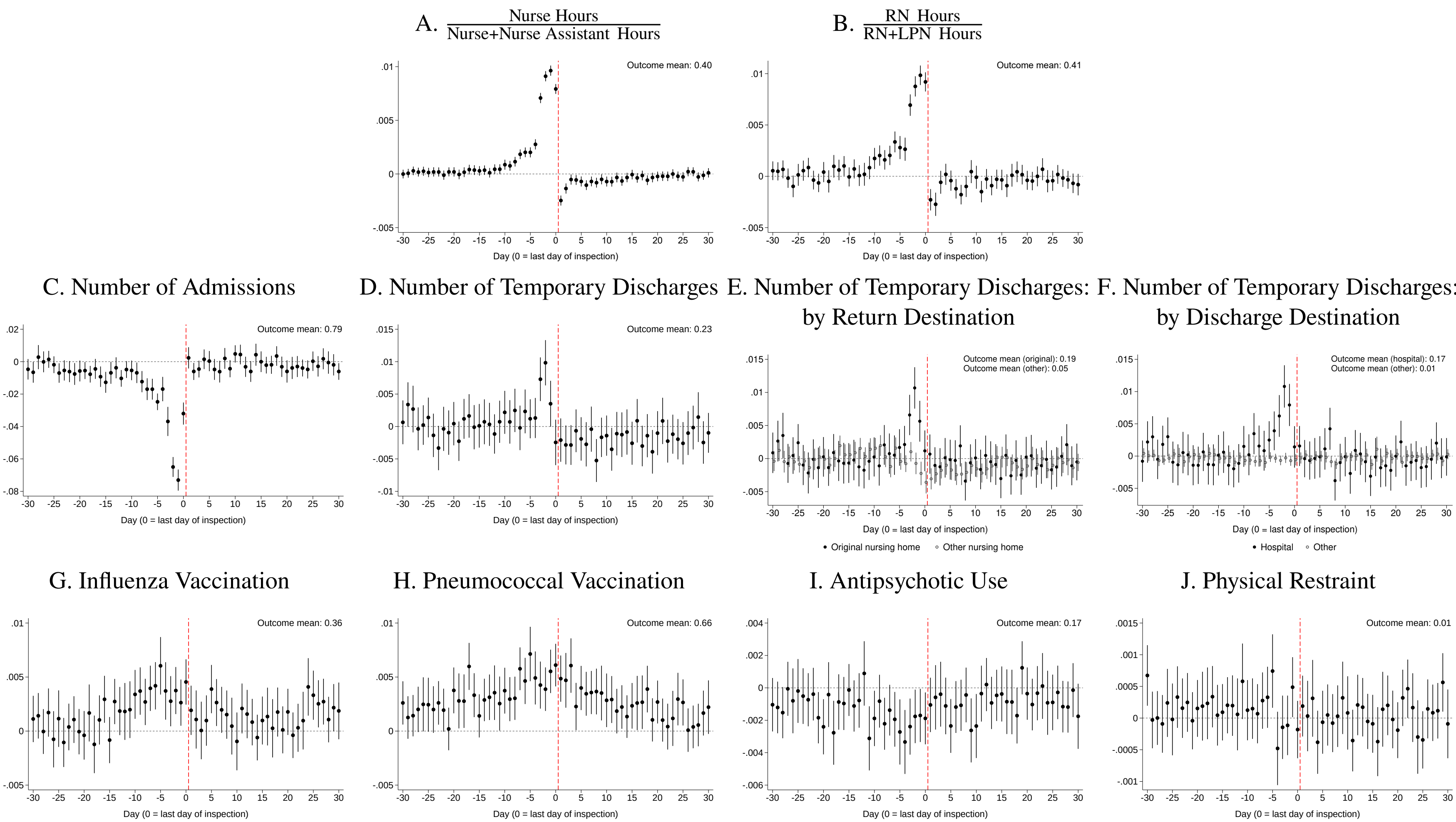

*Notes*: This figure shows changes in staff work hours in response to health inspections. Panels A–D present the results for nurses, nurse assistants, RNs, and LPNs, respectively. Each panel reports estimates and associated 95 percent confidence intervals from a separate regression of the examined work hours on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure 2: Responses to Health Inspections: Quantity of Labor Inputs (Additional Staff)



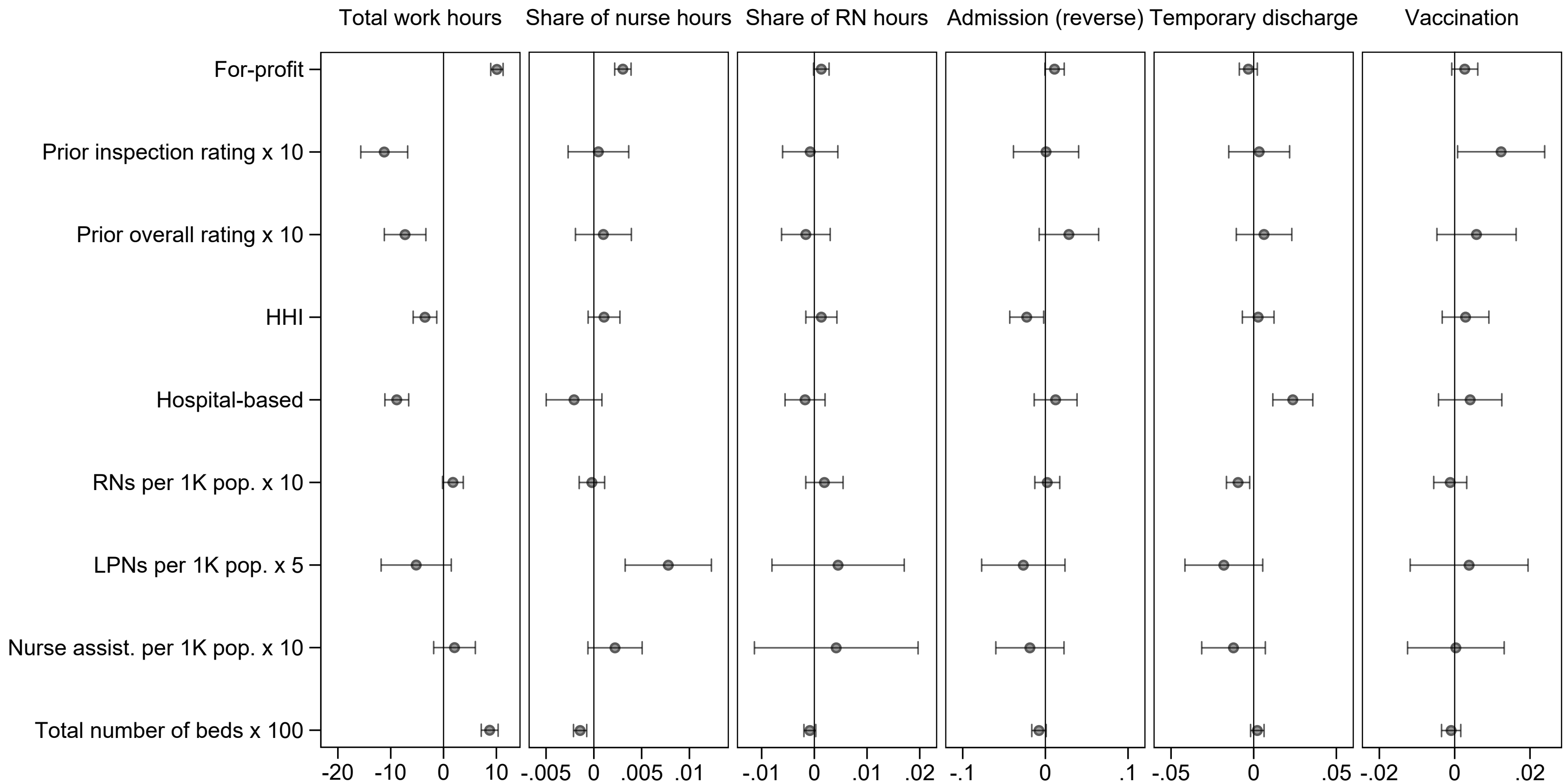

*Notes*: This figure shows changes in staff work hours in response to health inspections, with each panel representing a different type of staff as shown in the panel title. Each panel reports estimates and associated 95 percent confidence intervals from a separate regression of the examined work hours on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure 3: Responses to Health Inspections: Quality of Labor Inputs, Admissions, Temporary Discharges, and Patient Care




*Notes*: This figure reports changes in multiple outcomes in response to health inspections. Panels A–D show results for the share of nursing staff hours provided by nurses, the share of nurse hours provided by RNs, the number of admissions, and the number of temporary discharges, respectively. Panel E decomposes temporary discharges by whether the patient returned to the original nursing homes from which they were discharged or to other nursing homes. Panel F decomposes temporarily discharged cases who returned to the original nursing homes by the destination to which they were temporarily discharged: hospitals or other settings (e.g., homes). Panels G–J show results for whether patients received influenza vaccinations at the facility, pneumococcal vaccinations, antipsychotics, and physical restraints, respectively. Each panel reports estimates and associated 95 percent confidence intervals from a separate regression of the examined outcome on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

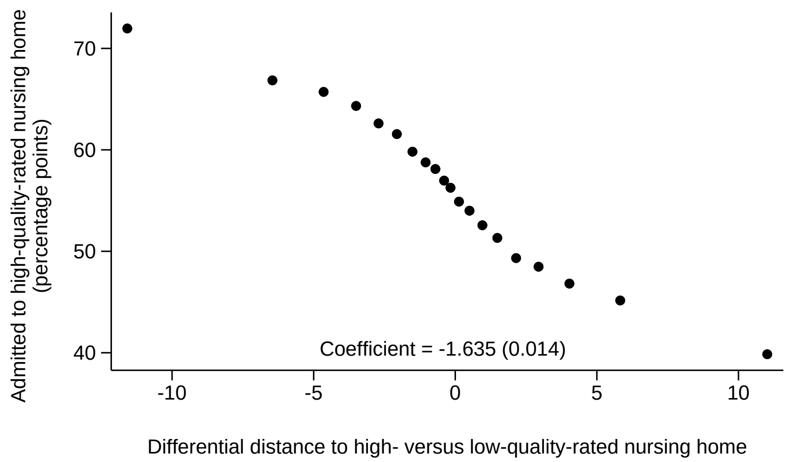
Figure 4: Correlates of Nursing Homes' Responses to Health Inspections



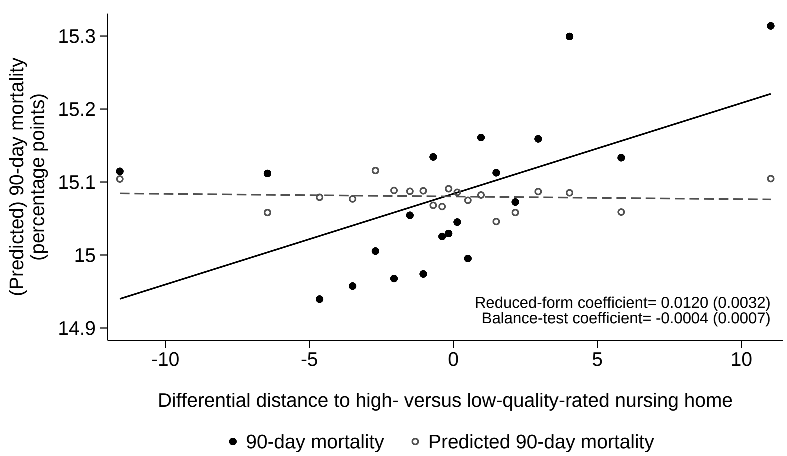

*Notes*: This figure shows correlations between individual nursing homes' responses to health inspections and nursing home and area characteristics. To construct the figure, we first estimate individual nursing homes' responses to health inspections in each outcome listed at the top of each panel. We then correlate these responses with nursing home and area characteristics listed on the *y*-axis using bivariate regressions. The dots and horizontal bars present estimates and associated 95 percent confidence intervals, respectively. To limit estimation noise, we restrict the sample to nursing homes with at least three health inspections during our study period. Appendix A.1 provides details of the analysis. For readability, some coefficients (and their confidence intervals) are scaled up by 5, 10, and 100, as indicated by "× 5," "× 10," and "× 100" on the *y*-axis, respectively. The examined responses from the leftmost to the rightmost panels are, respectively, total staff hours, the share of nursing staff hours provided by nurses, the share of nurse hours provided by RNs, the number of admissions (with the sign reoriented so that a higher value represents a larger admission decline), the number of temporary discharges, and a *z*-score index of influenza and pneumococcal vaccinations (specifically, we first standardize the responses in influenza and pneumococcal vaccinations separately to have a mean of zero and a standard deviation of one; we then take the sum of the two standardized measures).

Figure 5: Relationship between Quality Rating and Quality: First Stage, Balance, and Reduced Form

A. First Stage

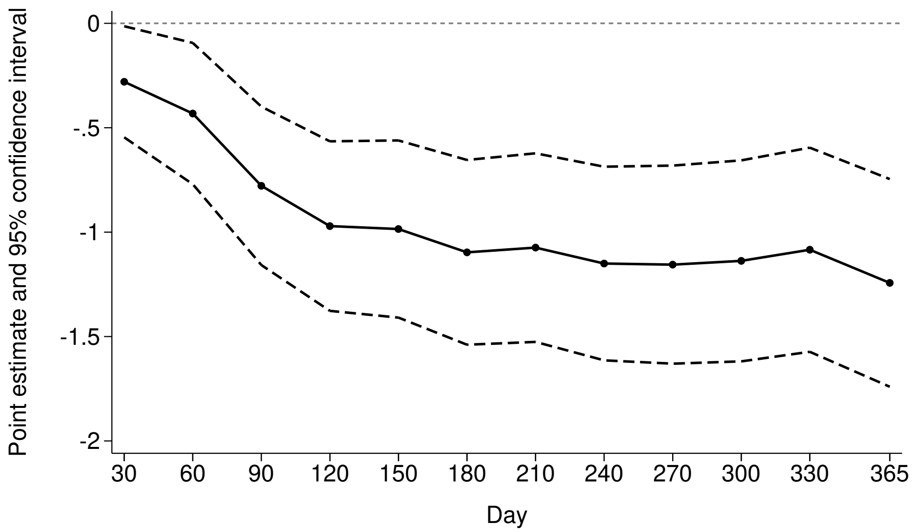



B. Balance and Reduced Form



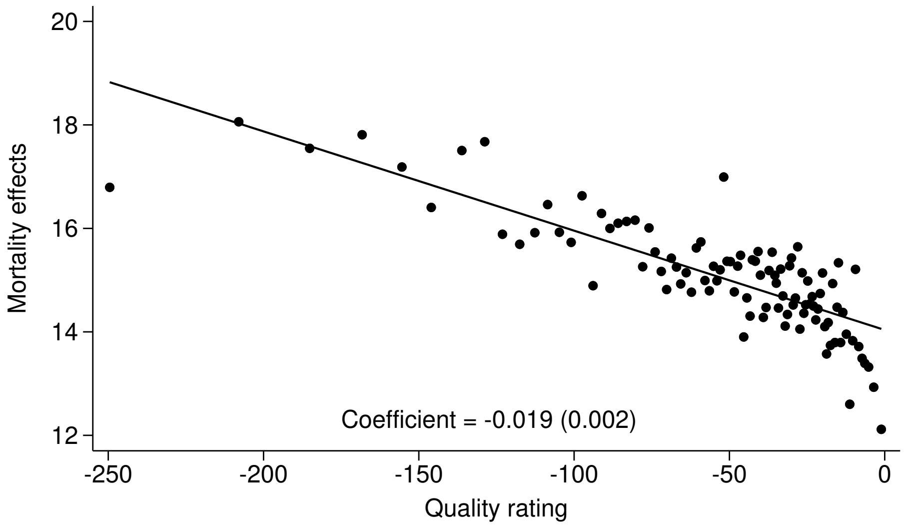

*Notes*: Panel A shows a binned scatter plot of whether the patient is admitted to a high-quality-rated (above-median) nursing home versus the instrument, i.e., the differential distance from the patient's zip code of residence to the nearest high-quality-rated nursing home versus the nearest low-quality-rated nursing home, controlling for the baseline control vector for the IV estimation (i.e., county-by-year indicators). Panel B shows binned scatter plots of patient actual and predicted 90-day mortality on the *y*-axis versus the instrument on the *x*-axis, controlling for the baseline control vector. The solid circles and line represent patients' actual 90-day mortality. The hollow circles and dashed line represent patients' predicted 90-day mortality generated based on patient characteristics $\mathbf{X}_c$ included in Equations (2) and (3), including five-year age bins, gender, black race, Medicaid coverage in the prior year, disability status, and 27 chronic conditions reported in the Chronic Conditions file of Medicare claims at the start of the year of the nursing home admission. To construct the binned scatter plots, we first residualize both the *y*-axis and *x*-axis variables with respect to the baseline controls and then add means back for ease of interpretation. The coefficients report the estimated slope of the best-fit line between the *y*-axis and *x*-axis variables (conditional on the baseline control vector), with standard errors clustered by patient zip code reported in parentheses.

Figure 6: Relationship between Quality Rating and Quality: IV Estimates for Alternative Mortality Outcomes




*Notes*: This figure reports IV estimates of the relationship between admission to high-quality-rated (above-median) nursing homes and patient mortality measured in a series of windows, from 30 days to one year since admission as shown on the *x*-axis. The estimations follow the IV specification in Equations (2) and (3). The solid line connects the point estimates. The dashed lines show the 95 percent confidence intervals. Standard errors are clustered by patient zip code. The sample mean mortality rates for follow-up windows in the order shown on the *x*-axis are, respectively, 6.8, 11.7, 15.1, 17.7, 19.8, 21.5, 23.1, 24.6, 25.9, 27.2, 28.4, and 29.7 percentage points.

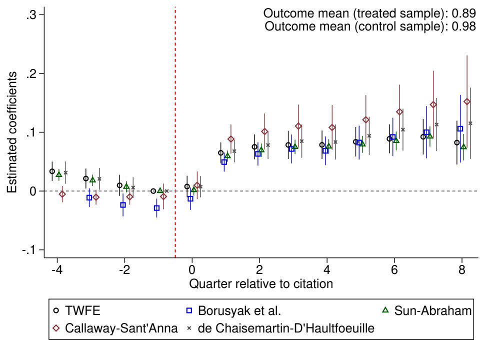
Figure 7: Relationship between Quality Rating and Quality: Individual Nursing Home Mortality Effects and Quality Rating



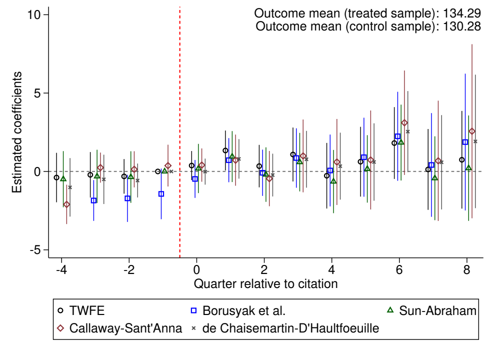

*Notes*: This figure shows a binned scatter plot of individual nursing homes' effects on patient 90-day mortality versus nursing homes' quality ratings from health inspections (i.e., the negative of inspection scores). The coefficient reports the estimated slope of the best-fit line between the *y*-axis and *x*-axis variables, with standard errors clustered by choice set reported in parentheses. Section 5.4 and Appendix A.4 provide details of this analysis.

# Figure 8: Effect of Quality Deficiency Citations

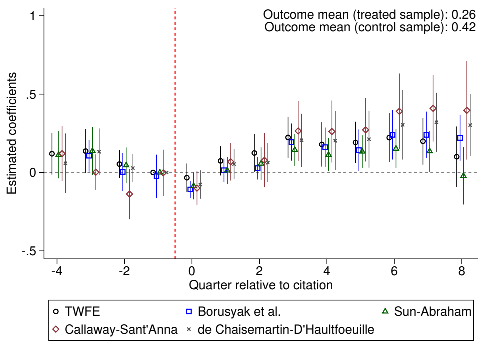
A. RN



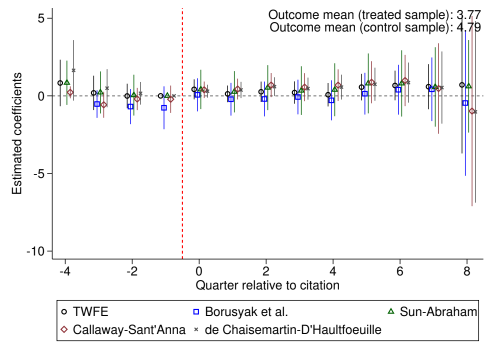

B. Nurse



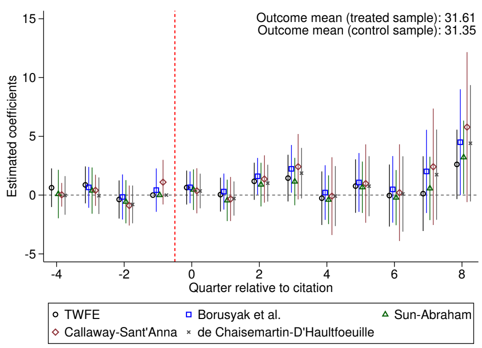

C. Social Worker



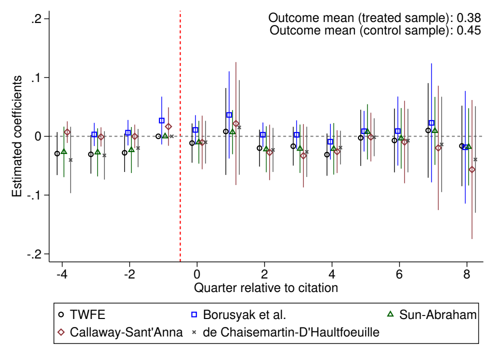

D. Dietary Support Staff



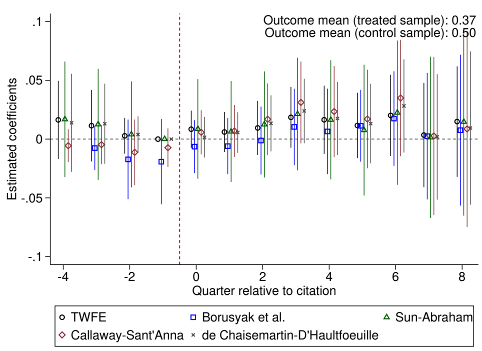

E. Rehabilitative Service




F. Medication Review




G. Pharmacist




H. Pressure Ulcer

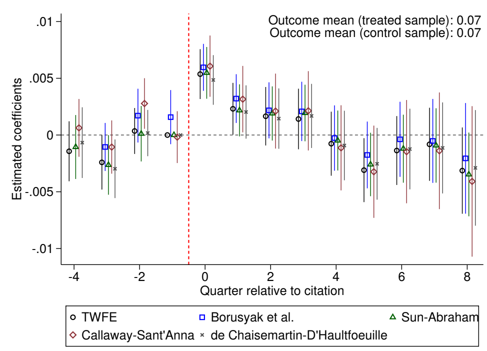
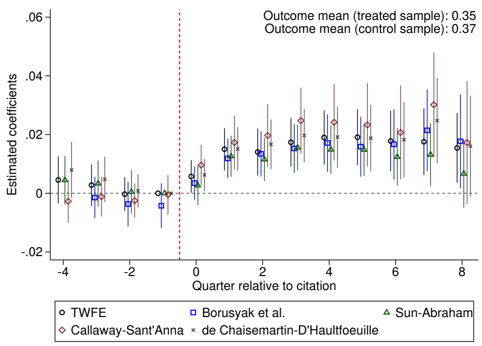
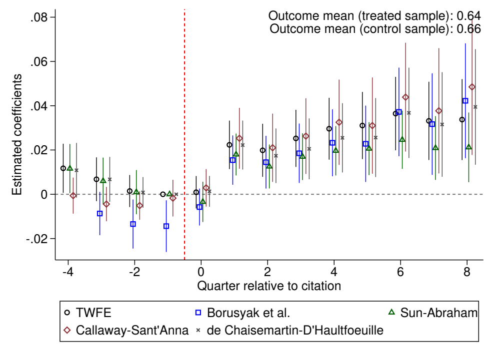


Figure 8: Effect of Quality Deficiency Citations (Continued)

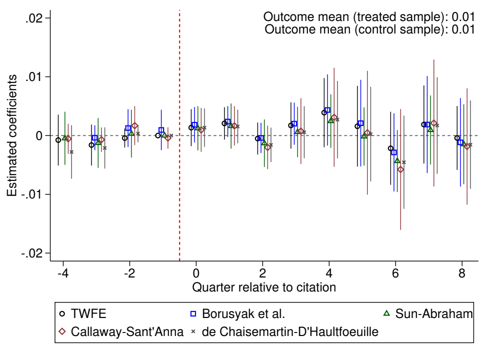
I. Vaccination: Influenza



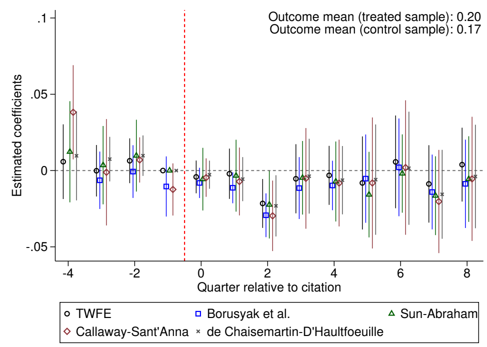

J. Vaccination: Pneumococcal



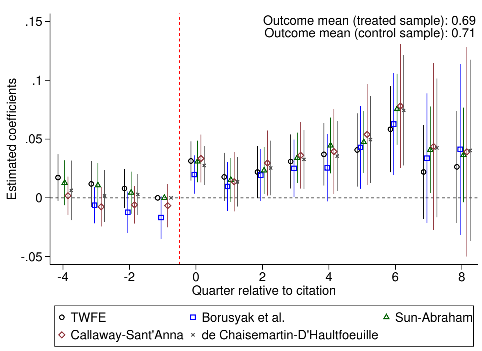

K. Physical Restraint



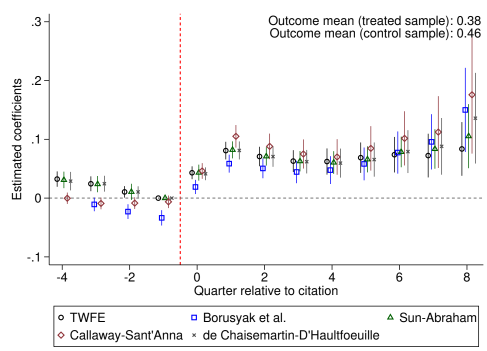

L. Chemical Restraint




M. Assessment Every 3 Months




N. Assessment Transmitted in ≤ 7 days




*Notes*: This figure overlays event-study estimates (along with 95 percent confidence intervals) of the effect of quality deficiency citations using five different estimators: TWFE, Borusyak, Jaravel, and Spiess (2024), Sun and Abraham (2021), Callaway and Sant'Anna (2021), and De Chaisemartin and d'Haultfoeuille (2020) (see details in Section 6.1). Each panel shows results for the citation listed in the panel title, with detailed descriptions of the citation and the examined outcome described in Table 3. For the Borusyak, Jaravel, and Spiess (2024) estimator, we use three preperiods because including more preperiods considerably increases the standard errors for the preperiod coefficients (Borusyak, Jaravel, and Spiess 2024). The unit of observation is at the nursing-home-by-quarter level. Standard errors are clustered by nursing home.

Table 1: Summary Statistics

| | Mean | 25th percentile | Median | 75th percentile | Standard deviation | Nursing home count | Observations |
|---|---|---|---|---|---|---|---|
| **Panel A. Quantity of labor inputs (daily work hours)** | | | | | | | |
| Nursing staff | 322.05 | 190.00 | 283.45 | 403.68 | 198.97 | 15,496 | 6,972,789 |
| Nurse assistant | 195.00 | 113.97 | 170.79 | 244.14 | 124.77 | 15,496 | 6,972,789 |
| Nurse | 127.06 | 71.68 | 109.89 | 161.50 | 81.92 | 15,496 | 6,972,789 |
| RN | 51.70 | 23.50 | 40.55 | 67.00 | 45.42 | 15,496 | 6,972,789 |
| LPN | 75.35 | 37.50 | 65.66 | 100.60 | 53.82 | 15,496 | 6,972,789 |
| Social worker | 7.69 | 0.00 | 7.75 | 10.00 | 9.19 | 15,496 | 6,972,789 |
| Activities professional | 13.67 | 5.03 | 8.75 | 18.50 | 14.53 | 15,496 | 6,972,789 |
| Administrator | 8.75 | 0.00 | 8.00 | 8.00 | 15.60 | 15,496 | 6,972,789 |
| Dietitian | 2.58 | 0.00 | 0.00 | 3.00 | 6.23 | 15,496 | 6,972,789 |
| Physical therapist | 6.61 | 0.00 | 4.00 | 8.55 | 9.32 | 15,496 | 6,972,789 |
| Occupational therapist | 6.34 | 0.00 | 3.68 | 8.42 | 8.89 | 15,496 | 6,972,789 |
| Feeding assistant | 4.91 | 0.00 | 0.00 | 0.00 | 18.04 | 15,496 | 6,972,789 |
| Physical therapy assistant | 9.77 | 0.00 | 6.88 | 15.00 | 12.11 | 15,496 | 6,972,789 |
| Occupational therapy assistant | 7.37 | 0.00 | 5.10 | 11.25 | 9.64 | 15,496 | 6,972,789 |
| Speech/Language therapist | 4.05 | 0.00 | 1.85 | 7.00 | 5.39 | 15,496 | 6,972,789 |
| Respiratory therapist | 1.63 | 0.00 | 0.00 | 0.00 | 10.62 | 15,496 | 6,972,789 |
| Medical aide/Technician | 6.64 | 0.00 | 0.00 | 0.00 | 15.78 | 15,496 | 6,972,789 |
| **Panel B. Quality of labor inputs** | | | | | | | |
| Share of nurse hours | 0.40 | 0.34 | 0.39 | 0.45 | 0.09 | 15,496 | 6,965,695 |
| Share of RN hours | 0.41 | 0.25 | 0.38 | 0.54 | 0.21 | 15,496 | 6,962,863 |
| **Panel C. Admissions** | | | | | | | |
| Number of admissions | 0.79 | 0.00 | 0.00 | 1.00 | 1.19 | 16,034 | 17,158,039 |
| **Panel D. Temporary discharges** | | | | | | | |
| Number of temporary discharges | 0.23 | 0.00 | 0.00 | 0.00 | 0.52 | 16,034 | 17,158,039 |
| **Panel E. Patient care** | | | | | | | |
| Influenza vaccination | 0.36 | 0.00 | 0.00 | 1.00 | 0.48 | 16,017 | 33,952,607 |
| Pneumococcal vaccination | 0.66 | 0.00 | 1.00 | 1.00 | 0.47 | 16,017 | 33,554,117 |
| Antipsychotic use | 0.17 | 0.00 | 0.00 | 0.00 | 0.37 | 16,018 | 34,575,039 |
| Physician restraint | 0.01 | 0.00 | 0.00 | 0.00 | 0.11 | 16,018 | 34,575,331 |

*Notes*: This table presents summary statistics for outcomes in the analysis of responses to health inspections, including the mean, 25th percentile, median, 75th percentile, standard deviation, number of unique nursing homes, and number of observations, from the leftmost to the rightmost column. The number of unique nursing homes and observations varies across outcomes, with detailed explanations provided in footnote 18.

Table 2: Admission to High-Quality-Rated Nursing Homes and 90-Day Mortality

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Panel A: Reduced-form estimates | | | | |
| Differential distance | 0.0120 | 0.0158 | 0.0154 | 0.0126 |
| | (0.0032) | (0.0032) | (0.0032) | (0.0031) |
| Panel B: IV estimates | | | | |
| High-quality-rated nursing home | -0.7333 | -0.9707 | -0.9462 | -0.7781 |
| | (0.1962) | (0.1981) | (0.1980) | (0.1935) |
| County-by-year indicators | Yes | Yes | Yes | Yes |
| Demographics | No | Yes | Yes | Yes |
| Medicaid coverage (prior year) | No | No | Yes | Yes |
| Comorbidities, disability | No | No | No | Yes |
| Outcome mean | 15.08 | 15.08 | 15.08 | 15.08 |
| Outcome S.D. | 35.79 | 35.79 | 35.79 | 35.79 |
| Observations | 5,578,510 | 5,578,510 | 5,578,510 | 5,578,510 |

*Notes*: This table shows the relationship between admission to high-quality-rated (above-median) nursing homes and patient 90-day mortality. Panel A reports the reduced-form estimates; Panel B reports the IV estimates. Patient demographic controls include five-year age bins, gender, and black race. We control for Medicaid coverage in the prior year rather than the current year because nursing home stays may affect Medicaid coverage. Comorbidities include the 27 chronic conditions reported in the Chronic Conditions file of Medicare claims at the start of the year of the nursing home admission. For each patient covariate with missing values, we add an indicator for missing values and replace the missing values with zero. Standard errors clustered by patient zip code are reported in parentheses.

Table 3: Examined Quality Deficiency Citations and Outcome Measures

| Label | Version | Description | Outcome measure |
|---|---|---|---|
| RN | 1 | Use a RN at least 8 hours a day, 7 days a week | Share of days with RN hours $\geq 8$ |
| | 2 | Have a RN on duty 8 hours a day; and select a RN to be the director of nurses on a full time basis | |
| Nurse | 1 | Have enough nurses to care for every resident in a way that maximizes the resident's well being | Average daily nurse hours |
| | 2 | Provide enough nursing staff every day to meet the needs of every resident; and have a licensed nurse in charge on each shift | |
| Social worker | 1 | Hire a qualified full-time social worker in a facility with more than 120 beds | Have qualified full-time social worker (sample: facilities with > 120 beds at baseline) |
| | 2 | Hire a qualified full-time social worker in a facility with more than 120 beds | |
| Dietary support staff | 1 | Hire sufficient dietary support personnel | Average daily feeding assistant hours |
| | 2 | Provide sufficient support personnel to safely and effectively carry out the functions of the food and nutrition service | |
| Rehabilitative service | 1 | Give or get specialized rehabilitative services per the patient's assessment or plan of care | Average daily hours of occupational/ physical therapists and occupational/ physical therapy assistants |
| | 2 | Provide or get specialized rehabilitative services as required for a resident | |
| | 1 | Give specialized rehabilitative services that are medically necessary by qualified personnel, when ordered by a doctor | |
| | 2 | Provide specialized rehabilitative services by qualified personnel, when ordered for a resident by a doctor | |
| Medication review | 1 | At least once a month, have a licensed pharmacist review each resident's medication(s) and report any irregularities to the attending doctor | Average daily pharmacist hours |
| | 2 | Ensure a licensed pharmacist perform a monthly drug regimen review, including the medical chart, following irregularity reporting guidelines in developed policies and procedures | |
| Pharmacist | 1 | Provide routine and emergency drugs through a licensed pharmacist and only under the general supervision of a licensed nurse | Average daily pharmacist hours |
| | 2 | Provide pharmaceutical services to meet the needs of each resident and employ or obtain the services of a licensed pharmacist | |
| Pressure ulcer | 1 | Give residents proper treatment to prevent new bed (pressure) sores or heal existing bed sores | Share of residents with pressure ulcers |
| | 2 | Provide appropriate pressure ulcer care and prevent new ulcers from developing | |
| Vaccination | 1 | Develop policies and procedures for influenza and pneumococcal immunizations | Share of residents with influenza/ pneumococcal vaccination |
| | 2 | Develop and implement policies and procedures for flu and pneumonia vaccinations | |
| Physical restraint | 1 | Keep each resident free from physical restraints, unless needed for medical treatment | Share of residents with physical restraints |
| | 2 | Ensure that each resident is free from the use of physical restraints, unless needed for medical treatment | |
| Chemical restraint | 1 | Keep each resident free from drugs that restrain them, unless needed for medical treatment | Share of residents with antipsychotics |
| | 2 | Ensure that each resident is free from medications that restrain them, unless needed for medical treatment | |
| Assessment every 3 months | 1 | Assure that each resident's assessment is updated at least once every 3 months | Share of residents with the next assessment within 3 months |
| | 2 | Assure that each resident's assessment is updated at least once every 3 months | |
| Assessment transmitted in $\leq 7$ days | 1 | Encode each resident's assessment data and transmit these data to the State within 7 days of assessment | Share of assessments transmitted in $\leq 7$ days |
| | 2 | Encode each resident's assessment data and transmit these data to the State within 7 days of assessment | |

*Notes*: This table presents the quality deficiency citations included in our analyses. Column 1 shows the label we add to each citation. Column 2 shows the version of the citation descriptions provided by CMS. As CMS modified citation descriptions during our study period, we include both versions. Versions "1" and "2" represent the earlier and updated versions, respectively. Column 3 shows the citation descriptions provided by CMS. Column 4 shows the examined outcomes.

# Appendix

## A.1 Correlates of Nursing Homes' Responses to Health Inspections

In this appendix, we describe our analysis of the correlates of nursing homes' responses to health inspections. We first estimate responses for each nursing home separately using the following specification:

$$y_{it} = \sum_{k=-30}^{k=30} \beta_k^i d_{it}^k + \mathbf{I}_{it}\theta^i + \epsilon_{it}, \tag{A.1}$$

where $y_{it}$ represents the outcome of interest on calendar day $t$ at nursing home $i$, and $d_{it}^k$ denotes event day indicators that take the value of one if day $t$ is $k$ days away from the last day of an inspection. The analysis includes observations ranging from -100 to 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. The coefficients of interest are $\beta_k^i$, representing differences in outcomes between a day that is $k$ days away from the last day of an inspection and days in the reference group for nursing home $i$. The vector $\mathbf{I}_{it}$ contains the baseline controls, i.e., year, month, and day-of-the-week indicators, which capture potential systematic differences across time categories in the nursing home. Finally, $\epsilon_{it}$ is the error term. To limit estimation noise, we include only nursing homes with at least three health inspections during our study period.

Next, we estimate the correlation between individual nursing homes' responses to health inspections and nursing home and area characteristics:

$$y_i = \rho x_i + \nu_i, \tag{A.2}$$

where $y_i$ is nursing home $i$'s response to health inspections, $x_i$ is the characteristic of interest, and $\nu_i$ is the error term. The coefficient of interest is $\rho$. Since event day -1 is the day of peak response for most outcomes examined and presents statistically significant responses to health inspections for all outcomes that show responses to inspections (see Section 4), we use $\beta_{-1}^i$ estimated from Equation (A.1) to represent a nursing home's response to health inspections. We set $\beta_{-1}^i$ to zero if it is statistically insignificant at the 10 percent level. Results are qualitatively similar when we use an alternative cutoff of 5 percent. Since the responses in antipsychotic and physical restraint use to health inspections are less precise, we do not include them in this analysis.

## A.2 Selection on Patient Unobservables

In this appendix, we describe our analysis that explores the robustness of the estimates presented in Section 5.2 to selection on patient unobservable characteristics. Specifically, using an approach by Oster (2019), we report the adjusted coefficient estimate $\sigma^*$ constructed as

$$\sigma^* = \tilde{\sigma} - \kappa[\mathring{\sigma} - \tilde{\sigma}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}, \tag{A.3}$$

where $\tilde{\sigma}$ and $\mathring{\sigma}$ are the coefficient estimates from the regressions with the full set of patient covariates and no patient covariates, respectively; $\tilde{R}$ and $\mathring{R}$ are the corresponding $R$-squared values. Following Oster (2019), we set $\kappa$, i.e., the relative degree of selection on observed and unobserved characteristics, equal to one and set $R_{max}$, the maximum $R$-squared possible, equal to $min(1, \Pi * \tilde{R})$. We use the rule of thumb of $\Pi = 1.3$. Since our baseline specification already controls for county-by-time-category fixed effects, we use the $R$-squared after we partial out county-by-time-category fixed effects. Column 3 of Appendix Table A.4 shows that the estimates adjusted for potential selection on patient unobservables are similar to our main estimates, suggesting that our main estimates are unlikely to be explained away by unobserved patient selection.

## A.3 Characterizing Compliers

This appendix describes the examination of complier characteristics. Following the approach developed by Abadie (2003), we estimate complier characteristics using the following 2SLS model:

$$x_c \times \text{high}_c = \pi_1 \text{high}_c + \mathbf{L}_c \lambda_1 + \varepsilon_{1,c}, \tag{A.4}$$

$$\text{high}_c = \pi_2 Z_c + \mathbf{L}_c \lambda_2 + \varepsilon_{2,c}, \tag{A.5}$$

where $\pi_1$ is the parameter of interest, and $x_c \times \text{high}_c$ denotes the interaction between each patient characteristic and the indicator for admission to a high-quality-rated nursing home. Following the main estimation in Equations (2) and (3), we cluster standard errors by patient zip code. Results are shown in Appendix Table A.3. For each patient characteristic, we also compute the ratio between the complier mean and the overall sample mean, along with the 95% confidence interval of each ratio.

## A.4 Individual Nursing Homes' Mortality Effects and Quality Rating

### A.4.1 Estimating Individual Nursing Homes' Effects on Patient 90-day Mortality

In this appendix, we describe the estimation of each nursing home's effect on patient 90-day mortality. To account for non-random sorting of patients to nursing homes, we adopt a control function approach following Dubin and McFadden (1984). We leverage distances between the patient's residence and each nursing home in their choice set as excluded instruments as in Einav, Finkelstein, and Mahoney (2025). Specifically, assuming that the utility of case $c$ from nursing home $i$ is

$$U_{ci} = \zeta_i(\mathbf{X}_c) - \tau d_{ci} + \xi_{ci}, \tag{A.6}$$

where $\zeta_i(\mathbf{X}_c)$ denotes the average utility from nursing home $i$ for patients with characteristics $\mathbf{X}_c$, $d_{ci}$ denotes the distance between case $c$'s zip code of residence and nursing home $i$'s zip code, and $\xi_{ci}$ denotes an error term drawn i.i.d. from a Type I Extreme Value distribution.

Assuming that the control function is linear in the demeaned logit errors, then conditional on the choice of nursing home $n_c$, patient observable characteristics $\mathbf{X}_c$, and the demand shocks $\xi_{ci}$, expected 90-day mortality for case $c$ can be written as

$$E[m_c|\mathbf{X}_c, n_c = i, \xi_{c1}, ..., \xi_{c|J_c|}] = \alpha_i + \mathbf{X}_c\beta + \left[\sum_{\ell \in J_c} \phi_\ell(\xi_{c\ell} - \mu_\xi)\right] + \varphi(\xi_{ci} - \mu_\xi), \tag{A.7}$$

where $J_c$ represents case $c$'s choice set, and $\mu_\xi$ denotes the mean logit errors. As shown by Dubin and McFadden (1984) and Einav, Finkelstein, and Mahoney (2025), integrating out the $\xi$'s yields

$$E[m_c|\mathbf{X}_c, n_c = i, \xi_{c1}, ..., \xi_{c|J_c|}] = \alpha_i + \mathbf{X}_c\beta + \left[\sum_{\ell \in J_c} \phi_\ell \gamma_{c\ell}\right] + \varphi\gamma_{ci}, \tag{A.8}$$

where $\gamma_{c\ell}$'s are functions of $\hat{p}_{c\ell}$, i.e., the predicted probability that case $c$ chooses nursing home $\ell$ based on the demand model in Equation (A.6):

$$\gamma_{c\ell}(i) = \begin{cases} -log\ \hat{p}_{c\ell} & \ell = i \\ \frac{\hat{p}_{c\ell}}{1-\hat{p}_{c\ell}} log\ \hat{p}_{c\ell} & \text{otherwise} \end{cases}. \tag{A.9}$$

As discussed in Einav, Finkelstein, and Mahoney (2025), $\phi_\ell\gamma_{c\ell}$ controls for the unobserved correlation between demand shocks for nursing home $\ell$ and the value added of the nursing home. For example, $\phi_\ell\gamma_{c\ell}$ can correct for the bias that would arise if patients in unobservably worse health sort into higher-quality nursing homes. The term $\varphi\gamma_{ci}$ corrects for potential correlation between demand shocks for the chosen nursing home $i$ and the value added of the nursing home beyond that captured by the $\phi_\ell\gamma_{c\ell}$'s.

We define the choice set of nursing homes for each patient using an approach similar to Einav, Finkelstein, and Mahoney (2025). Specifically, we start with Hospital Referral Regions (HRRs) and split large HRRs into sub-HRRs for computational tractability. We define sub-HRRs as having over 80 percent of patients in the 25 largest nursing homes. Of the 306 HRRs, 161 satisfy this criterion and are not split. The remaining 145 HRRs are divided into sub-HRRs as follows: (i) split each HRR into two groups which are geographically connected Hospital Service Areas (HSAs), using $k$-means clustering on the longitude and latitude of HSA centroids to ensure that sub-HRRs are geographically connected; (ii) repeat (i) until every sub-HRR satisfies the criterion or has only one HSA. In total, we obtain 886 sub-HRRs.

For a detailed discussion about the validity of this estimation approach, see Einav, Finkelstein, and Mahoney (2025).

### A.4.2 Correlation between Individual Nursing Homes' Mortality Effects and Quality Rating

We next correlate the estimated individual nursing homes' effects on patient 90-day mortality with quality ratings from health inspections as follows:

$$\hat{\alpha}_i = \lambda \text{rating}_i + \varepsilon_i, \tag{A.10}$$

where $\hat{\alpha}_i$ is nursing home $i$'s effect on patient 90-day mortality estimated in Appendix Section A.4.1, and $\text{rating}_i$ is nursing home $i$'s quality rating from health inspections. In health inspection ratings, higher scores represent lower quality performance; therefore, we take the negative of the scores so that higher values represent higher quality ratings. We define $\text{rating}_i$ as the average of the nursing home's ratings during our study period 2014-2019. Finally, $\varepsilon_i$ is the error term. We cluster standard errors by choice set, i.e., sub-HRR. To limit the influence of estimation noise, we restrict the sample to nursing homes with at least 50 cases when estimating $\lambda$.

## A.5 Estimating Variance of Nursing Home Mortality Effects

In this appendix, we describe our estimation of the variance of nursing home mortality effects, $\alpha_i$'s. To account for sampling error due to the fact that $\alpha_i$'s are estimated on a finite sample, we leverage a split-sample approach resembling that employed in earlier studies (e.g., Silver 2021; Chan and Chen 2022). Specifically, we first randomly split a nursing home's patients into two equal-sized partitions. We then estimate individual nursing homes' mortality effects using each partition separately as in Section A.4.1, yielding two fixed effect estimates for each facility: $\hat{\alpha}_{i,1}$ and $\hat{\alpha}_{i,2}$. Suppressing the subscript $i$ for simplicity, we have

$$\hat{\alpha}_p = \alpha + e_p, p \in \{1, 2\},$$

where $p$ denotes partitions, and $e_p$ denotes partition-specific sampling error. We have $\text{Cov}(\alpha, e_1) = \text{Cov}(\alpha, e_2) = 0$. By random split, we also have $\text{Cov}(e_1, e_2) = 0$. Thus, the variance of nursing home mortality effects can be computed as the covariance of $\hat{\alpha}_1$ and $\hat{\alpha}_2$:

$$\begin{aligned}
\text{Cov}(\hat{\alpha}_1, \hat{\alpha}_2) &= \text{Cov}(\alpha + e_1, \alpha + e_2) \\
&= \text{Cov}(\alpha, \alpha) + \text{Cov}(\alpha, e_2) + \text{Cov}(e_1, \alpha) + \text{Cov}(e_1, e_2) \\
&= \text{Var}(\alpha).
\end{aligned}$$

We first apply the split-sample approach to estimate the variance of $\alpha_i$'s, yielding a variance estimate of 204.9. Next, we apply the split-sample approach to estimate the variance of $\alpha_i$'s residualized with respect to quality ratings from health inspections, yielding a variance estimate of 183.8. Taken together, 10.3 percent $(1 - \frac{183.8}{204.9})$ of the variance in nursing home mortality effects is explained by quality ratings.

## References

**Abadie, Alberto.** 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–263.

**Chan, David, and Yiqun Chen.** 2022. "The Productivity of Professions: Evidence from the Emergency Department." NBER Working Paper 30608.

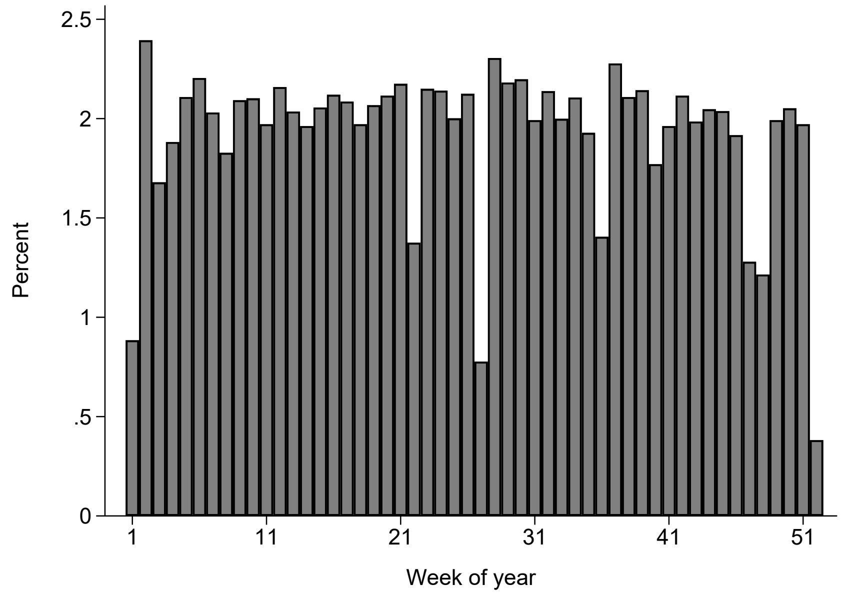
**Dubin, Jeffrey, and Daniel McFadden.** 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica,* 345–362.

**Einav, Liran, Amy Finkelstein, and Neale Mahoney.** 2025. "Producing Health: Measuring Value Added of Nursing Homes." *Econometrica* 93 (4): 1225–1264.

**Oster, Emily.** 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

**Silver, David.** 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Review of Economic Studies* 88 (3): 1385–1417.

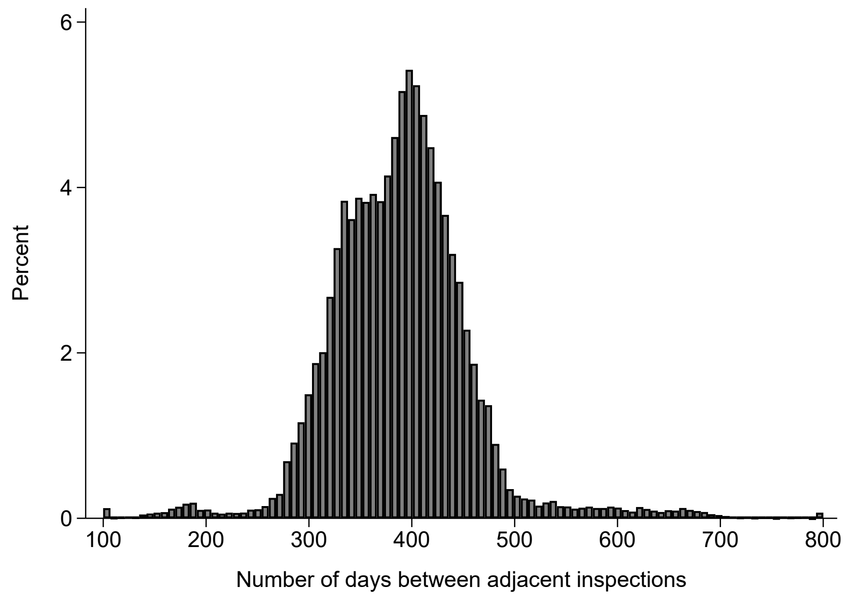
Figure A.1: Health Inspection Timing



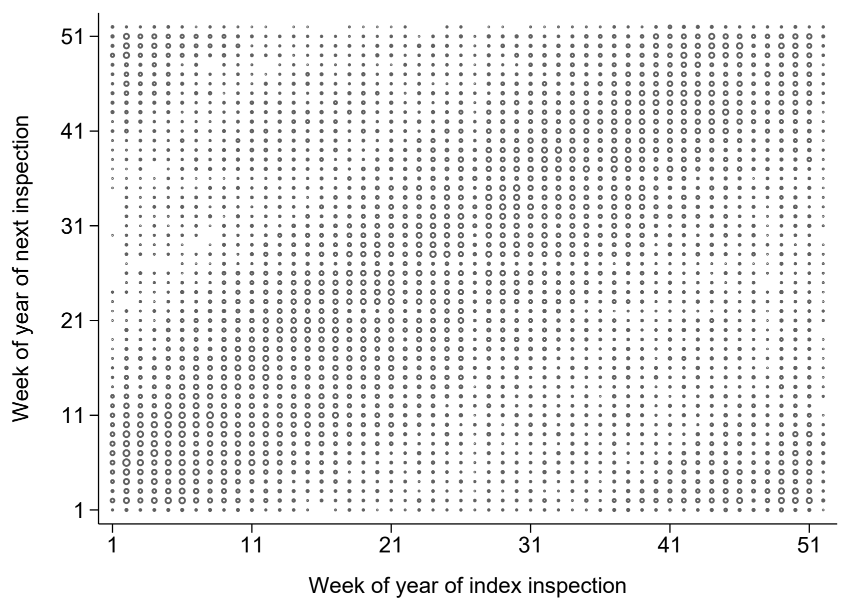

*Notes*: This figure shows a histogram of the week of the year of health inspections based on inspection end dates.

Figure A.2: Time between Adjacent Health Inspections

A. Number of Days between Adjacent Inspections



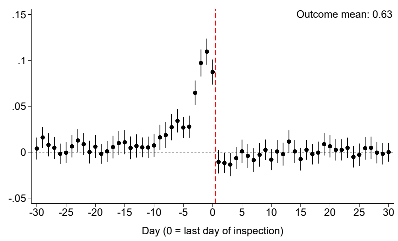

B. Week of the Year of Adjacent Inspections



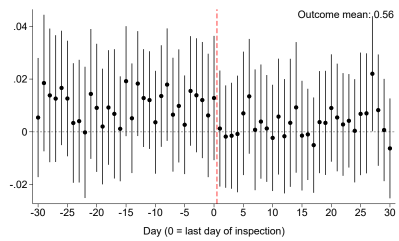

*Notes*: Panel A shows the number of days between adjacent health inspections. Panel B shows the week of the year of adjacent health inspections; circle sizes are proportional to the number of observations.

Figure A.3: Responses to Health Inspections: Quantity of Labor Inputs (Additional Staff: Average Daily Work Hours below One Hour)

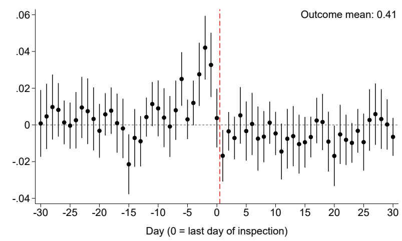
A. Therapeutic Recreation Specialist



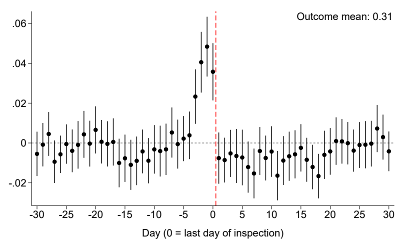

B. Physician



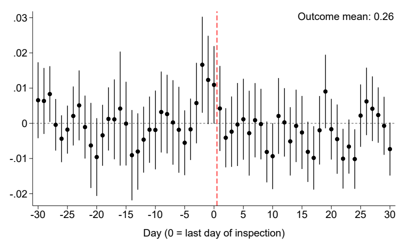

C. Pharmacist



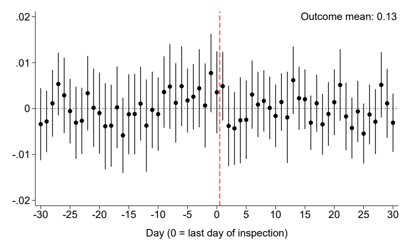

D. Advanced Practice RN



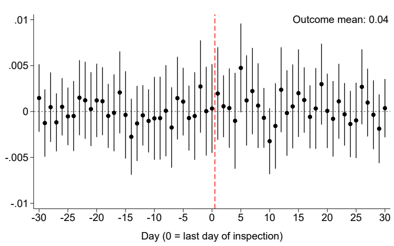

E. Mental Health Service Worker




F. Respiratory Therapy Technician

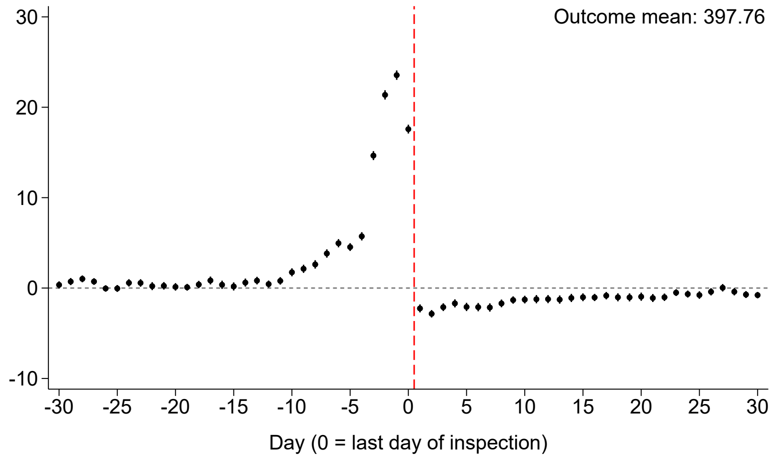



G. Physician Assistant




*Notes*: This figure shows changes in work hours in response to health inspections for staff with average daily work hours below one hour, with each panel representing a different type of staff as shown in the panel title. Each panel presents estimates and associated 95 percent confidence intervals from a separate regression of the examined work hours on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure A.4: Responses to Health Inspections: Quantity of Labor Inputs (Total Staff Hours)




*Notes*: This figure shows changes in total staff hours (the sum of all staff hours in Panels A and B of Figure 1, Figure 2, and Appendix Figure A.3) in response to health inspections. The figure presents estimates and associated 95 percent confidence intervals from regressing total staff hours on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure A.5: Responses to Health Inspections: Quantity of Labor Inputs (Employee versus Contract Worker Hours)

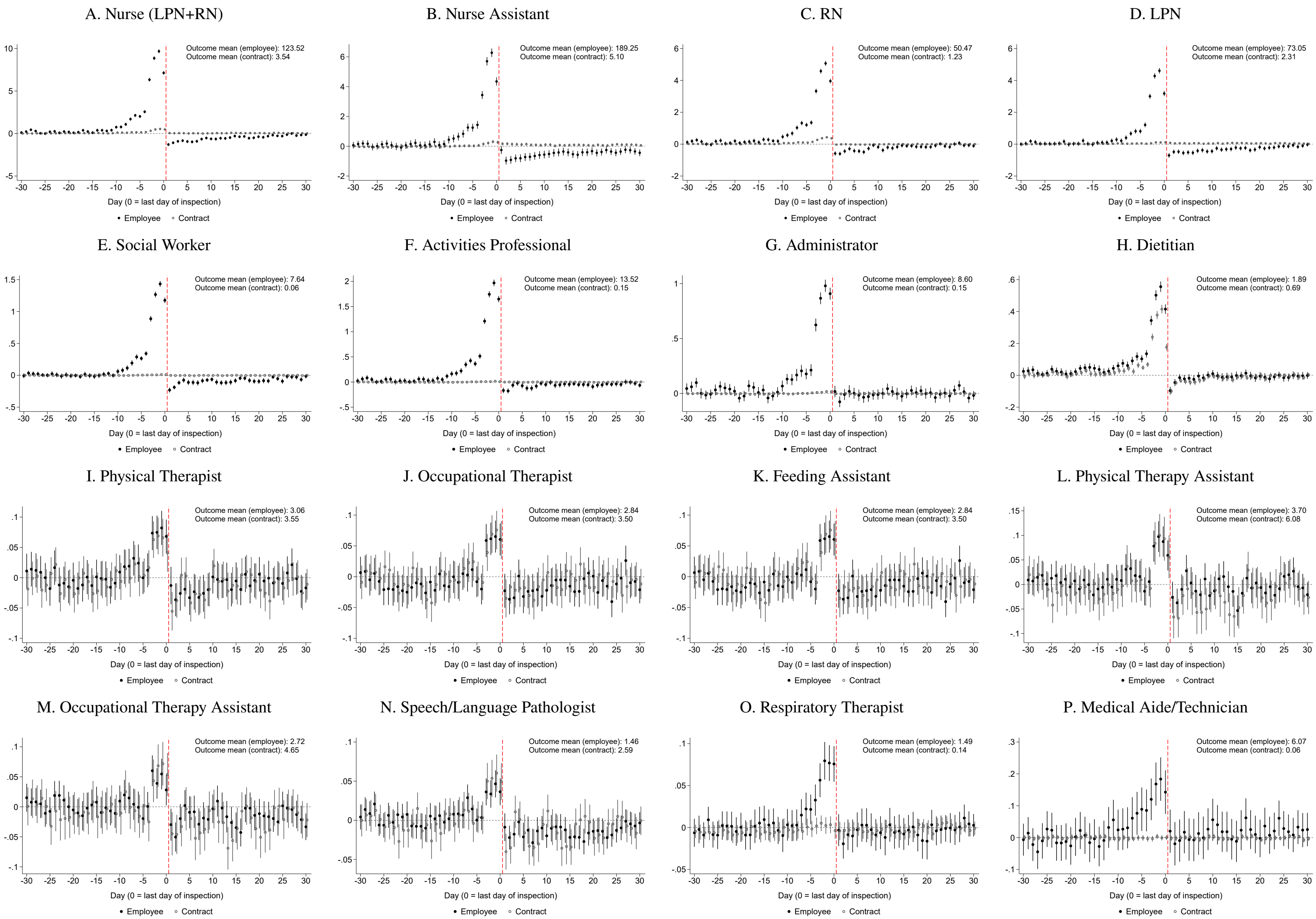
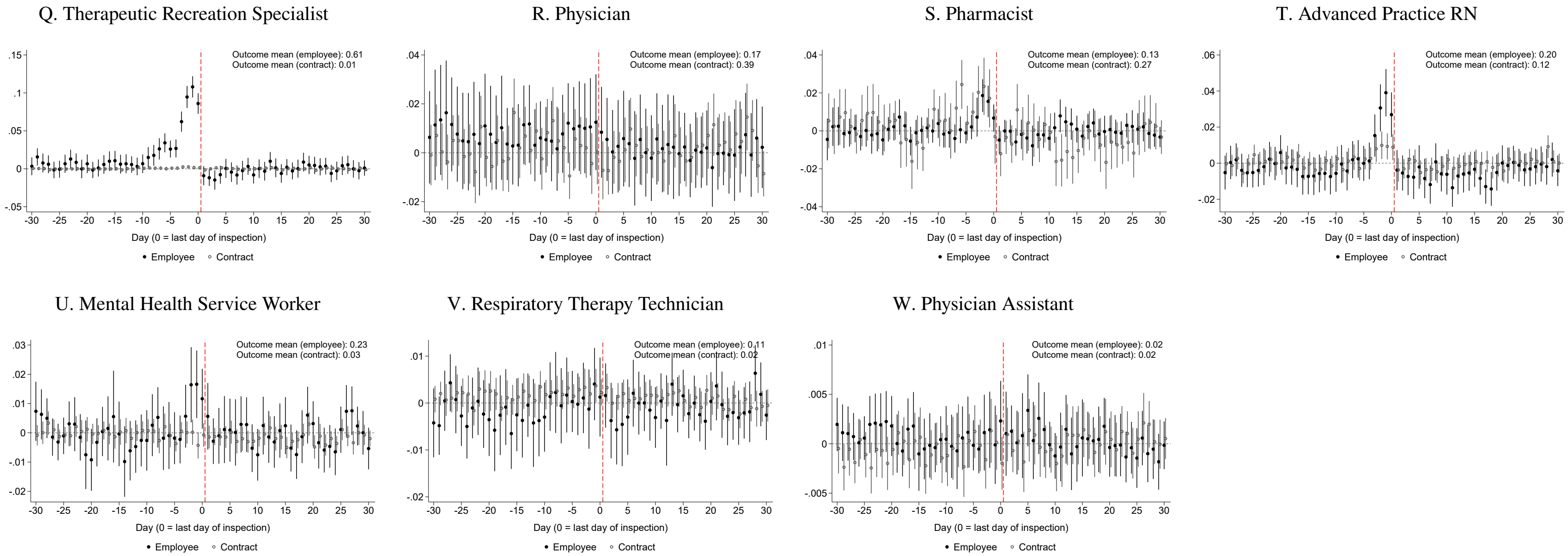
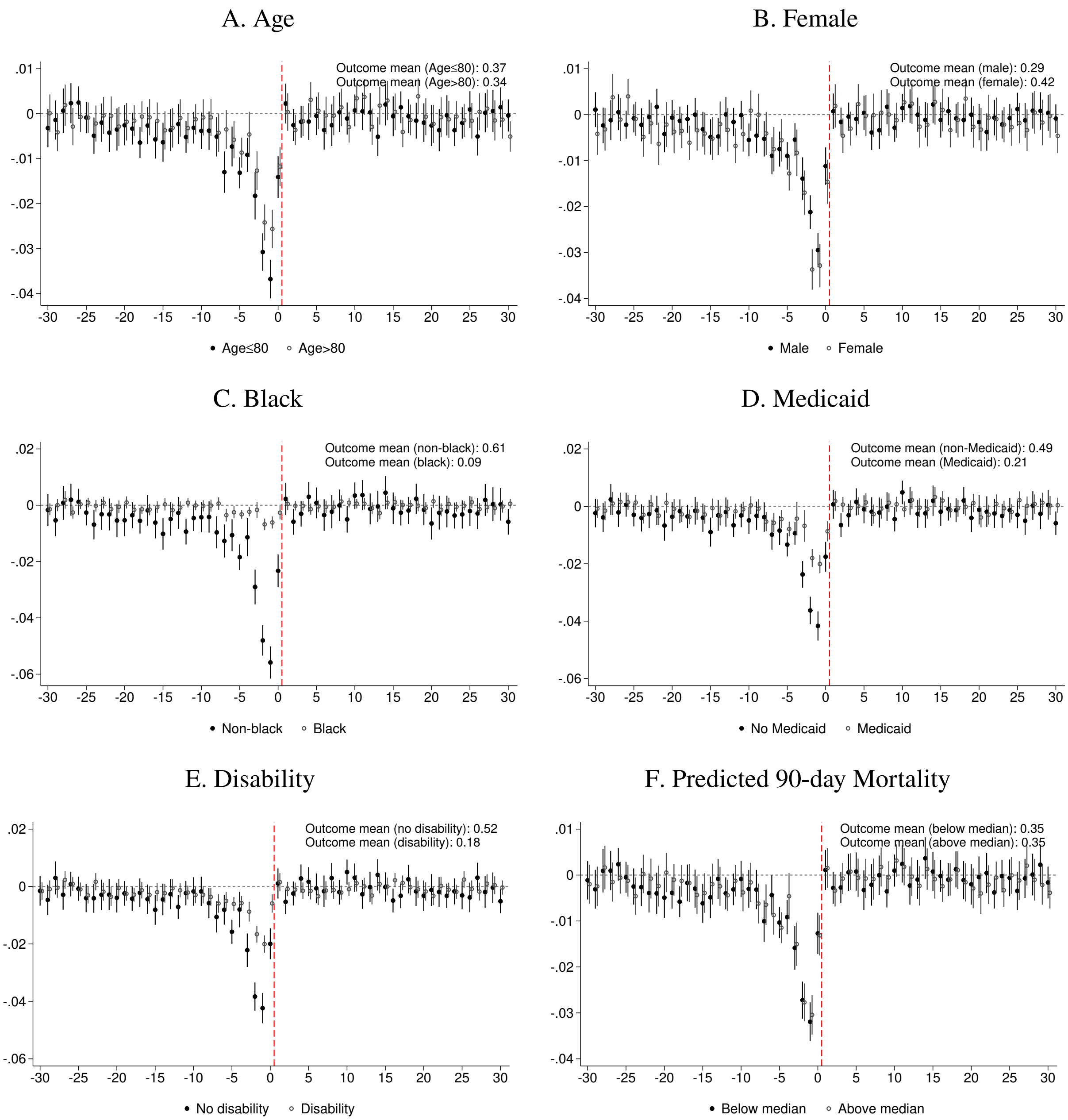


Figure A.5: Responses to Health Inspections: Quantity of Labor Inputs (Employee versus Contract Worker Hours, Continued)




*Notes*: This figure shows changes in staff hours in response to health inspections, decomposing hours of each type of staff shown in the panel title into those provided by employees versus contract workers. Each panel presents estimates and associated 95 percent confidence intervals from a separate regression of the examined employee/contract worker hours on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

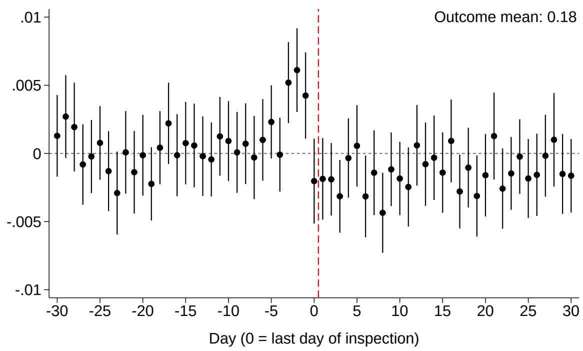
Figure A.6: Responses to Health Inspections: Admissions by Patient Characteristics



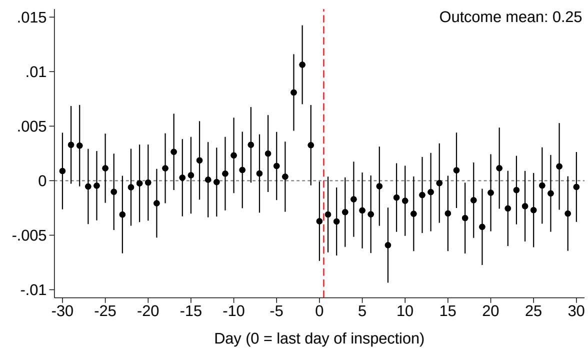

*Notes*: This figure reports changes in the number of admissions in response to health inspections by patient characteristics. Panels A–F decompose admissions by patient age (≤ 80 versus > 80), gender, race, Medicaid coverage in the prior year, disability status, and predicted 90-day mortality (below versus above median), respectively. The sum of mean admission rates in each panel is slightly below the mean shown in Panel C of Figure 3 due to a small share of missing values in the studied patient characteristic. Each panel presents estimates and associated 95 percent confidence intervals from a separate regression of the examined admissions on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure A.7: Responses to Health Inspections: Temporary Discharges Using Alternative Definitions

A. Return in One Week

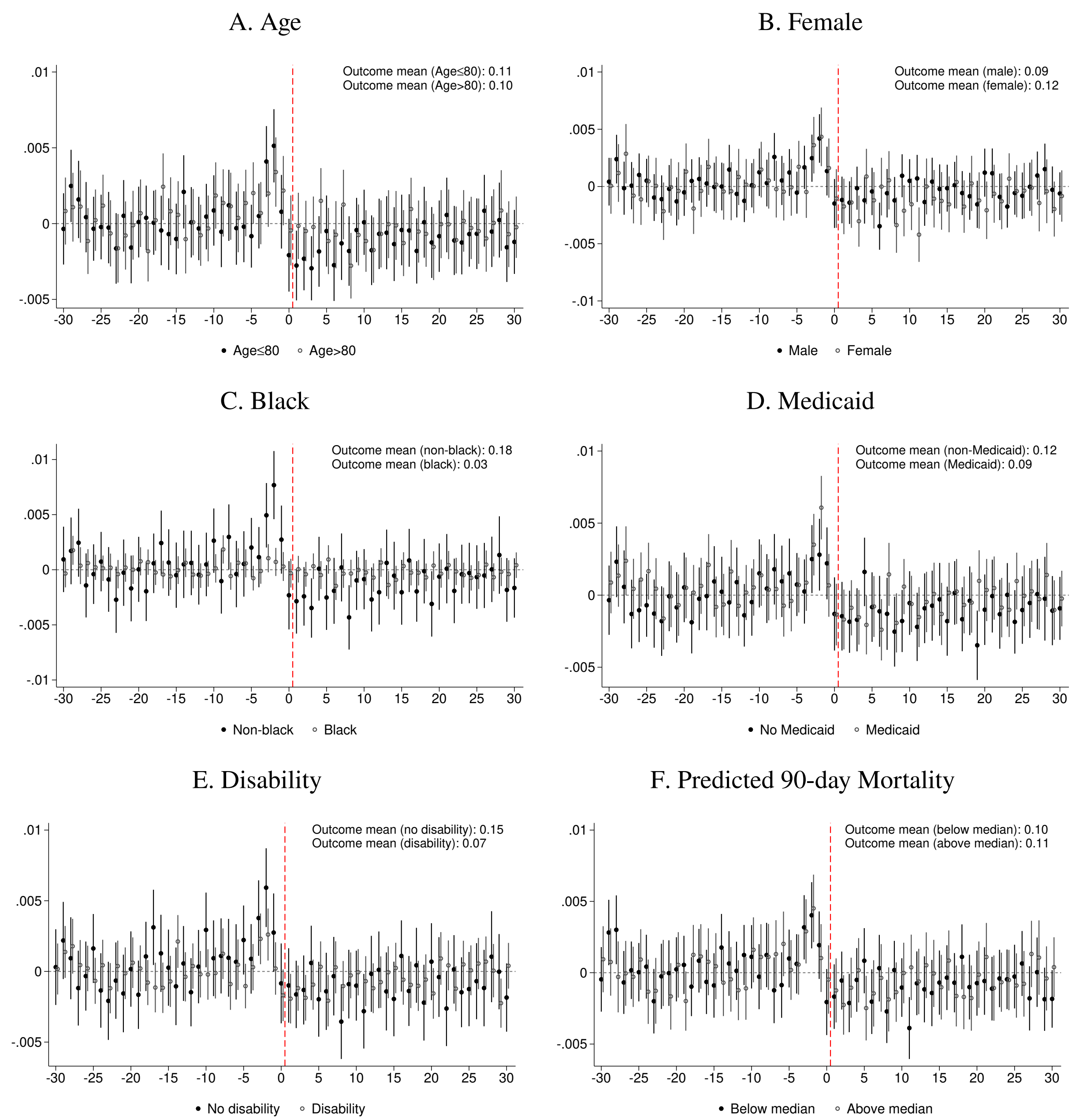



B. Return in Three Weeks



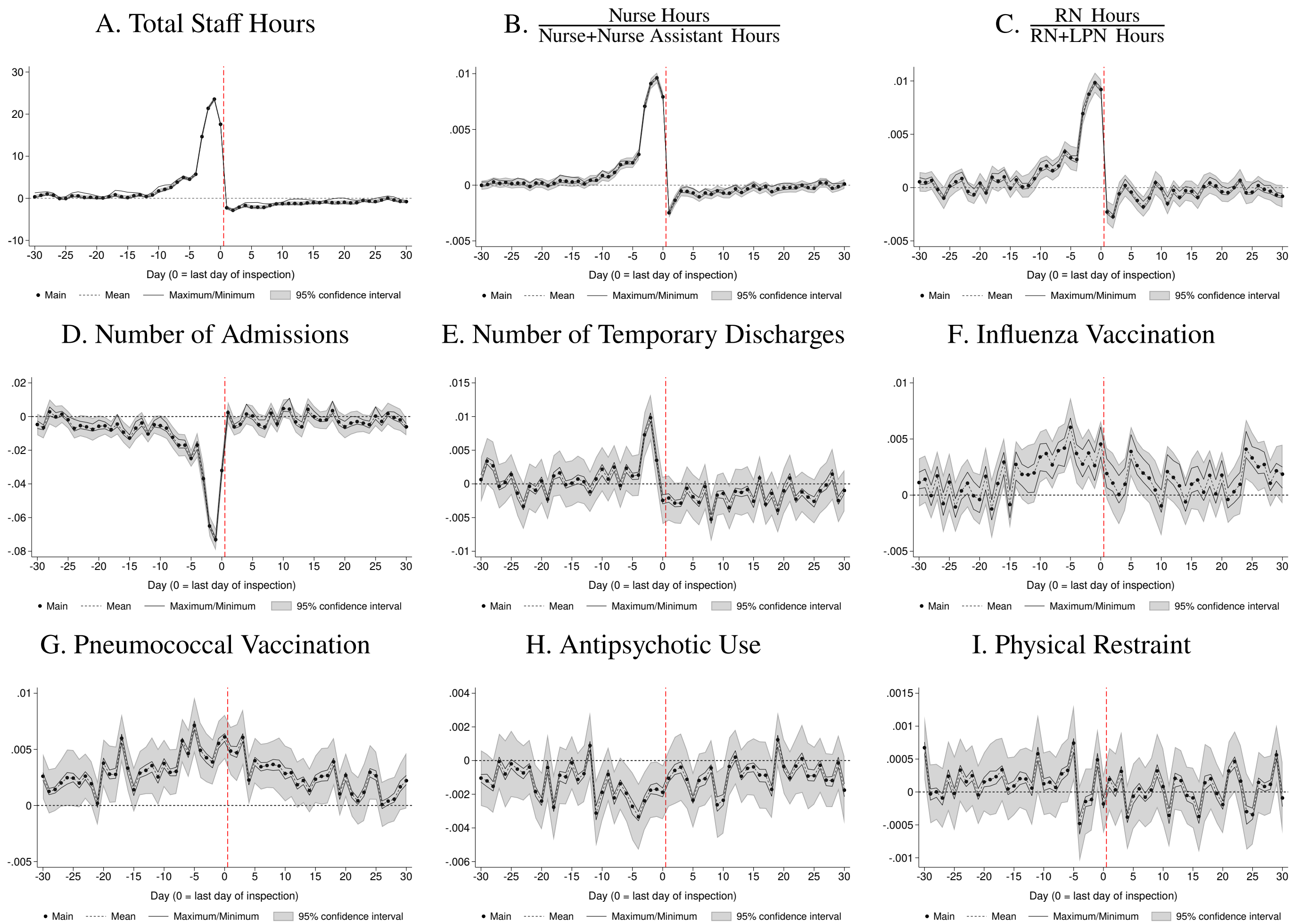

*Notes*: This figure reports changes in the number of temporary discharges in response to health inspections, using alternative definitions of temporary discharges: patients who were discharged and returned to a nursing home within one week (Panel A) and three weeks (Panel B). Each panel presents estimates and associated 95 percent confidence intervals from a separate regression of the examined outcome on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure A.8: Responses to Health Inspections: Temporary Discharges by Patient Characteristics



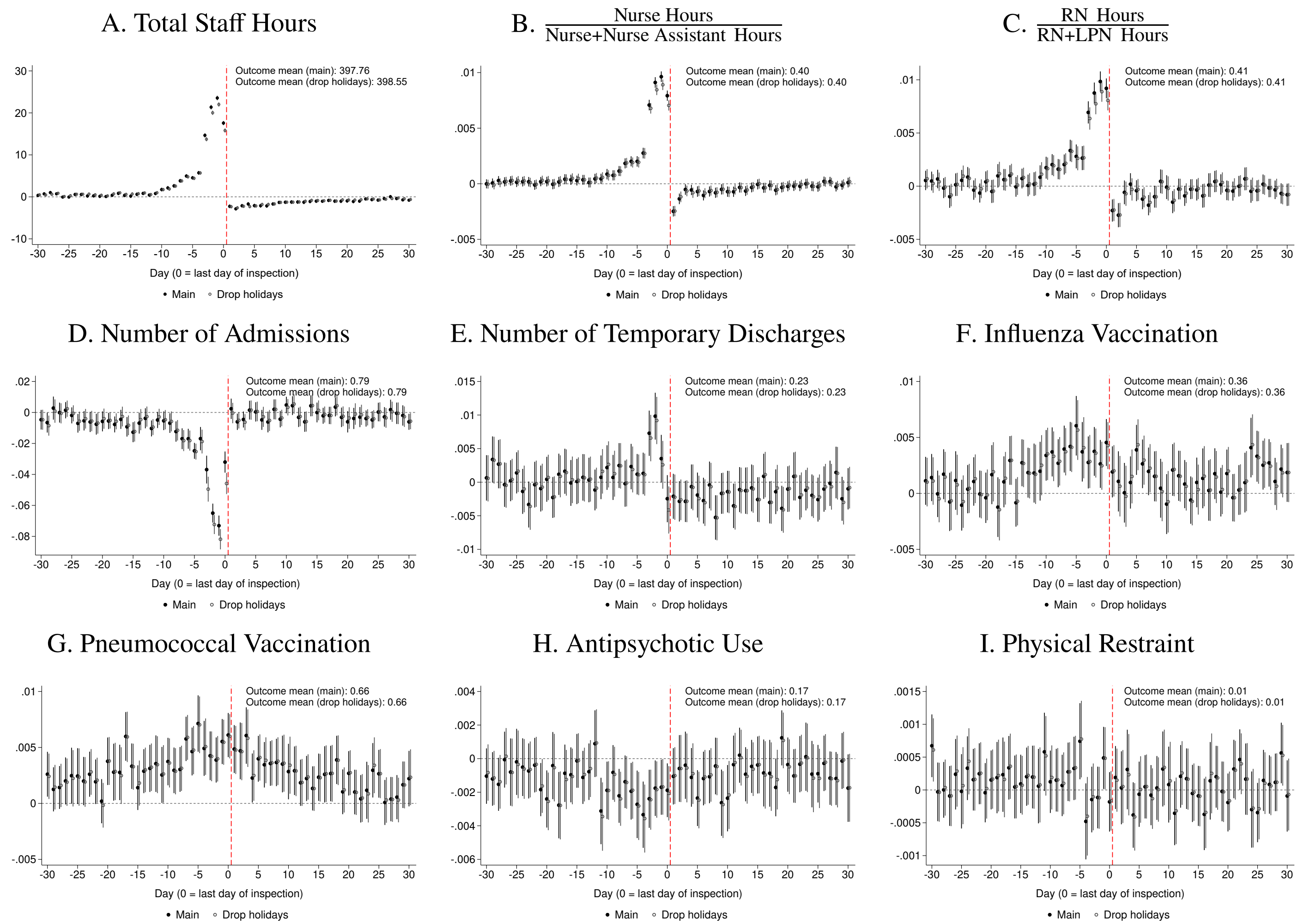

*Notes*: This figure reports changes in the number of temporary discharges in response to health inspections by patient characteristics. Panels A–F decompose temporary discharges by patient age ($\leq 80$ versus $> 80$), gender, race, Medicaid coverage in the prior year, disability status, and predicted 90-day mortality (below versus above median), respectively. The sum of mean temporary discharges in each panel is slightly below the mean shown in Panel D of Figure 3 due to a small share of missing values in the studied patient characteristic. Each panel presents estimates and associated 95 percent confidence intervals from a separate regression of the examined temporary discharges on indicators for days relative to the last day of the inspection, conditional on the baseline controls (nursing-home-by-year, nursing-home-by-month, and nursing-home-by-day-of-the-week indicators). Standard errors are clustered by nursing home. The unit of observation is at the nursing-home-by-day level. The sample includes observations between -100 and 100 days from the last day of each inspection, with days [-100, -30) and (30, 100] set as the reference group. To avoid overlap, we exclude a small share (0.5 percent) of inspections that have another inspection within 150 days.

Figure A.9: Responses to Health Inspections: Robustness to Excluding Inspections Conducted in Any Week of the Year



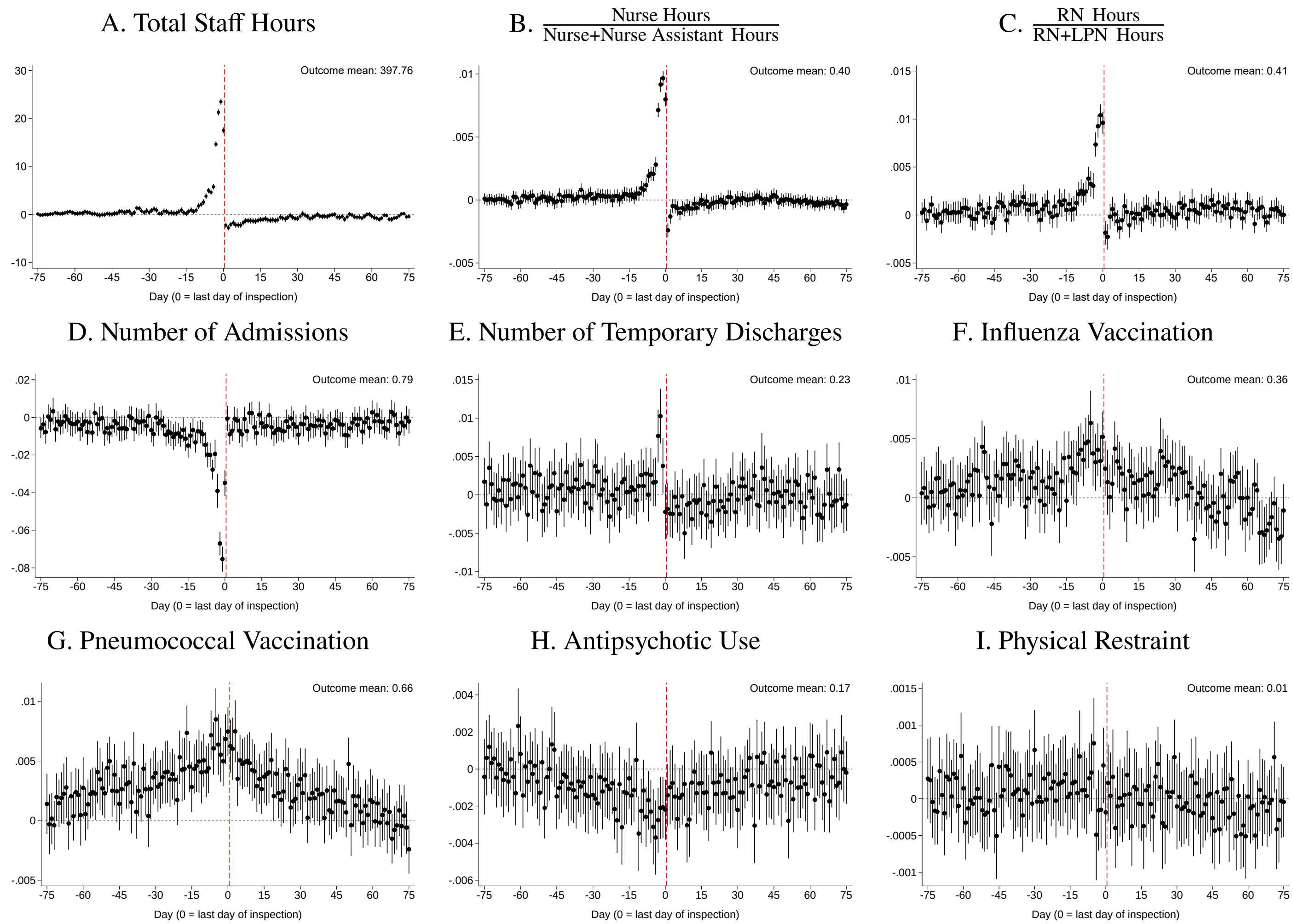

*Notes*: This figure shows the robustness of our estimated responses to health inspections to excluding inspections conducted in any week of the year. To construct the figure, we first run separate regressions that exclude inspections conducted in each week of the year, using Equation (1). We then connect the maximum and minimum of the point estimates from all regressions in the solid lines and the mean of the point estimates in the dotted lines. The dots and gray areas plot, respectively, point estimates and associated 95 percent confidence intervals from our main estimation that includes all inspections. Panels A–I report results for total staff hours, the share of nursing staff hours provided by nurses, the share of nurse hours provided by RNs, the number of admissions, the number of temporary discharges, influenza vaccinations, pneumococcal vaccinations, antipsychotic use, and physical restraint use, respectively.

Figure A.10: Responses to Health Inspections: Robustness to Excluding Holidays



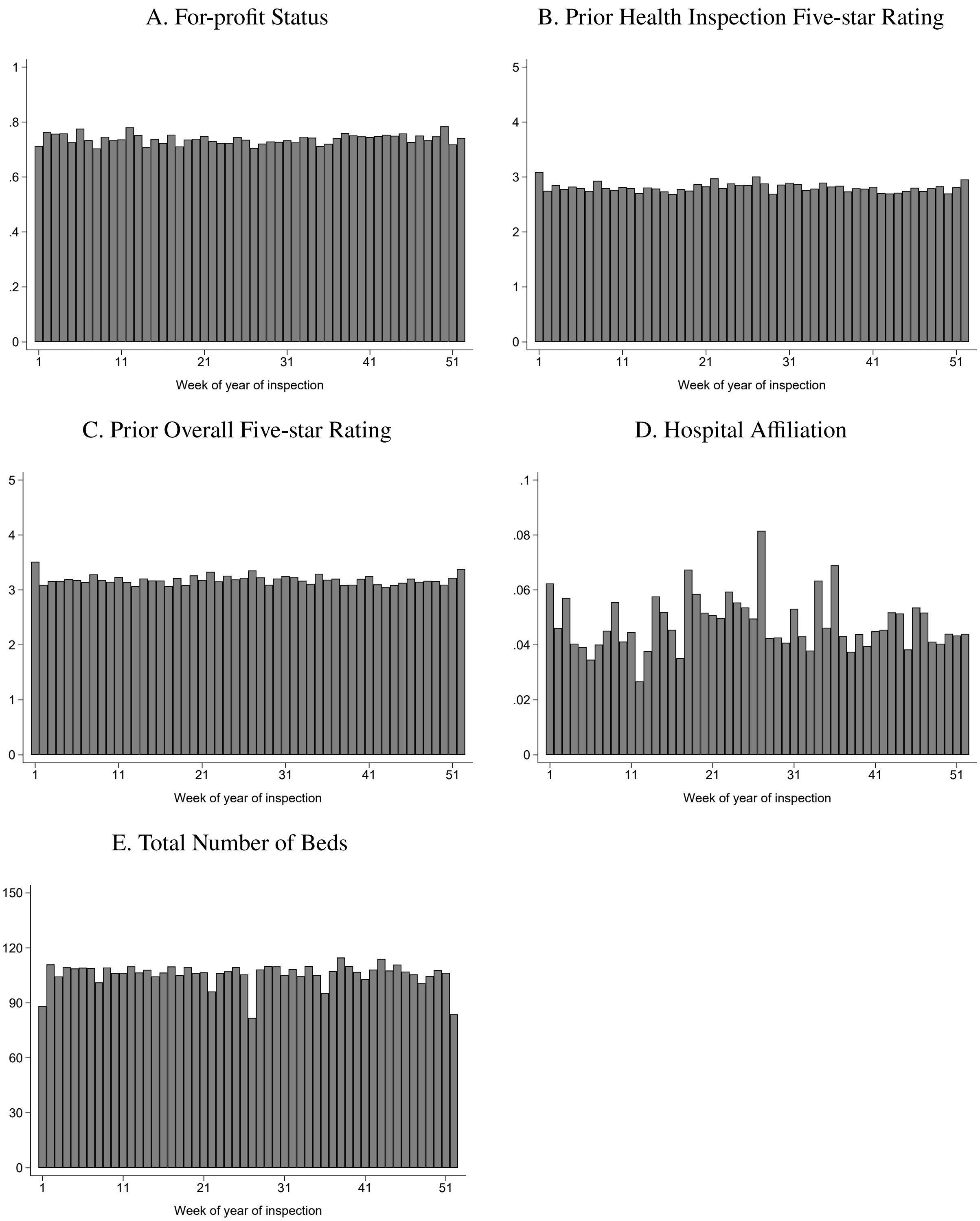

*Notes*: This figure shows the robustness of our estimated responses to health inspections to excluding holidays from the analysis sample. The black dots and lines report, respectively, estimates and associated 95 percent confidence intervals based on our main sample. The gray dots and lines report, respectively, estimates and associated 95 percent confidence intervals based on the sample that excludes holidays. Panels A–I report results for total staff hours, the share of nursing staff hours provided by nurses, the share of nurse hours provided by RNs, the number of admissions, the number of temporary discharges, influenza vaccinations, pneumococcal vaccinations, antipsychotic use, and physical restraint use, respectively.

Figure A.11: Responses to Health Inspections: Reporting Additional Days



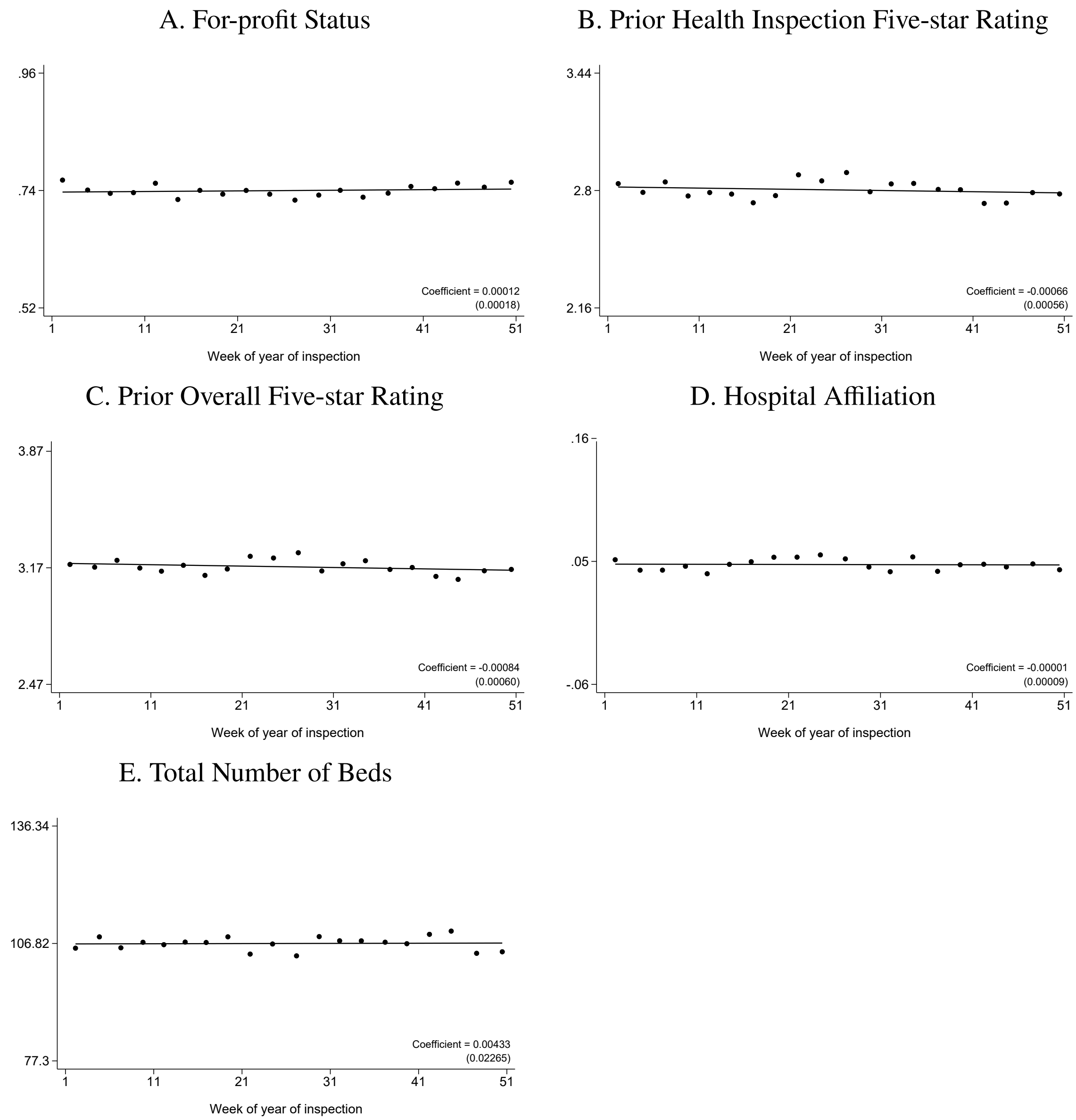

*Notes*: This figure shows the robustness of our estimated responses to health inspections to extending the reporting window from 30 to 75 days before and after the inspection ends. Panels A–I report results for total staff hours, the share of nursing staff hours provided by nurses, the share of nurse hours provided by RNs, the number of admissions, the number of temporary discharges, influenza vaccinations, pneumococcal vaccinations, antipsychotic use, and physical restraint use, respectively.

Figure A.12: Inspection Timing and Nursing Home Characteristics



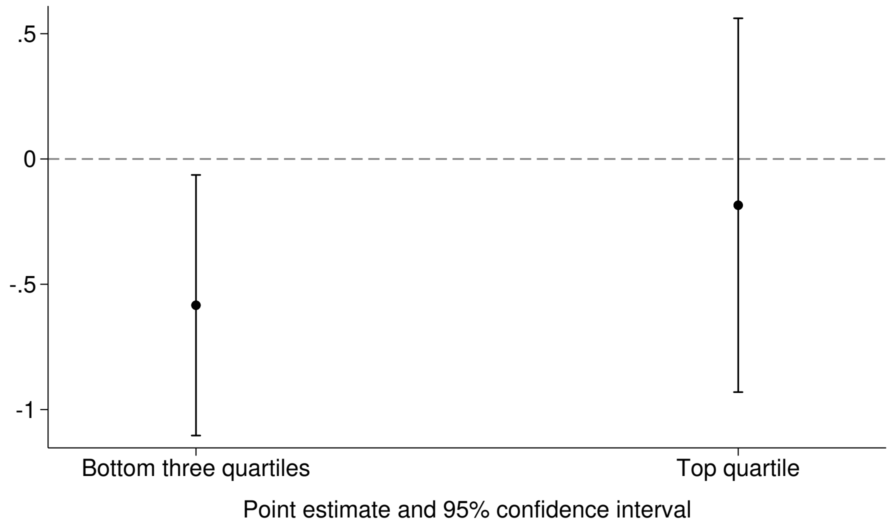

*Notes*: This figure shows mean characteristics of nursing homes versus the week of the year of the inspection. Panels A–E report results for for-profit status, five-star rating from the prior health inspection, overall five-star rating in the prior cycle, hospital affiliation, and total number of beds, respectively.

Figure A.13: Inspection Timing and Nursing Home Characteristics (Binned Scatter Plots)



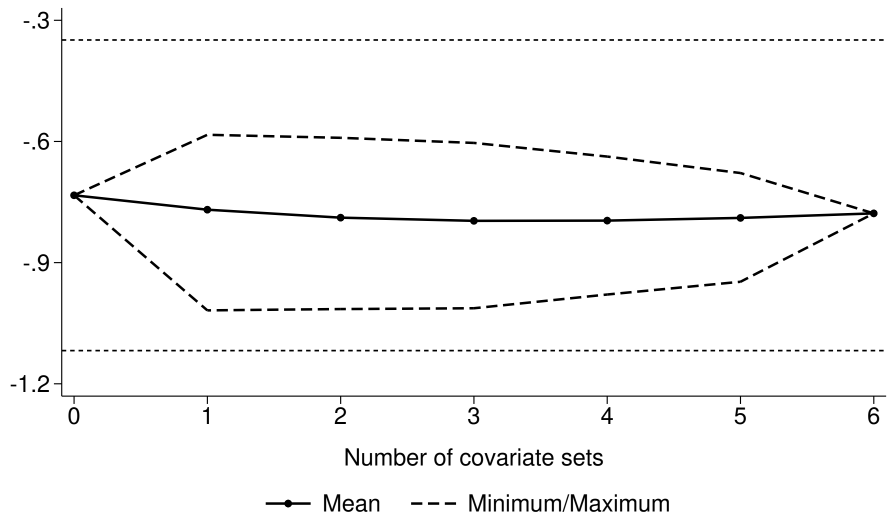

*Notes*: This figure shows binned scatter plots of nursing home characteristics versus the week of the year of the inspection, conditional on state-by-year fixed effects. To construct these binned scatter plots, we first residualize both the *y*-axis and *x*-axis variables with respect to state-by-year indicators and then add means back. The coefficients report the estimated slope of the best-fit line between the *y*-axis and *x*-axis variables (conditional on state-by-year fixed effects), with standard errors clustered by nursing home reported in parentheses. Panels A–E report results for for-profit status, five-star rating from the prior health inspection, overall five-star rating in the prior cycle, hospital affiliation, and total number of beds, respectively.

Figure A.14: Relationship between Quality Rating and Quality: Heterogeneity by Response to Health Inspections



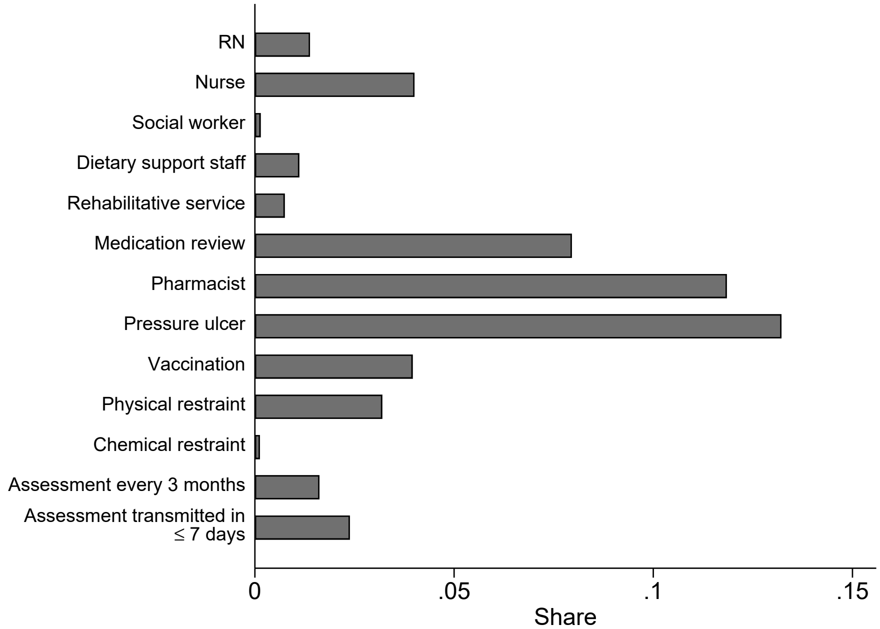

*Notes*: This figure reports heterogeneity in IV estimates of the relationship between admission to high-quality-rated nursing homes (above-median) and patient 90-day mortality. The bottom three quartiles and top quartile refer to nursing homes whose responses in total staff hours are in the bottom three quartiles and top quartile, respectively.

Figure A.15: Relationship between Quality Rating and Quality: Stability of IV Estimates with Varying Patient Controls




*Notes*: This figure shows the robustness of our IV estimate of the relationship between admission to high-quality-rated (above-median) nursing homes and patient 90-day mortality to the inclusion of different sets of patient controls. We divide patient characteristics into six subsets: (i) five-year age-bin indicators; (ii) gender; (iii) black race; (iv) Medicaid coverage in the prior year; (v) disability status; and (vi) indicators for 27 chronic conditions at the start of the year of the nursing home admission. We then run separate regressions that control for each of the $2^6 = 64$ different combinations of patient covariates. Each $n$ on the $x$-axis represents the number of covariate subsets included. For each $n$, the figure reports the mean, minimum, and maximum of the estimated IV coefficients using all possible combinations with $n$ (out of six) subsets of patient covariates. The connected line shows the mean of the estimated IV coefficients. The dashed lines connect the minimum and maximum of the estimated IV coefficients. The dotted lines show the 95 percent confidence interval of the IV estimate with only the baseline controls (i.e., county-by-year indicators), with the standard error clustered by patient zip code.

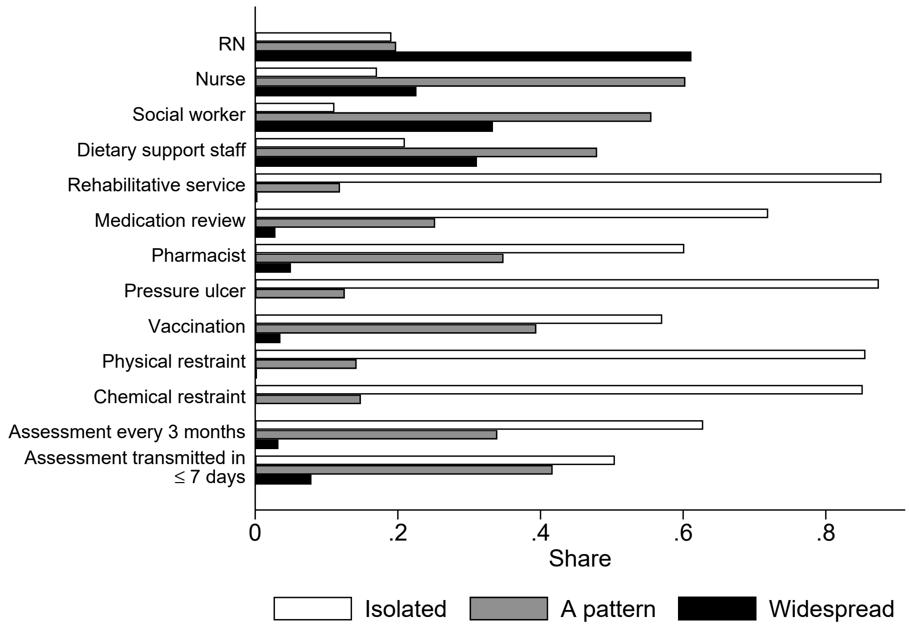
Figure A.16: Share of Health Inspections with Each Quality Deficiency Citation



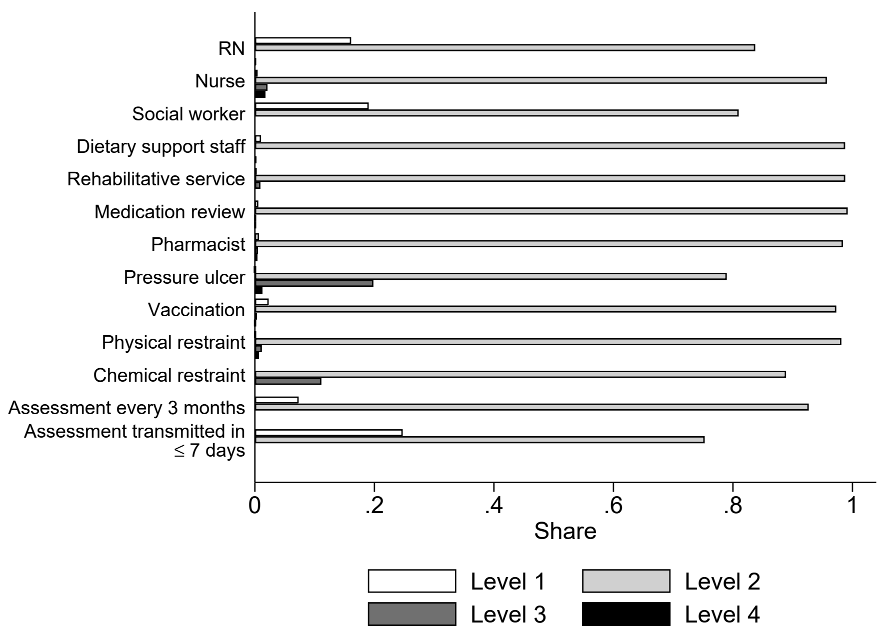

*Notes*: This figure reports the share of health inspections with each citation listed on the *y*-axis. Table 3 describes each citation in detail.

Figure A.17: Scope and Severity of Quality Deficiency Citations

A. Scope



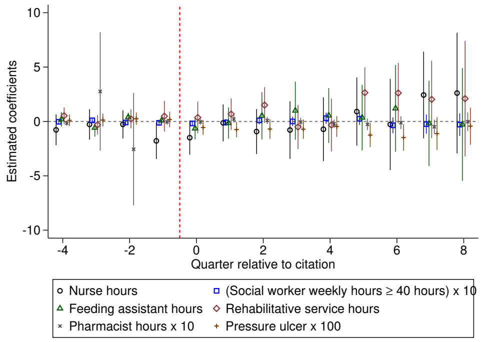

B. Severity



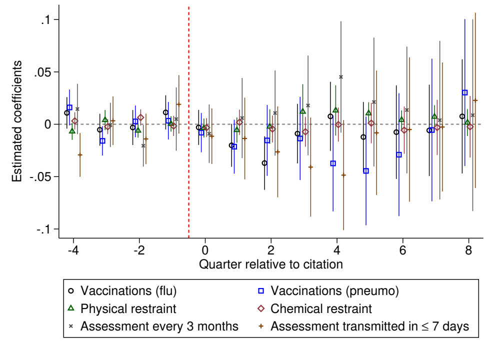

*Notes*: This figure reports the scope and severity levels of each citation listed on the *y*-axis. Table 3 describes each citation in detail.

Figure A.18: Effect of Quality Deficiency Citations on Other Quality Dimensions

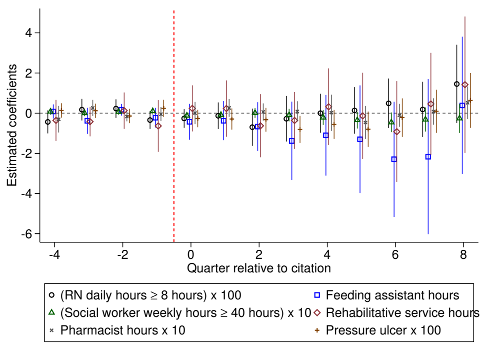
A. RN



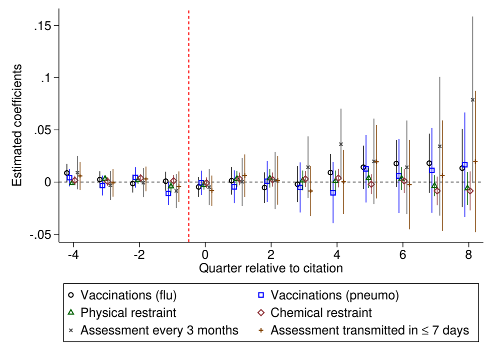

B. RN



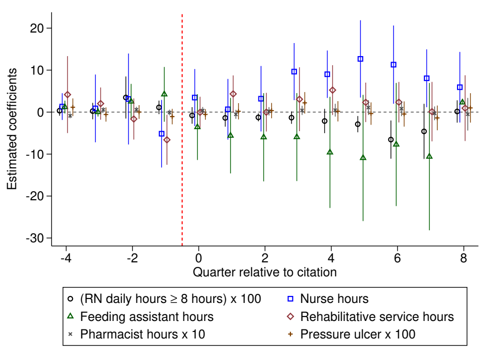

C. Nurse



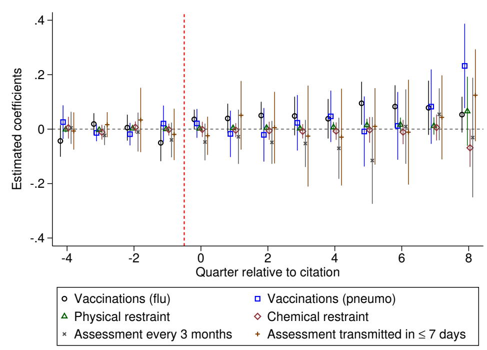

D. Nurse



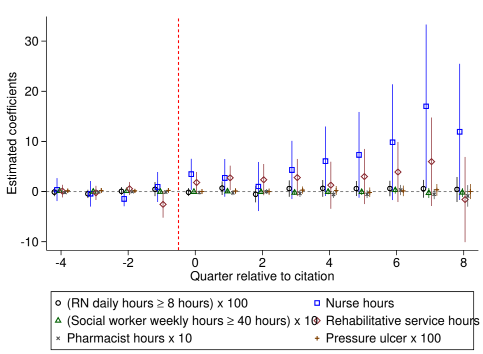

E. Social Worker




F. Social Worker




G. Dietary Support Staff




H. Dietary Support Staff

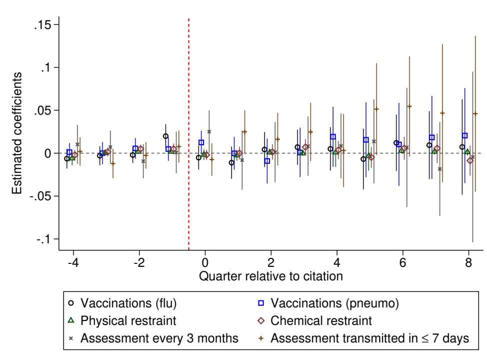
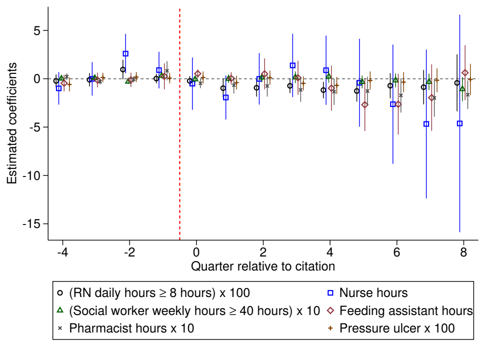
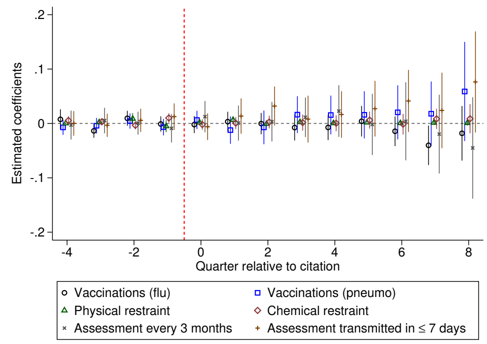


Figure A.18: Effect of Quality Deficiency Citations on Other Quality Dimensions (Continued)

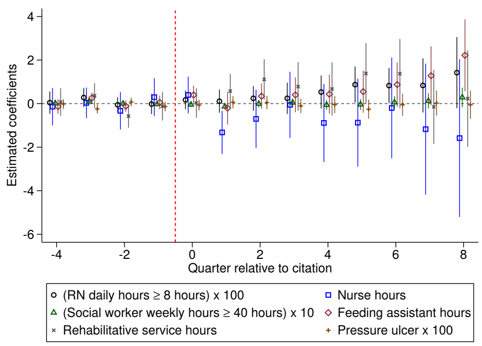
I. Rehabilitative Service



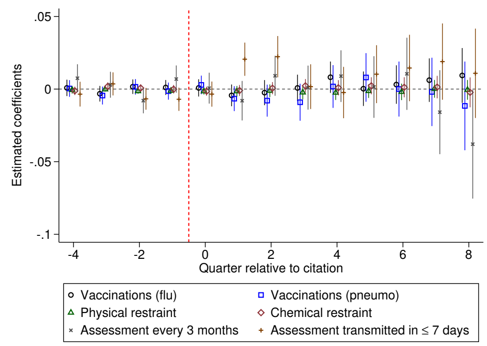

J. Rehabilitative Service



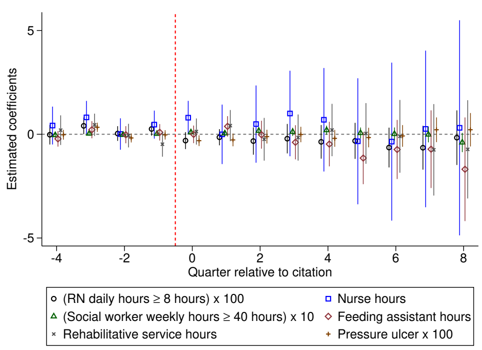

K. Medication Review



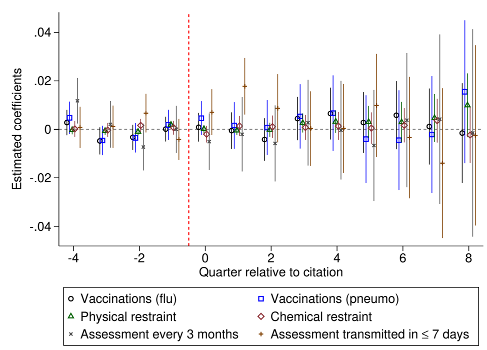

L. Medication Review



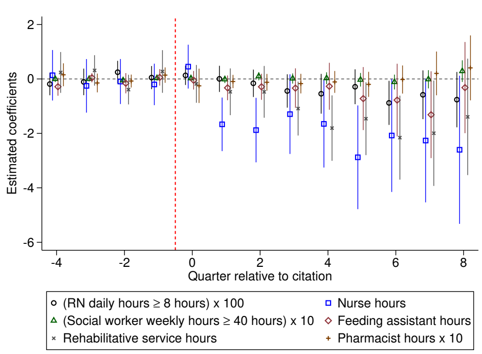

M. Pharmacist




N. Pharmacist




O. Pressure Ulcer




P. Pressure Ulcer

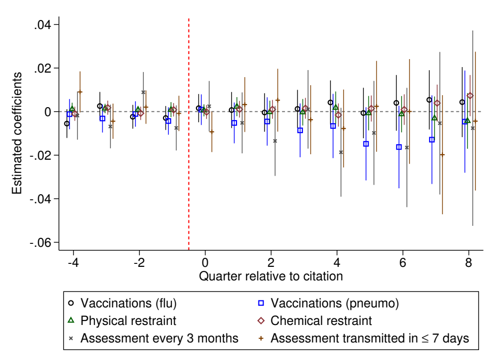
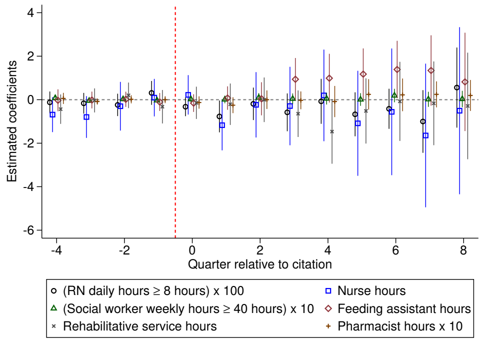
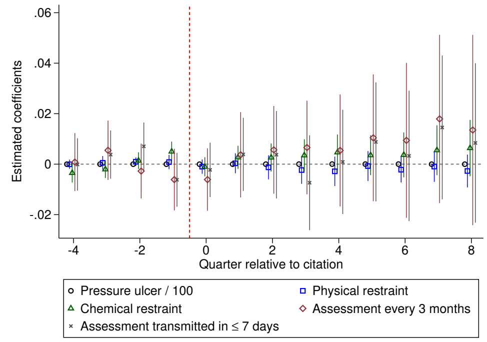


Figure A.18: Effect of Quality Deficiency Citations on Other Quality Dimensions (Continued)

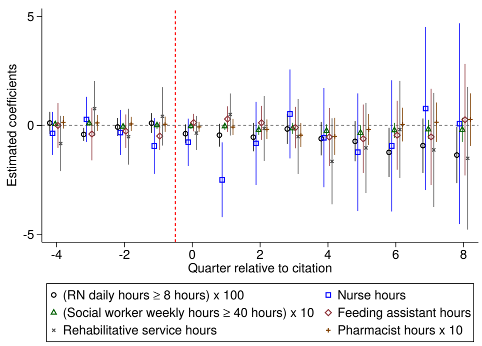
Q. Vaccination



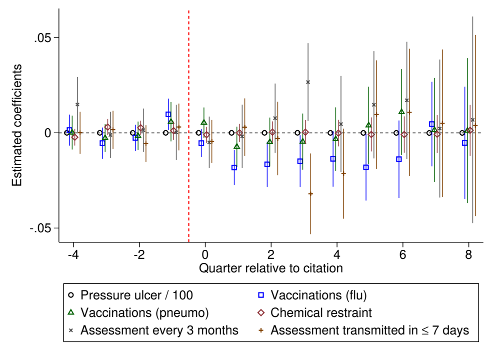

R. Vaccination



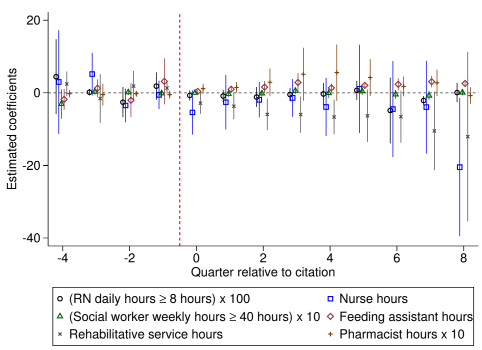

S. Physical Restraint



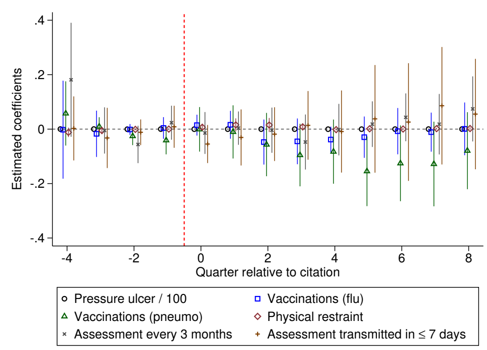

T. Physical Restraint



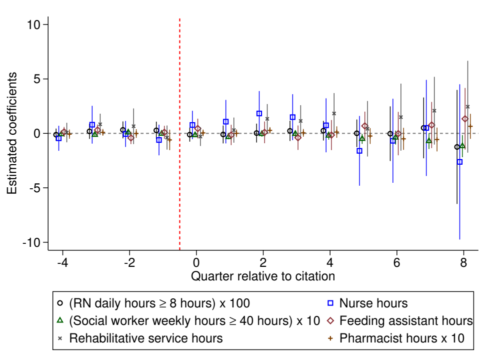

U. Chemical Restraint




V. Chemical Restraint




W. Assessment Every 3 Months




X. Assessment Every 3 Months

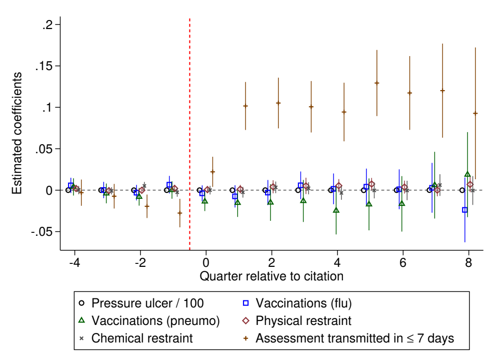
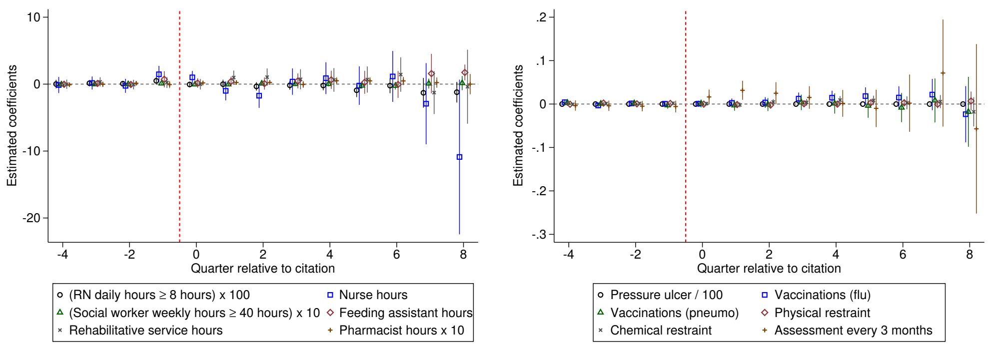


Figure A.18: Effect of Quality Deficiency Citations on Other Quality Dimensions (Continued)

Y. Assessment Transmitted in ≤ 7 days Z. Assessment Transmitted in ≤ 7 days




*Notes*: This figure overlays event-study estimates (along with 95 percent confidence intervals) of the effect of receiving a citation on outcomes not directly related to the citation. Since results are qualitatively similar when using various estimators in Figure 8, for readability this figure reports results using one estimator: Callaway and Sant'Anna (2021). Each panel shows estimates of the effect of the citation described in the panel title on outcomes described in the legend. For readability, some estimates (and their confidence intervals) are scaled up and down by, e.g., 100, as indicated by "× 100" and "/100" in the legend, respectively. Table 3 provides details of each citation and outcome. The unit of observation is at the nursing-home-by-quarter level. Standard errors are clustered by nursing home.

Figure A.19: Effect of Quality Deficiency Citations on Nearby Facilities

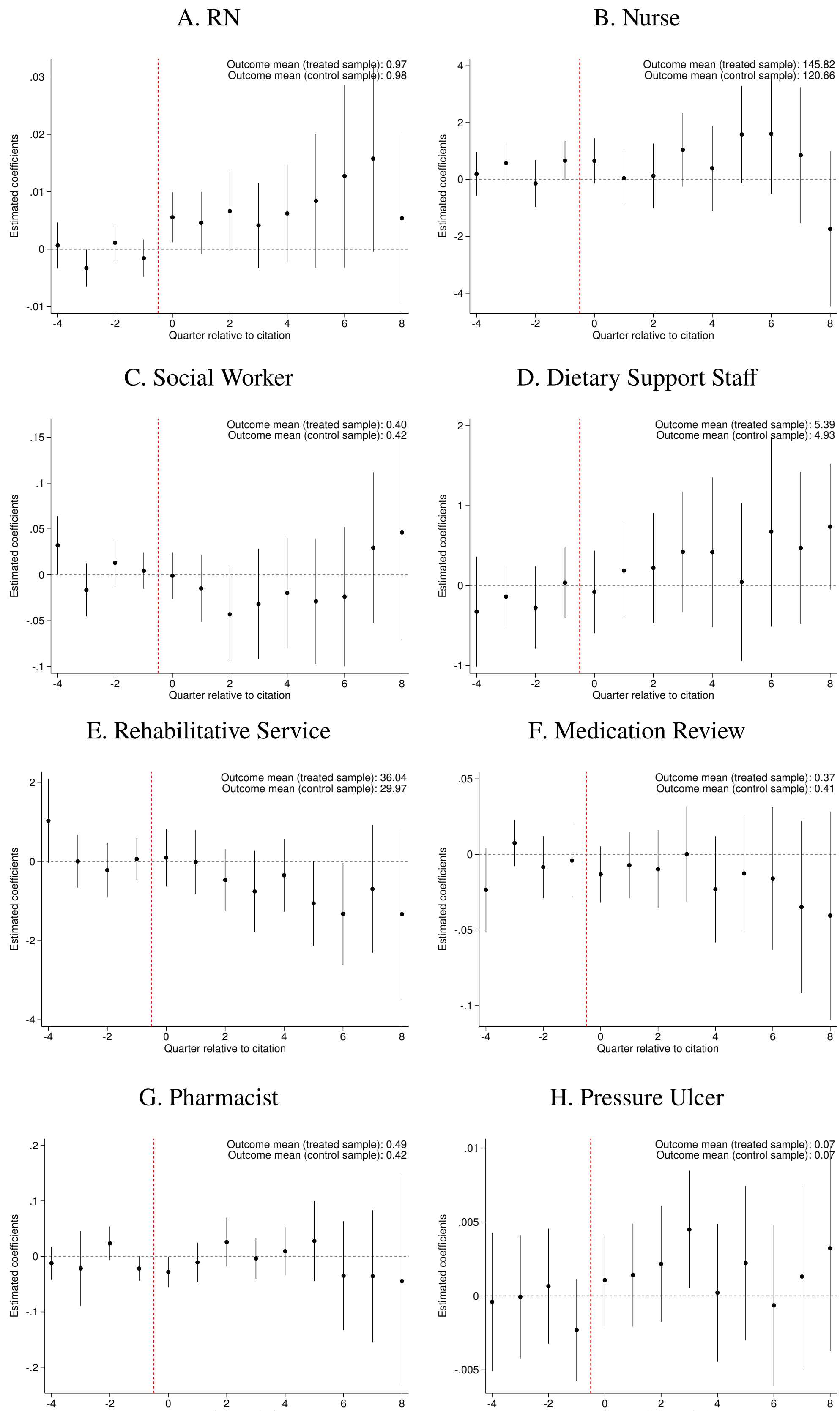
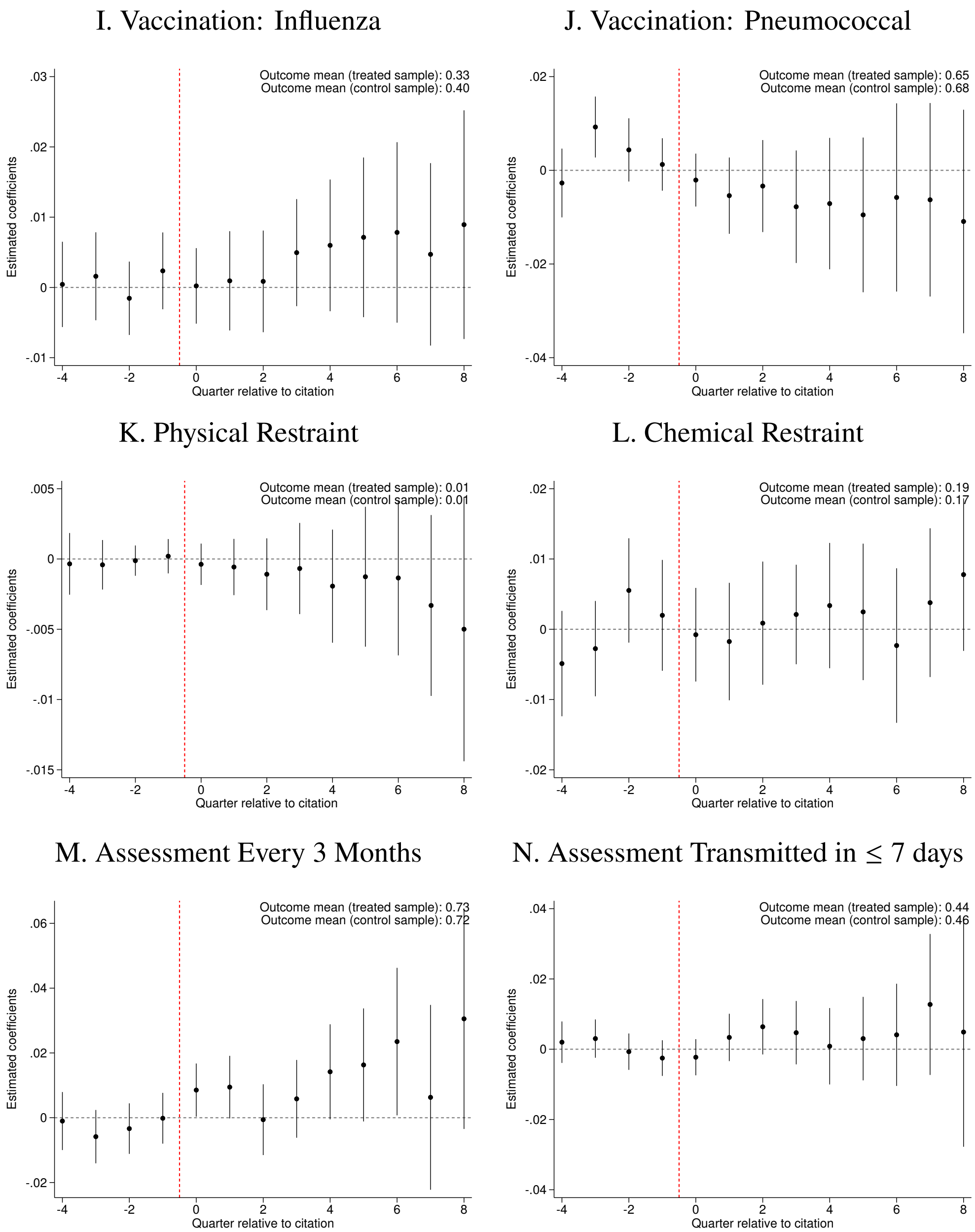


Figure A.19: Effect of Quality Deficiency Citations on Nearby Facilities (Continued)




*Notes*: This figure displays event-study estimates (along with 95 percent confidence intervals) of the effect of a citation on nearby nursing homes (within five miles of cited nursing homes). Since results are qualitatively similar when using various estimators in Figure 8, this figure reports results using one estimator: Callaway and Sant'Anna (2021). Each panel shows estimates of the effect of the citation described in the panel title on the corresponding outcome described in Table 3. The unit of observation is at the nursing-home-by-quarter level. Standard errors are clustered by nursing home.

Figure A.20: Effect of Quality Deficiency Citations on the Cited Quality Dimension: Robustness to Excluding Days around Health Inspections, Nearby Facilities, and Facilities with Repeated Citations

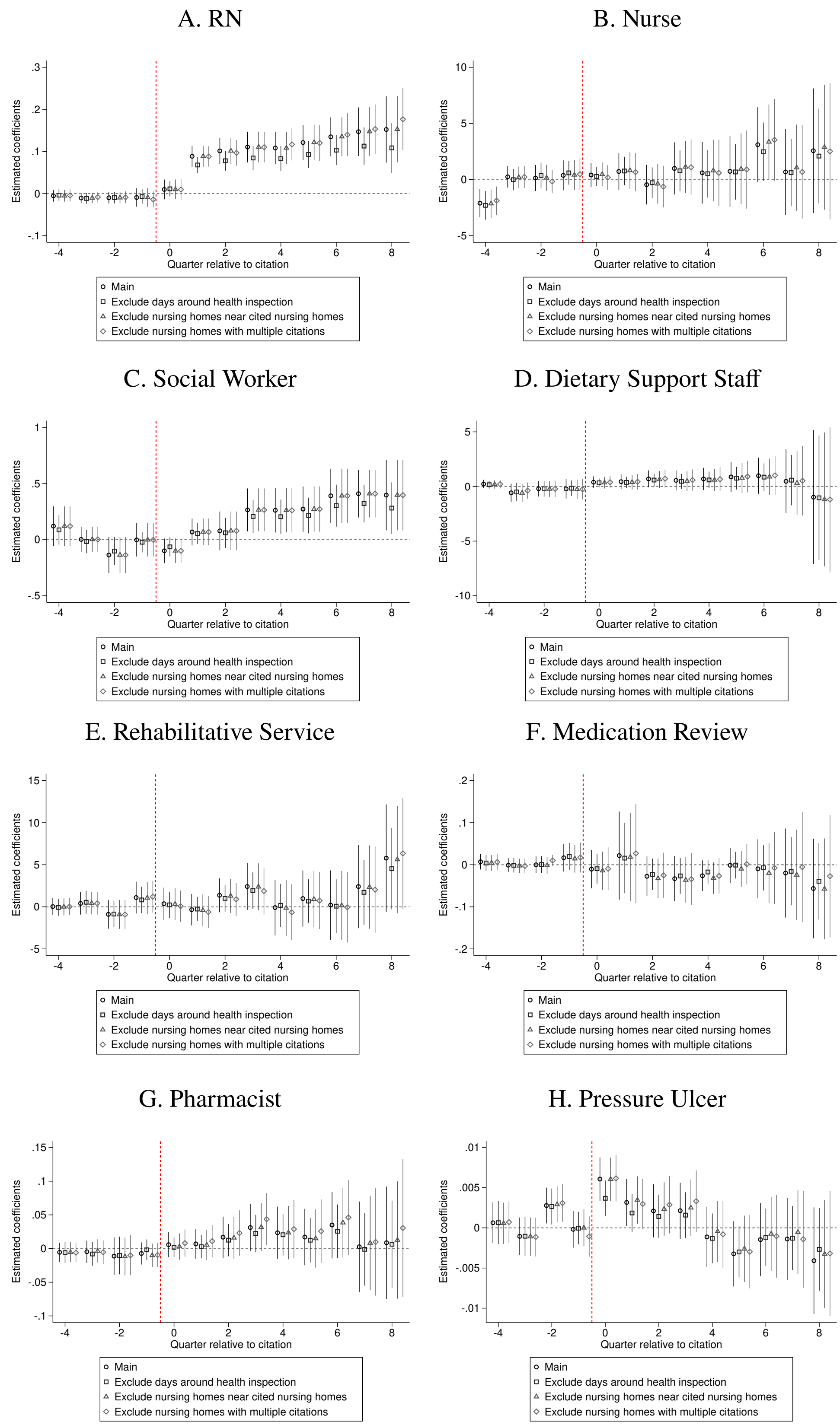
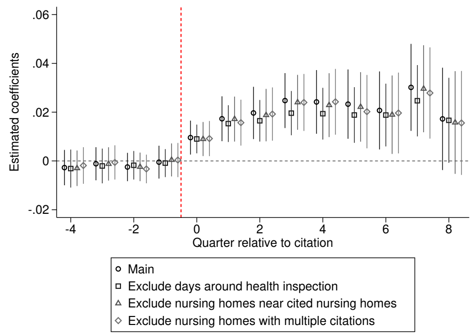
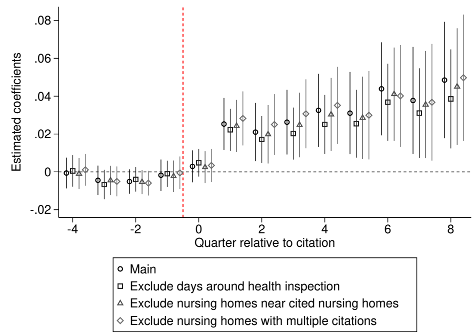


Figure A.20: Effect of Quality Deficiency Citations on the Cited Quality Dimension: Robustness to Excluding Days around Health Inspections, Nearby Facilities, and Facilities with Repeated Citations (Continued)

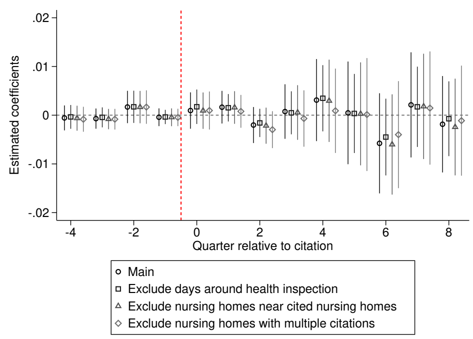
I. Vaccination: Influenza



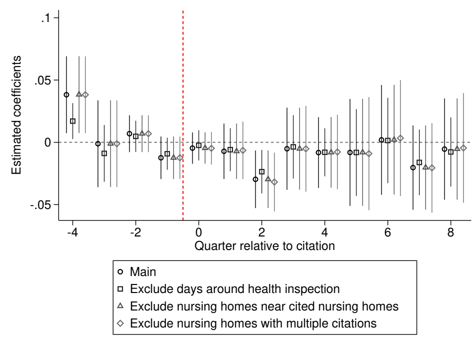

J. Vaccination: Pneumococcal



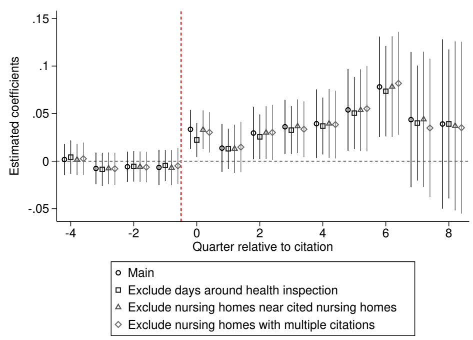

K. Physical Restraint



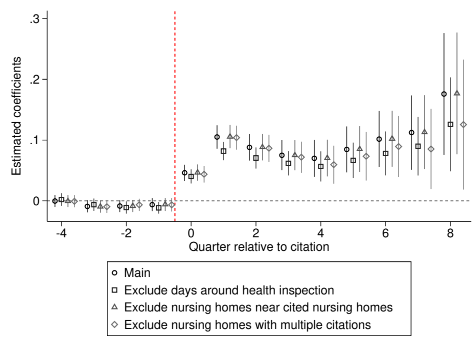

L. Chemical Restraint




M. Assessment Every 3 Months




N. Assessment Transmitted in ≤ 7 days




*Notes*: Each panel shows estimates (along with 95 percent confidence intervals) of the effect of the citation described in the panel title on the corresponding outcome described in Table 3. Since results are qualitatively similar when using various estimators in Figure 8, for readability this figure reports results using one estimator: Callaway and Sant'Anna (2021). The circles, squares, triangles, and diamonds show, respectively, our main estimates reported in Figure 8, the estimates excluding from the analysis sample days around health inspections (from two weeks before to two weeks after an inspection), the estimates excluding from the control group nursing homes near cited nursing homes (within five miles), and the estimates excluding from the treatment group nursing homes receiving the same citation again in our study period. The unit of observation is at the nursing-home-by-quarter level. Standard errors are clustered by nursing home.

Figure A.21: Effect of Quality Deficiency Citations on the Cited Quality Dimension: Robustness to Patient and Nursing Home Covariates

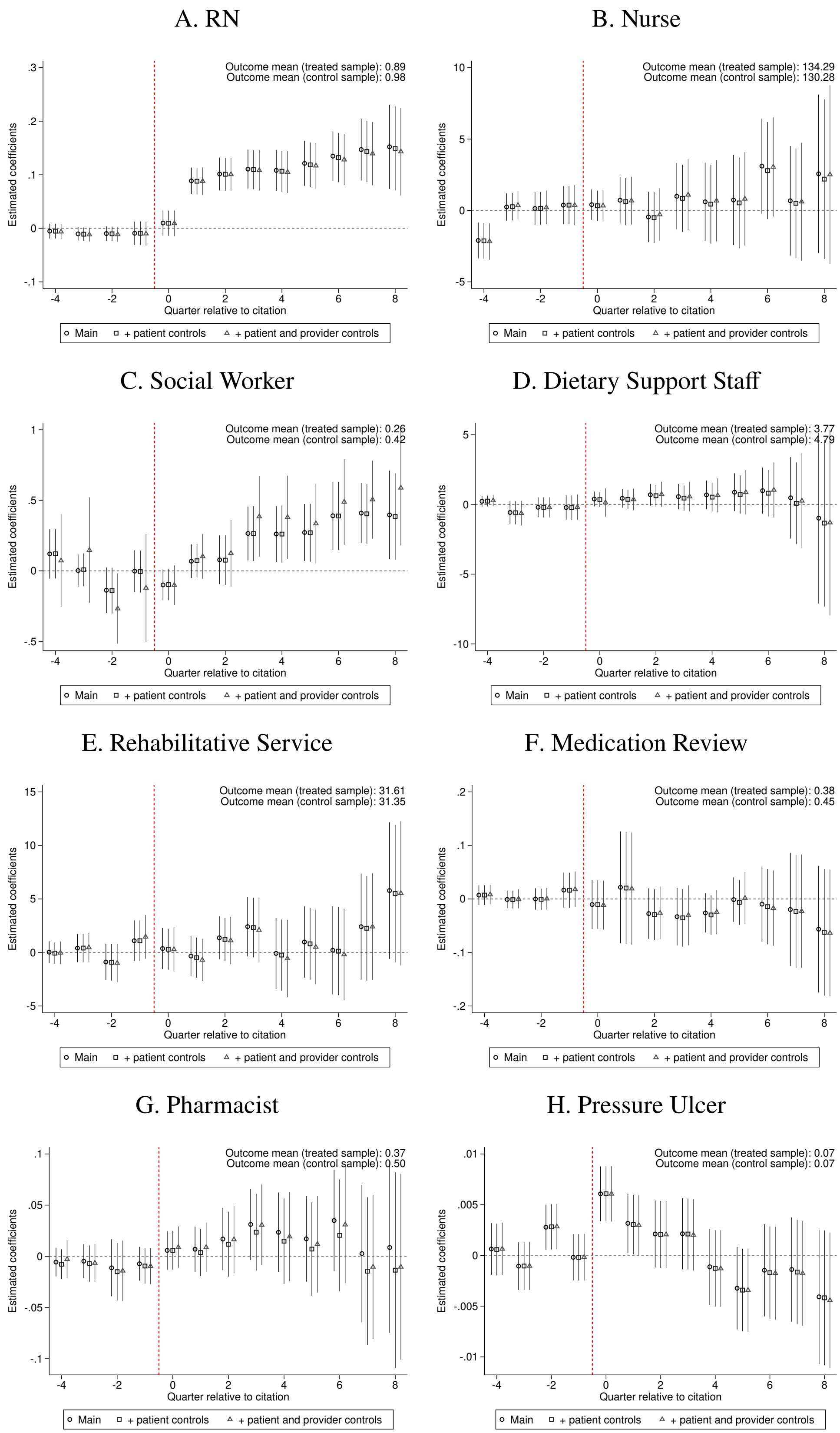
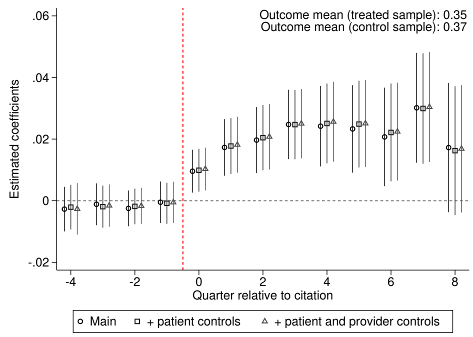
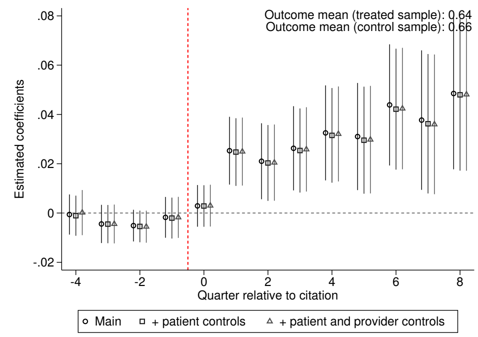


Figure A.21: Effect of Quality Deficiency Citations on the Cited Quality Dimension: Robustness to Patient and Nursing Home Covariates (Continued)

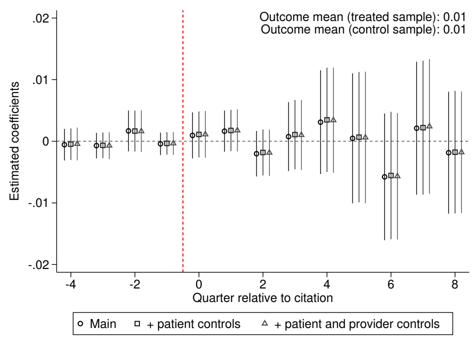
I. Vaccination: Influenza



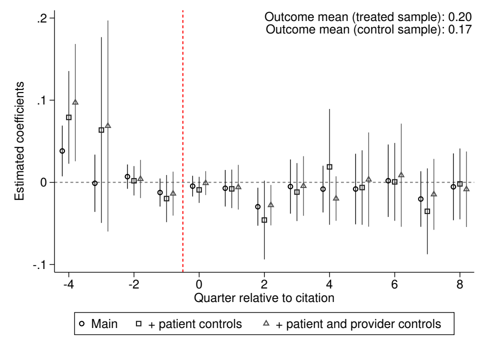

J. Vaccination: Pneumococcal



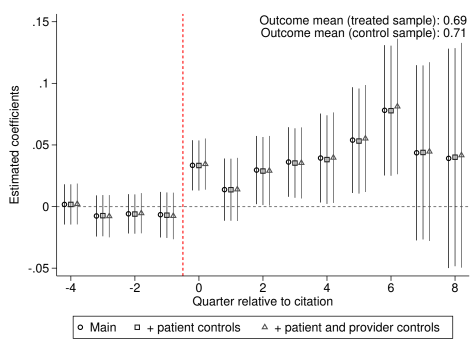

K. Physical Restraint



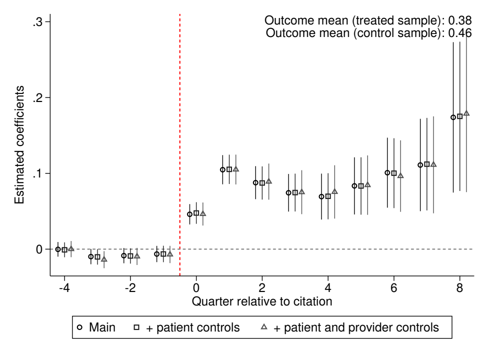

L. Chemical Restraint




M. Assessment Every 3 Months




N. Assessment Transmitted in $\leq$ 7 days




*Notes*: This figure shows the robustness of the estimated effect of citations to including patient characteristics and nursing home time-varying characteristics as controls. Since results are qualitatively similar when using various estimators in Figure 8, for readability this figure reports results using one estimator: Callaway and Sant'Anna (2021). The circles, squares, and triangles show, respectively, the main estimates in Figure 8, the estimates controlling for patient characteristics, and the estimates controlling for patient and nursing home time-varying characteristics. Patient characteristics include age, gender, black race, disability status, Medicaid coverage in the prior year, and the number of 27 chronic conditions reported in the Chronic Conditions file of Medicare claims at the start of the year of the nursing home admission, collapsed at the nursing-home-by-quarter level. Nursing home time-varying characteristics include an indicator for for-profit status, an indicator for hospital affiliation, the number of beds, and average daily patient census. Each panel shows estimates and 95 percent confidence intervals of the effect of the citation described in the panel title on the corresponding outcome described in Table 3. The unit of observation is at the nursing-home-by-quarter level. Standard errors are clustered by nursing home.

Table A.1: Relationship between Quality Rating and Quality: Characteristics of Analysis Sample

| | | By nursing home quality rating | | |
|---|---|---|---|---|
| | All | Low | High | $p$-value |
| Age | 81.02 | 80.79 | 81.21 | <0.01 |
| | [8.31] | [8.35] | [8.28] | |
| Female | 0.61 | 0.60 | 0.61 | <0.01 |
| | [0.49] | [0.49] | [0.49] | |
| Black | 0.09 | 0.11 | 0.08 | <0.01 |
| | [0.29] | [0.31] | [0.28] | |
| Medicaid coverage (prior year) | 0.18 | 0.20 | 0.16 | <0.01 |
| | [0.38] | [0.40] | [0.37] | |
| Number of comorbidities | 5.34 | 5.36 | 5.32 | <0.01 |
| | [2.94] | [2.97] | [2.91] | |
| Disability | 0.14 | 0.15 | 0.13 | <0.01 |
| | [0.34] | [0.36] | [0.33] | |
| 90-day mortality | 15.08 | 15.99 | 14.37 | <0.01 |
| | [35.79] | [36.65] | [35.08] | |
| Observations | 5,578,510 | 2,439,441 | 3,139,069 | |

*Notes*: Column 1 shows average characteristics of all patients in our main IV analysis of the relationship between admission to high-quality-rated (above-median) nursing homes and patient 90-day mortality. Columns 2 and 3 show average characteristics of patients admitted to low-quality-rated (below-median) and high-quality-rated (above-median) nursing homes, respectively. Standard deviations are reported in brackets; $p$-values of $t$-tests for the equivalence of means between patients in Columns 2 and 3 are shown in the last column.

Table A.2: Relationship between Quality Rating and Quality: First Stage (Split Sample)

| | Observations | Percent admitted to high-quality-rated | First-stage estimate |
|---|---|---|---|
| Age >80 | 2,954,510 | 57.23 | -1.80 |
| | | | (0.02) |
| Age <= 80 | 2,623,032 | 55.19 | -1.46 |
| | | | (0.01) |
| Female | 3,382,265 | 57.06 | -1.65 |
| | | | (0.01) |
| Male | 2,195,289 | 55.05 | -1.62 |
| | | | (0.02) |
| Black | 520,737 | 49.55 | -1.55 |
| | | | (0.03) |
| Non-black | 5,028,704 | 56.97 | -1.63 |
| | | | (0.01) |
| Medicaid coverage (prior year) | 986,001 | 50.45 | -1.66 |
| | | | (0.02) |
| No Medicaid coverage (prior year) | 4,588,891 | 57.53 | -1.62 |
| | | | (0.02) |
| Disability | 762,284 | 52.38 | -1.51 |
| | | | (0.02) |
| No disability | 4,814,630 | 56.89 | -1.66 |
| | | | (0.01) |
| Predicted mortality above median | 2,788,764 | 55.66 | -1.73 |
| | | | (0.02) |
| Predicted mortality below median | 2,788,756 | 56.88 | -1.54 |
| | | | (0.01) |

*Notes*: This table shows the first-stage estimates for patients of different characteristics. Column 1 reports the number of observations for each subset of patients. The sum of observations in subsamples split by patient characteristics is smaller than the total number of observations shown in Table 2 due to missing values in patient characteristics. Column 2 reports the share of patients in each subset admitted to high-quality-rated nursing homes. The share of patients admitted to high-quality-rated nursing homes is larger than 50 percent, reflecting that high-quality-rated nursing homes on average have higher patient volumes than low-quality-rated nursing homes. Column 3 reports the first-stage estimates for each subset of patients conditional on the baseline control vector (county-by-year indicators), with standard errors clustered by patient zip code reported in parentheses.

Table A.3: Relationship between Quality Rating and Quality: Complier Characteristics

| | All | Compliers | Ratio |
|---|---|---|---|
| Age | 81.02 | 82.30 | 1.02 |
| | | (0.056) | [1.01, 1.02] |
| Female | 0.606 | 0.611 | 1.01 |
| | | (0.002) | [1.00, 1.01] |
| Black | 0.094 | 0.035 | 0.37 |
| | | (0.003) | [0.31, 0.43] |
| Medicaid coverage (prior year) | 0.177 | 0.177 | 1.00 |
| | | (0.003) | [0.96, 1.04] |
| Number of comorbidities | 5.341 | 5.277 | 0.99 |
| | | (0.017) | [0.98, 0.99] |
| Disability | 0.137 | 0.125 | 0.91 |
| | | (0.002) | [0.89, 0.94] |
| Predicted 90-day mortality | 15.08 | 15.55 | 1.03 |
| | | (0.031) | [1.03, 1.04] |

*Notes*: Columns 1 and 2 report average characteristics of the overall sample and compliers, respectively. Standard errors are reported in parentheses. Column 3 reports the ratio between the complier mean and the overall sample mean. 95% confidence intervals of each ratio are shown in brackets. Predicted 90-day mortality is generated from a linear regression of actual 90-day mortality on patient characteristics $\mathbf{X}_i$ included in Equations (2) and (3).

Table A.4: Relationship between Quality Rating and Quality: Robustness to Selection on Unobservables

| | Baseline | Main | Bias-corrected Oster (2019) | Control for local mean Altonji and Mansfield (2018) |
|---|---|---|---|---|
| Panel A: First-stage estimates (unit: percentage points) | | | | |
| Point estimate | -1.6347 | -1.6218 | -1.6000 | -1.6080 |
| Standard error | (0.0144) | (0.0142) | | (0.0140) |
| Panel B: Reduced-form estimates | | | | |
| Point estimate | 0.0120 | 0.0126 | 0.0128 | 0.0086 |
| Standard error | (0.0032) | (0.0031) | | (0.0031) |
| Panel C: IV estimates | | | | |
| Point estimate | -0.7333 | -0.7781 | -0.8000 | -0.5376 |
| Standard error | (0.1962) | (0.1935) | | (0.1926) |

*Notes*: This table shows the robustness of our estimates for the relationship between admission to high-quality-rated nursing homes and patient 90-day mortality to accounting for potential selection on patient unobservables using approaches by Altonji and Mansfield (2018) and Oster (2019). Column 1 reports the baseline estimates with only the baseline control vector for the IV analysis (county-by-year indicators). Column 2 reports the main estimates with the full set of controls included in Equations (2) and (3). Column 3 reports the estimates adjusting for potential selection on patient unobservables using an approach by Oster (2019) (details described in Appendix A.2). Column 4 follows Altonji and Mansfield (2018) and reports the estimates adjusting for potential selection on patient unobservables by controlling for average demographics and comorbidities that include five-year bin indicators for mean age, the shares of patients who are female, black, disabled, and covered by Medicaid in the prior year, and the shares of patients with each of the 27 chronic conditions described in Section 5.1 obtained using other patients living in the index patient's zip code of residence, as well as average socioeconomic characteristics that include per capita income, the share of individuals with a college degree, and unemployment rates at the zip code level obtained from the 2015-2019 American Community Survey. Panels A, B, and C report the first-stage, reduced-form, and IV estimates, respectively. The unit of the outcome variable in the first-stage estimates is percentage points.

Table A.5: Relationship between Quality Rating and Quality: Robustness Checks

| | Main | Exclude overlap | Prior year residence | Std. error clustered by county |
|---|---|---|---|---|
| Panel A: First-stage estimates (unit: percentage points) | | | | |
| Differential distance | -1.6218 | -1.5843 | -1.5282 | -1.6218 |
| | (0.0142) | (0.0146) | (0.01357) | (0.0299) |
| Panel B: Reduced-form estimates | | | | |
| Differential distance | 0.0126 | 0.0132 | 0.0118 | 0.0126 |
| | (0.0031) | (0.0036) | (0.0031) | (0.0037) |
| Panel C: IV estimates | | | | |
| High-quality-rated nursing home | -0.7781 | -0.8350 | -0.7801 | -0.7781 |
| | (0.1935) | (0.2295) | (0.2014) | (0.2238) |
| Full controls | Yes | Yes | Yes | Yes |
| Outcome mean | 15.08 | 16.25 | 15.08 | 15.08 |
| Outcome S.D. | 35.79 | 36.89 | 35.79 | 35.79 |
| Observations | 5,578,510 | 4,507,940 | 5,574,890 | 5,578,510 |

*Notes*: Column 1 replicates the main estimates in Column 4 of Table 2. Column 2 shows the robustness of our estimates to excluding patients whose stays overlap with a health inspection, i.e., those whose stays overlap with the interval from two weeks before to two weeks after the end date of a health inspection. Column 3 shows the robustness of our estimates to defining the instrument based on the patient's prior-year zip code of residence. Columns 1-3 cluster standard errors by patient zip code of residence. Column 4 shows the robustness of our estimates to clustering standard errors by patient county of residence. Panels A, B, and C report the first-stage, reduced-form, and IV estimates, respectively. The unit of the outcome variable in the first-stage estimates is percentage points. The set of full controls is that included in Equations (2) and (3), including county-by-year indicators, five-year age bins, gender, race, disability status, Medicaid coverage in the prior year, and the 27 chronic conditions reported in the Chronic Conditions file of Medicare claims at the start of the year of the nursing home admission. For each patient covariate with missing values, we add an indicator for missing values and replace the missing values with zero.

Table A.6: Effects of Quality Deficiency Citations

| | TWFE | Borusyak et al. (2024) | Sun and Abraham (2021) | Callaway and Sant'Anna (2021) | De Chaisemartin and d'Haultfoeuille (2020) | Synthetic difference-in-differences |
|---|---|---|---|---|---|---|
| RN | 0.0445 | 0.0591 | 0.0494 | 0.0925 | 0.0693 | 0.0541 |
| | (0.0072) | (0.0096) | (0.0026) | (0.0143) | (0.0106) | (0.0171) |
| Nurse | 0.7807 | 0.8503 | 0.7093 | 0.7469 | 0.5065 | -1.0217 |
| | (0.6611) | (0.7368) | (0.4332) | (0.8394) | (0.6350) | (1.2113) |
| Social worker | 0.0478 | 0.0868 | 0.0865 | 0.1782 | 0.1331 | 0.1017 |
| | (0.0492) | (0.0423) | (0.0197) | (0.0726) | (0.0610) | (0.0631) |
| Dietary support staff | 0.2190 | 0.1550 | -0.2143 | 0.5400 | 0.4962 | 0.8045 |
| | (0.4066) | (0.4835) | (0.3715) | (0.5020) | (0.3567) | (0.5009) |
| Rehabilitative service | 0.2934 | 0.9686 | 0.6095 | 1.0007 | 0.6029 | 0.0844 |
| | (0.6982) | (0.7591) | (0.4553) | (1.1106) | (0.8200) | (1.9336) |
| Medication review | 0.0090 | 0.0067 | 0.0010 | -0.0139 | -0.0102 | -0.0009 |
| | (0.0120) | (0.0135) | (0.0102) | (0.0195) | (0.0151) | (0.0305) |
| Pharmacist | 0.0054 | 0.0055 | 0.0029 | 0.0162 | 0.0119 | 0.0094 |
| | (0.0104) | (0.0113) | (0.0118) | (0.0134) | (0.0106) | (0.0193) |
| Pressure ulcer | -0.0006 | -0.0022 | 0.0011 | -0.0231 | 0.0010 | -0.0005 |
| | (0.0009) | (0.0010) | (0.0006) | (0.0044) | (0.0011) | (0.0017) |
| Vaccinations (influenza) | 0.0126 | 0.0140 | 0.0105 | 0.0195 | 0.0159 | 0.0218 |
| | (0.0027) | (0.0030) | (0.0019) | (0.0043) | (0.0033) | (0.0056) |
| Vaccinations (pneumococcal) | 0.0153 | 0.0188 | 0.0157 | 0.0253 | 0.0195 | 0.0141 |
| | (0.0050) | (0.0056) | (0.0025) | (0.0068) | (0.0055) | (0.0125) |
| Physical restraint | 0.0018 | 0.0013 | 0.0008 | 0.0002 | 0.0004 | 0.0011 |
| | (0.0013) | (0.0016) | (0.0010) | (0.0024) | (0.0019) | (0.0016) |
| Chemical restraint | -0.0080 | -0.0110 | -0.0084 | -0.0100 | -0.0078 | -0.0005 |
| | (0.0070) | (0.0073) | (0.0054) | (0.0108) | (0.0084) | (0.0111) |
| Assessment every 3 months | 0.0201 | 0.0268 | 0.0243 | 0.0355 | 0.0317 | 0.0248 |
| | (0.0084) | (0.0095) | (0.0055) | (0.0116) | (0.0097) | (0.0135) |
| Assessment transmitted in $\leq$ 7 days | 0.0475 | 0.0525 | 0.0531 | 0.0808 | 0.0642 | 0.0901 |
| | (0.0065) | (0.0072) | (0.0044) | (0.0094) | (0.0074) | (0.0170) |

*Notes*: This table reports the simple difference-in-differences estimate of the effect of each citation on the corresponding outcome described in Table 3 using TWFE, Borusyak, Jaravel, and Spiess (2024), Sun and Abraham (2021), Callaway and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2020), and synthetic difference-in-differences (Arkhangelsky et al. 2021) (see details in Section 6). The unit of observation is at the nursing-home-by-quarter level. Standard errors are clustered by nursing home.