Input Regulation and the Production of Hospital Quality

Chandni Raja *

November 21, 2024

Please click here for the latest version of the draft

Abstract

We have a limited understanding of how nurses, physicians, and patients interact to produce high quality care but these interactions are central to efficient regulatory design. This paper estimates a value-added production model for hospital quality in nurses per patient, physicians per patient, and patient health using identifying variation from the 1999 California nurse staffing mandate – the first and to date one of few pieces of comprehensive legislation of nurse staffing levels in hospitals. I find nurses and physicians to be highly complementary (near Leontief) in production. I show that minimum nurse-to-patient ratios that do not account for these complementarities increase healthcare labor costs by 1.4 percent holding quality constant amounting to $24 million in costs across hospitals affected by the mandate. On average, I do not find evidence of across-hospital misallocation of nurses to low productivity hospitals due to ratio regulation – low staffing hospitals are as productive as their high staffing neighbors – but I find allocative gains can be made by reallocating nurses to hospitals with higher severity patients where they are more valuable.

1 Introduction

Low quality hospitals are a source of regulatory concern.[1] The question of "how" to efficiently regulate these hospitals is consequential for production misallocation both within and across firms – a sizeable literature investigates the linkages between regulation and misallocation (Restuccia and Rogerson, 2013; Chandra et al., 2023a). Efficient regulation

*Department of Economics, University of California, Los Angeles (email: chandni.raja@gmail.com). I am grateful for the encouragement and support of my advisors: John Asker, Martin Hackmann, and Will Rafey. I thank Maya Ayoub, Jesper Böjeryd, Naomi Crowther, David Henning, Hugo Hopenhayn, Jonathan Kowarski, Michael Rubens, Yotam Shem-Tov, Nick Tilipman, and seminar participants for helpful comments. I thank Harry Dhami at the California Department of Healthcare Access and Information (HCAI) for answering my questions regarding the financial reporting data. This work benefited from secure data storage facilities and support provided by the California Center for Population Research (P2CHD041022). The content is solely the responsibility of the author and does not necessarily represent the views of any organizations listed. All errors are mine.

[1]In the U.S., the lowest rated hospitals have in-hospital mortality, short-term complication, and infection rates that are at least twice as large as those at the highest rated hospitals. See Rosenberg et al. (2016) for these magnitudes.

is all the more consequential considering the magnitude of the healthcare sector which constitutes 17 percent of U.S. GDP and employs nearly 10 percent of its workforce.[2]

In this paper, I estimate the misallocation from a widely debated input regulation of hospitals: minimum nurse-to-patient ratios.[3] Input regulation may be appealing relative to the direct regulation of quality for several reasons: political feasibility, lower monitoring costs for inputs relative to quality, and inadvertent consequences of direct quality regulation (Gupta, 2021) are examples.[4] On the other hand, input regulation is in theory misallocative because hospitals use multiple inputs in production (regulation of a single input leads to misallocation across inputs within the hospital) and hospitals have heterogeneous productivities (regulation leads to misallocation of the regulated input across hospitals). These dimensions are illustrated in Figure 1.[5]

We have a limited understanding of how the primary inputs into hospital production – nurses, physicians, and patients – interact to produce high quality care but these interactions are central to understanding the misallocation from ratio regulation. To study these relationships, I estimate a value-added model of hospital quality production in nurses per patient, physicians per patient, and patient case mix. There are several challenges: measurement of hospital quality given endogenous patient selection into hospitals, measurement of physician labor given U.S. hospitals infrequently hire physicians directly, and the endogeneity of inputs to unobserved hospital productivity.

I overcome these challenges by exploiting a unique setting of California hospitals which deliver medical care to 3 million patients each year. I construct risk-adjusted hospital quality (measured as the 30-day non-readmission rate) from administrative patient-level discharge data and I observe nurse and physician labor from detailed hospital financial reporting data. I leverage identifying variation in nurse labor from the 1999 California nurse staffing mandate which was the first and to date one of few pieces of comprehensive legislation worldwide establishing minimum nurse staffing ratios in hospitals. The mandate affects the provision of medical care for the nearly 3 million patients admitted to California hospitals each year.

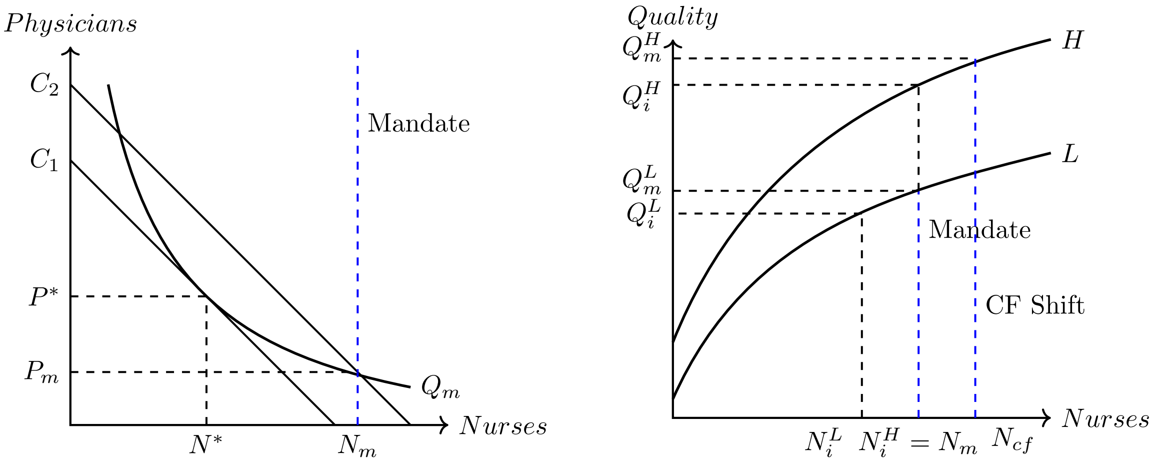
I use the recovered production primitives to quantify the within-hospital misallocation between nurses and physicians and the across-hospital misallocation of nurses to low marginal product hospitals as shown in Figure 1. I find that nurses and physicians are highly complementary (near Leontief) in production. I show that regulation that does not

[2]See CMS (2024) and BLS (2024), respectively, for the 17 percent and 10 percent figures.

[3]Minimum nurse-to-patient ratios have received significant attention as a way to regulate low performers. Ratios are in consideration at the federal and state levels in the U.S. with active bills S.1567 (U.S. Senate), SB240 (Pennsylvania Senate), and S6855 (New York Senate).

[4]With respect to political feasibility – nurse unions are a notably powerful force and responsible for several political drives for minimum staffing ratios (Semuels, 2014). With respect to monitoring costs – monitoring quality may be relatively costly because it requires the development and calculation of risk-standardized quality measures. On the other hand, nurse unions play an active role in monitoring compliance with ratios. With respect to inadvertent consequences of direct quality regulation – Gupta (2021) finds that nearly half of the average reduction in readmission due to Medicare's Hospital Readmissions Reductions Program was due to selective readmission.

[5]Ratios lead to input misallocation within hospitals if the targeted quality can be achieved at lower cost using a lower level of the regulated input. Depending on the magnitude of this misallocation, we may prefer alternate policies that do not intervene in factor decisions. At the same time, ratios lead to misallocation of labor across hospitals if the targeted hospitals have low marginal products relative to another hospital in close proximity that could treat the same patients. My paper focuses on the differential in quality gains but acknowledges that the valuation of these gains is important.



(a) Within-Hospital Misallocation

(b) Across-Hospital Misallocation

Figure 1: Misallocation Within and Across Hospitals

Notes: In panel (a), I show hypothetical isoquant and isocost curves associated with the quality Q_m produced under the mandate. N_m and P_m represent the nurse- and physician-to-patient ratios used to produce Q_m. However, Q_m can be produced at lower cost under isocost curve C_1 using the cost-minimizing input vector $\{N^*, P^*\}$ with the difference in costs $(C_2 - C_1)$ representing the within-hospital misallocation. In panel (b), I show a hypothetical quality production function for two hospitals in close proximity to one another: H and L. L is affected by the mandate and shifts nurse staffing from N_i^L to N_m while hospital H has an initial level of N_m and is unaffected. If we consider moving the mandate-induced number of nurses from hospital L to H ("CF Shift"), the production differential due to the heterogeneous marginal products is $(Q_m^H - Q_i^H) - (Q_m^L - Q_i^L)$.

account for these complementarities leads to within-hospital misallocation of 1.4 percent of healthcare labor costs equaling $180,000 for the average hospital and $24 million in costs across hospitals affected by the mandate. I find that labor per patient and patient health are substitutable and the marginal product of nurses is higher when the patients they treat are sicker. On average, I find no evidence of across-hospital misallocation due to the ratio regulation because low staffing are as productive as their peers and have a higher marginal product of labor due to their lower nurse levels. However, relative to a mandate that does not take into account differences in patient mix across hospitals, there are allocative gains to be made for specific pairs of hospitals by reallocating nurses to the hospital with higher severity patients where nurses are more valuable.

This paper contributes to a deep literature on productivity in healthcare. *The Dartmouth Atlas* and others find wide disparities in health outcomes across healthcare providers (Chandra et al., 2016a,c; Einav et al., 2022; Fisher et al., 2009). Understanding the within-hospital drivers of productivity differentials across hospitals is central to regulating low performers. The healthcare literature has focused on inefficient treatment use as a driver of these productivity differentials (Chandra et al., 2023a; Skinner and Staiger, 2015; Chandra and Skinner, 2012; Garber and Skinner, 2008) but labor use is understudied from an efficiency perspective despite labor's central role in production and the policy concerns over the need for efficient allocation given healthcare labor shortages (KFF, 2023).[6] My

[6]The provision of healthcare is labor intensive and labor costs are estimated to make up two-thirds of total healthcare

paper addresses this gap using methods from the rich industrial organization literature on production functions that allow me to recover structural objects and quantify misallocation in labor use both absent and due to regulatory design.

The paper proceeds as follows. First, I construct the risk-adjusted, 30-day hospital-wide non-readmission rate as my measure of quality using administrative patient discharge data for California hospitals from 1995-2008. I document descriptive facts about the incidence of the 1999 California nurse staffing mandate that I leverage to identify the production primitives. Using a difference-in-differences design, I show that the mandate led to a 12 percent increase in the nurse-to-patient ratio and a 0.7 percent increase in the non-readmission rate among treated hospitals within one year of implementation. I provide descriptive evidence that nurses were added to hospitals that varied widely in terms of their existing levels of patient health and physicians per patient and that there were large changes in the ratio of nurses to physicians as a consequence of the mandate. Using this identifying variation, I show evidence of heterogeneous treatment effects of the mandate on quality suggestive of interaction between inputs: the quality gains were larger for hospitals with more physicians and with sicker patients.

Next, to understand the underlying mechanisms and quantify the two dimensions of misallocation, I estimate a value-added model of hospital quality production featuring heterogeneous productivities across hospitals and flexible elasticities of substitution between the inputs which is operationalized using a translog parameterization. The industrial organization literature has outlined the challenges with the identification of production models due to the endogeneity of inputs to unobserved firm productivity (Ackerberg et al., 2015). I address the endogeneity issue using a statistical restriction on the productivity process, fixed effects to control for the time-invariant component of productivity, and the mandate and lagged input variables as instruments to address the time-varying component of productivity. I estimate the model using IV2SLS with fixed effects.

I recover the marginal quality product of nurse labor and the elasticity of substitution between nurses and physicians from the production estimates. I find that nurse and physician labor are highly complementary in the production of quality with an elasticity of substitution ranging from zero ("perfect complements") to 0.2 depending on the input levels.[7] High complementarity between the inputs reconciles prior estimates of large negative consequences of negative staffing shocks (Friedrich and Hackmann, 2021; Gruber and Kleiner, 2012; Propper and Van Reenen, 2010) and muted positive consequences of positive staffing shocks. Consistent with reduced-form evidence (Chan and Chen, 2022), I find that nurses and physicians are more substitutable when patients have less severe cases. I find that labor per patient and patient health are substitutable and that the marginal product of nurse labor decreases in patient health. Intuitively, nurses are more valuable in settings where patients have more severe cases.

expenditures (World Health Organization, 2000)

[7] The elasticity of substitution between nurses and physicians is the percent change in the nurse to physician ratio divided by the percent change in the physician to nurse marginal product ratio. A low elasticity of substitution indicates that even large changes in the relative marginal products do not lead to large changes in labor allocations.

I validate the production model using the reduced-form treatment effects of the mandate estimated from a difference-in-differences model. I show that the implied quality effects from feeding the reduced-form treatment effect of the mandate on nurse staffing into the structural model at different levels of physicians and patient health are qualitatively consistent with the reduced-form effects from the difference-in-differences model.

Using the recovered primitives, I estimate the magnitude of the within-hospital misallocation between nurses and physicians by solving a cost-minimization problem in which the hospital chooses the levels of nurses and physicians per patient to produce a mandated quality with and without a minimum staffing ratio regulation. I impose a mandated quality set at the average quality effect of the California mandate (a 0.5 percent increase in non-readmission relative to the pre-period non-readmission rate) and set the minimum staffing ratio threshold at the median of the pre-period nurse staffing level implying similar incidence to the California mandate. My findings from the counterfactuals indicate that one-fifth of the hospitals are unable to improve quality by 0.5 percent absent productivity gains implying that any regulation that targets labor even jointly would be misallocative for this subset. For the remainder of hospitals I compare the cost-minimizing and staffing mandate scenarios and find that ratios increase healthcare labor costs by 1.4 percent holding quality constant amounting to \$24 million in costs across hospitals affected by the mandate. When hospitals are given the option to choose the input mix, they prefer to increase the nurse and physician ratios by nearly equal percentages, consistent with fixed-proportions production.

I highlight that the estimated magnitude of the misallocation assumes efficient allocation of nurses and physicians absent the regulation. The model implies that the efficient ratio of full-time equivalent nurses to full-time equivalent physicians engaged in patient care is around 2.70-2.90. Prior to the mandate, I estimate that the mean of this ratio among treated hospitals was 3.12 with an interquartile range of 1.51 to 4.02 indicating a significant number of hospitals for whom the nurse to physician ratio was already inefficiently high. In this way, the model provides insights into why the mandate improves quality for some hospitals and not others and highlights the role of pre-existing levels of other inputs in driving the quality gains.

Finally, I use the recovered productivities to assess the magnitude of across-hospital misallocation. If we hold fixed the number of nurses added due to the regulation and change their allocation across hospitals could we produce higher quality of care? I consider the counterfactual that the nurses added to hospitals treated by the mandate are instead added to the nearest untreated hospital within 10 miles. I document that less than one-third of treated hospitals have an untreated hospital within 10 miles. For these hospitals, my model implies an average 1.1 percent gain in quality from the mandate compared to an average 0.7 percent counterfactual gain at the untreated hospitals (0.4 percentage point differential). I therefore find no evidence of misallocation on average owing to the fact that treated hospitals have lower nurse staffing levels and comparable productivities to untreated hospitals. There are, however, allocative improvements that can be made for specific pairs of hospitals located in densely populated counties where the untreated hospitals admit higher severity patients.

This paper contributes to two main literatures. First, I contribute to a deep literature on hospital productivity. Several aforementioned papers document the wide disparities in productivity across healthcare providers and study the role of inefficient treatment use in driving these productivity differences. Labor use is understudied from an efficiency perspective despite labor's central role in production. A large literature estimates the quality returns to the use of specific inputs including nurse labor (Gupta, 2021; Bloom et al., 2015; Raja, 2023; Friedrich and Hackmann, 2021; Bartel et al., 2014; Gruber and Kleiner, 2012; Propper and Van Reenen, 2010; Chan and Chen, 2022; Chandra et al., 2023a; Skinner and Staiger, 2015; Chandra and Skinner, 2012; Garber and Skinner, 2008). However, absent knowledge of how the returns to the input in question vary with its level, other inputs, and the firm's productivity we cannot study the efficiency of these allocations. Notably a few papers study how the returns to physicians vary based on the hospital's productivity and the implications for the allocation of physicians across hospitals (Mourot; Huckman and Pisano, 2006). Relative to these papers, I focus on the interaction between physicians and nurses and patient case mix in addition to physicians' interaction with the hospital's productivity.

Importantly, my paper uses methods from the rich industrial organization literature on production functions which have had limited application to the study of healthcare quality (Romley and Goldman, 2011; Grieco and McDevitt, 2017; Gertler and Waldman, 1992).[8] By allowing for flexible elasticities of substitution between multiple inputs and heterogeneous productivities across hospitals, my model captures to greater extent the organizational complexity of the hospital and the interaction between inputs in the production of medical care quality. [9] These modeling choices allow me to uncover the interaction between nurses, physicians, and patients and show that the organizational structure of the hospital affects the returns to the staffing mandate.

Second, I build upon the body of empirical work testing for misallocation within and across firms (Restuccia and Rogerson, 2013) by focusing on the design of input regulation which are considered in several regulated industries including education (minimum class sizes) and environmental markets (technology standards). Minimum staffing ratios in healthcare are notable in their own right and are under wide legislative consideration in the U.S. but understudied from an efficiency angle. I follow papers in the macroeconomics and industrial organization literatures by estimating the underlying production or cost primitives and using the recovered primitives to assess the misallocation relative to an efficient benchmark (Hsieh and Klenow, 2009; Asker et al., 2019). This approach also allows

[8] Lee et al. (2013) do not consider hospital quality but estimate a hospital revenue production function from which they consider the marginal products of information technology inputs. Chandra and Staiger (2020) allow the hospital's treatment use to vary as a function of its productivity.

[9] Health services research has highlighted the coordination and interaction between nurses and physicians in providing care (Havens et al., 2010).

me to test for within-firm misallocation contributing to several papers in healthcare which estimate misallocation in treatment use (Chandra et al., 2023a; Skinner and Staiger, 2015; Chandra and Skinner, 2012; Garber and Skinner, 2008). My findings have implications for the growing policy concerns associated with labor allocation within healthcare and between healthcare and other sectors (CHCF, 2021).

The remainder of the paper proceeds as follows. Section 2 introduces the data and provides a description of production in the hospital setting and the measurement of quality and inputs. Section 3 highlights the reduced-form findings that inform the structural model of production. Section 4 presents the structural model and parameterization. Section 5 discusses the model identification and empirical strategy. Section 6 presents the results from the estimation of the model and recovers the production primitives. Section 7 presents the results on within- and across-hospital misallocation. Section 8 concludes.

2 Data, Measurement, and Setting

In this section, I provide descriptions of the data sources, production in the hospital setting, and how I measure quality and inputs. This section serves several purposes. In addition to providing an introduction to the data sources, it provides an overview of hospital production that directly informs the assumptions that I make to identify and estimate the structural model in Section 5. Given the complexity of hospital quality measurement, it also provides a detailed description of the clinical measure that I use for hospital quality (30-day hospital-wide non-readmission) and discussion on why it is the appropriate outcome to use with respect to the research question at hand.

2.1 Data Sources

I use data from three main sources from 1995-2008. First, I use financial reporting data from the California Department of Healthcare Access and Information (HCAI)'s Hospital Annual Financial Disclosure Reports to measure input use at the hospital unit level for each hospital and year. The data are desk-audited and notable for their granularity and detail – these data report patient volumes, capacity, revenues, nurse and administrative labor hours, and expenditures on labor, materials, and capital for each inpatient hospital unit (e.g. Medical/Surgical Acute Care) of reporting hospitals. Hospital characteristics including ownership type, medical staff numbers and specialties, and services inventories are reported at the hospital level. Notably, these data have been used in other papers to study the effects of the nurse staffing mandate (Raja, 2023; Cook et al., 2012; Spetz et al., 2013; Mark et al., 2013; Munnich, 2014), to estimate a cost function for hospital quality (Romley and Goldman, 2011), and to study the returns to information technology by estimating a revenue production function (Lee et al., 2013).

Second, I link these data to administrative patient discharge data on California hospitals which include patient characteristics, date of admission and discharge, and primary and secondary diagnoses and procedure codes for each discharge at a California hospital between 1995-2008. These data allow me to construct my measure of hospital quality (30-day hospital-wide non-readmission rate) and risk-adjust the measure for each hospital and year. As far as I am aware, no other paper has linked these patient level health data with the financial reporting data. In the future, I plan to link these patient discharge data to death records in order to observe out-of-hospital mortality following inpatient stays and study risk-adjusted, 30-day survival rates as an additional clinical measure of quality.

Third, I use external data on the two sources of quasi-experimental variation that I use to identify the production primitives. I use the Dartmouth Atlas of Healthcare's publicly available hospital tracking file to compile a panel of hospital closures, mergers and acquisitions, and participation in CMS's Critical Access Hospital (CAH) program. I use data from the California Department of Health on the levels and timing of the 1999 California nurse staffing mandate.

2.2 Production in the Hospital Setting

2.2.1 *Hospitals*

Hospitals produce patient volumes (quantity) and clinical outcomes (quality) for patients with a myriad of initial diagnoses. For example, hospitals produce 30-day survival and 30-day non-readmission rates (two clinical outcomes that serve as measures of quality) for heart attack patients, pneumonia patients, and post-operative patients using labor, capital, and materials per patient. In this paper, I focus on a balanced panel of 208 non-federal, short-term general acute care (GAC) hospitals in California that report patient days in the Medical/Surgical Acute Care unit for each of the thirteen years in my sample between 1996-2008.[10] The year 1995 is used solely as a one-year lookback period to conduct risk-adjustment for patients admitted in 1996. The descriptive statistics for these hospitals for 1996-2002, the period prior to the implementation of the California nurse staffing mandate in 2003, are displayed in Table 1 according to the nurse-to-patient ratio distribution.

Size – The average hospital in my sample reports roughly 56,000 inpatient days and 10,000 inpatient discharges annually. The interquartile range is from 4,000 to 15,000 discharges annually implying significant dispersion in size. My sample includes hospitals designated as small and rural hospitals by the California Department of Health and those designated as CAH hospitals by CMS. Both designations indicate hospitals that serve a critical purpose of providing healthcare access to rural populations.

Service lines – The hospitals in my sample have multiple acute and intensive care

[10] Federal hospitals administered by the Veterans Administration, Department of Defense, or Public Health Service are exempt from California state reporting requirements for patient discharges by because they are not subject to state licensure. Kaiser hospitals were not required to submit hospital financial reporting data separately for their separate facilities until 2022. Therefore both federal and Kaiser hospitals are excluded from this analysis.

units. Examples of acute care units include Medical/Surgical Acute Care, Definitive Observation, and Obstetrics Acute. Examples of intensive care units include Coronary Care and Medical/Surgical Intensive Care. Intensive care units are used for treating patients of higher severity relative to acute care units and most GAC hospitals (97 percent) have at minimum a Medical/Surgical Acute Care unit and a Medical/Surgical Intensive Care unit. The Medical/Surgical Acute Care unit has the largest share of total inpatient days at 48 percent and an interquartile range falling between 34 and 63 percent. [11] However, beyond these two units the service line offerings vary across hospitals. The other units with the highest likelihoods of being offered are Definitive Observation (42 percent of hospitals) and Coronary Care (28 percent of hospitals).

In this paper, I focus on value-added production in the Medical/Surgical Acute Care unit (hereafter "acute care"). The focus on acute care allows me to exploit the exogenous shock in my setting which affected acute rather than intensive care nurse labor. Intensive care nurse-to-patient ratios have been in place in California beginning in the 1976-1977 fiscal year (Spetz et al., 2000) whereas ratios for acute care were established by the 1999 California nurse staffing mandate and implemented in 2003.[12]

Ownership and market structure – The hospitals have a range of ownership types and are located in markets with varied demand and competitive conditions. The majority of the 208 hospitals in my sample are not-for-profit owned (149 hospitals are not-for-profits at some point in the sample period) with far smaller numbers owned by investors (36 hospitals) or the government be it the state, county, city, or district government (47 hospitals). Slightly more than 20 of the hospitals changed ownership over the sample period. These hospitals are located in markets with varied market structures - at one extreme, CAH hospitals are by definition located at least 35 miles from the nearest hospital and at the other, a number of hospitals in populated counties such as Los Angeles and San Francisco are located within a few miles of the nearest hospital. These heterogeneities in ownership and market structure may give rise to heterogeneous incentives to provide quality.

2.2.2 *Quality*

I measure quality as the risk-adjusted 30-day, hospital-wide non-readmission rate reported in Table 1.

[11] In comparison, for hospitals with an intensive care unit, intensive care patient days made up 8 percent of the total inpatient days with the interquartile range falling between 5 and 10 percent. It should be noted that the share of total inpatient days associated with each hospital unit is close but not equivalent to the revenue share of each unit. In the average hospital-year, acute care revenues made up 40 percent of total revenue (interquartile range of 29-54 percent) whereas intensive care revenues made up 16 percent of total revenue (interquartile range of 12-20 percent). Given the majority of payments are made prospectively rather than on a cost-basis, the mismatch likely reflects the fact that the average patient in intensive care is higher severity and therefore the payor pays higher reimbursement for this patient's inpatient stay relative to the average patient in acute care.

[12] The parameters of interest to the audience may differ relative to the ones estimated in this paper if the policy of interest aims to increase nursing in intensive care units. Prior work estimating the returns to nursing (Friedrich and Hackmann, 2021; Gruber and Kleiner, 2012; Propper and Van Reenen, 2010) have not limited to nurse labor in acute care specifically therefore the underlying primitives driving the results in prior work may not be directly comparable to the ones I estimate.

Clinical outcome – I focus on readmission for a few reasons. First, readmission is likely to be sensitive to acute care staffing choices. Most patients spend significant time in acute care with patient days in acute care making up 85 percent of the patient days shared by the acute and intensive care units. Patients that spend time in multiple units are discharged from the hospital from acute care with discharges from acute care making up 93 percent of hospital discharges made from one of the two units. According to practitioners, the end of the inpatient stay is a particularly critical time for monitoring and discharge planning to avoid readmission. Nurses in particular play a well-documented role in preventing readmission (Needleman and Hassmiller, 2009).

Second, readmission is frequently studied by economists, for example in Friedrich and Hackmann (2021), Chandra et al. (2016b), and Gupta (2021), and by using readmission as the outcome I can more easily benchmark my findings to that of prior work.

Third, readmission is of consequence to regulators. Readmission is considered a "costly and often preventable event" with one-fifth of Medicare beneficiaries re-hospitalized within 30 days of discharge in 2003-2004 and one estimate that CMS spent more than $17 billion on payments for readmissions made within 30 days of discharge in 2004 (Horwitz et al., 2012; Jencks et al., 2009). As a result, CMS has several programs aimed to lower hospital readmission rates. As a part of the Hospital Inpatient Quality Reporting Program, CMS publicly reports hospital level risk-adjusted readmission rates for several diagnosis cohorts on its Hospital Care Compare website. Additionally, the Affordable Care Act of 2012 established several Medicare value-based purchasing programs including the Hospital Readmissions Reductions Program which ties hospitals' Medicare reimbursements to their readmission rates for specific diagnosis cohorts.

Patient population – I use a broad ("hospital-wide") patient population rather than a single diagnosis cohort (e.g. acute myocardial infarction) for a few reasons. In an ideal world, I would be able to observe the inputs allocated to each patient but given the limitations of the data I can only observe the average inputs per patient. To keep the inputs and outcomes at the same level, I aggregate the patient population across diagnosis cohorts. Intuitively, hospitals choose inputs to reflect the entire patient population rather than a single diagnosis cohort and the model should reflect this feature of production as closely as possible. Using a broader patient population additionally allows me to speak to broader implications of staffing decisions beyond a single cohort.[13] Finally, an aim of this paper is to model the interaction between patient case mix and labor per patient and this requires variation in diagnoses across the patient population. Details of the construction of the 30-day, hospital-wide non-readmission rate are included in the Appendix.

[13]If we could observe the inputs at the patient level, the estimation of a production model at the diagnosis cohort level can be advantageous if we expect the relationships between inputs and clinical outcomes to vary across cohorts or if we wish to adjust outcomes for individual-related risks but the determinants of risk vary across cohorts. For example, prior history of heart disease to be a larger risk factor in 30-day non-readmission for heart attack patients relative to other patients.

Risk-adjustment – A primary reason that production of healthcare quality is distinct from production of output in most other markets is that the patient who receives the output of production is also an input. A patient who is admitted to a hospital has a number of characteristics that could yield them more or less risky to the hospital for producing the quality outcome compared to other patients. The non-random sorting of patients to hospitals is known as the "selection in" problem and is commonly referred to in the healthcare literature (Chandra et al., 2023b).

I address the observable heterogeneity in patient mix in two ways. First, I risk-adjust hospital quality for age, gender, race, and history of inpatient care consist with the healthcare literature. Prior work analyzing health outcomes at the level of the diagnosis cohort has excluded admissions in which the patient has had with an inpatient stay for the same condition within the prior year (Chandra et al., 2016a; Friedrich and Hackmann, 2021) or within shorter time frames (Gupta et al., 2021). Given my patient population is broader than a single condition, I follow best practices by excluding any admission in which the patient had an inpatient stay for any condition within the prior year.[14]

In addition to excluding patients based on their history of inpatient care, I residualize the non-readmission rate using age, gender, and race interaction terms at granular age buckets as in the literature (Grieco and McDevitt, 2017). I regress the non-readmission rate on these interaction terms and recover the sum of the constant and residuals from the estimated equation which I term the risk-adjusted, non-readmission rate. In Appendix Figure A.1, I show the distributions of the non-readmission rate and the risk-adjusted non-readmission rate for this sample. Given that my hospital level quality measures are constructed from the patient level discharge data, I will be able to conduct robustness and make modifications on the risk-adjustment procedure in the future if needed.

Second, I explicitly model the relationship between risk-adjusted hospital quality and the inverse of CMI (which I term "patient health") in my production model. While the risk-adjustment step controls for variation in hospital quality due to variation in age, race, gender, and history of inpatient care, it does not address the variation in hospital quality due to variation in diagnoses and co-morbidities which are captured by the CMI. [15] In a departure from the existing literature, I do not assume that production is multiplicatively separable in patient health and instead estimate the elasticity of substitution between patient health and labor per patient. As I discussed earlier, recovering the substitutability between patient health and labor per patient is of importance for regulatory design.

Table 1 shows that both the factors used in risk-adjustment and the CMI vary widely across hospitals. Table 1 indicates variation in the CMI across the nurse-to-patient ratio distribution and indicates variation between the risk-residualized and non-residualized quality measures. Consistent with expectations, I find that patient characteristics are correlated with observable characteristics of hospitals including not-for-profit ownership,

[14]As a consequence, my sample of admissions has a much higher average rate of non-readmission of around 90 percent compared to the 80 percent rate that I obtain when I include patients with a history of inpatient care.

[15]The measure of hospital quality and CMI do not cover identical patient populations because I restrict my quality measure to patients that had not been admitted to an inpatient stay in the past year.

Table 1: Descriptive Statistics for California Hospitals from 1996-2002

	Nurse-to-Patient Ratio Distribution			
	Bottom 25	25-50	50-75	Top 25
Hospitals	52	52	52	52
Annual discharges	9,367	10,368	10,993	9,433
Annual inpatient revenue ($)	60,769,720	78,174,402	77,836,713	77,893,051
Acute share of revenue	0.366	0.401	0.417	0.448
Case Mix Index	1.03	1.08	1.10	1.14
Hospital-wide discharges				
Hospital-wide 30-day non-readmission rate	0.902	0.897	0.887	0.897
Hospital-wide risk-residualized rate	0.970	0.968	0.966	0.973
Hospital-wide length of stay	3.407	3.510	3.496	3.544
Inputs in acute care				
Nurses per 1,000 patient days	2.196	2.443	2.725	3.230
Physicians per 1,000 patient days	1.091	1.289	1.295	1.233
Materials expenditures per 1,000 patient days ($)	4,403	3,531	3,872	4,120
Capital expenditures per 1,000 patient days ($)	433,019	468,360	541,930	580,107
Patient care costs per 1,000 patient days ($)	399,171	473,644	547,243	626,019
Hospital characteristics				
Share church or non-profit	0.654	0.596	0.692	0.731
Share investor-owned	0.115	0.192	0.154	0.096
Share teaching hospitals	0.038	0.096	0.115	0.154
Share small/rural hospitals	0.173	0.115	0.135	0.212

Notes: This table includes the 208 hospitals in my balanced panel sample from 1996-2008. Hospitals are grouped into quartiles of the nurse-to-patient ratio distribution based on their average values from 2000-2002 (prior to the implementation of the California nurse staffing mandate in 2003). I follow CMS in the exclusion criteria for index admissions in the sample construction for the non-readmission rates (Horwitz et al., 2012) and additionally exclude admissions in which the patient had an inpatient stay for any condition within the prior year. The risk-adjusted rate is the residualized rate after controlling for interacted age, gender, and race indicators. Patient care costs include the costs accrued directly to the hospital unit and costs accrued centrally and then allocated to the unit.

small or rural hospital status, and teaching hospital status.

2.2.3 *Inputs in Production*

In this sub-section, I discuss the three inputs of interest (nurse labor, physician labor, patient health) and licensing restrictions between nurses and physicians which are an important institutional feature that governs the production relationships.

Nurse labor – Nurse labor is reported for each hospital unit, hospital, and year in hours of clinical nursing time. Reported hours are total paid hours including overtime less hours not on the job (vacation, sick leave, holidays, and other paid time-off). In California and most other U.S. states, there are two types of licensed nurses (Registered Nurses and Licensed Vocational Nurses) where RNs are the higher-skilled and higher-licensed nurse.

I aggregate RN and LVN labor in my analysis given that the majority of nurse labor in the hospital comes from RNs and their share of total nursing hours has only grown over time. In Appendix Table A.1, I present the average numbers of physicians and nurses for hospitals according to their location in the staffing distribution. Over 85 percent of nurses employed in the California hospital setting between 1996-2002 (prior to the mandate) were RNs.[16] By 2019, 90 percent were RNs indicating that the presence of vocational nurses is small.[17]

I aggregate the number of nurse hours to the number of nurses under the assumption that one nurse works 40 hours per week for 52 weeks of the year.[18] I then divide the number of acute care nurses by the number of acute care patient days and multiply by 1,000 days to obtain the nurses per 1,000 patient days.

The average hospital-year between 1996-2008 has 2.88 nurses per 1,000 patient days implying a nurse-to-patient ratio of 0.250. This average was far lower in the period prior to the mandate's implementation (0.222) compared to the period after implementation (0.287) but the dispersion across hospitals remained stable indicating that the hospital's input demand increased independently of the mandate on the upper end of the staffing distribution. The interquartile range in 2002 was 0.200-0.273 and in 2004 was 0.231-0.307.

Physician labor - Physician labor is reported for each hospital and year in terms of the number of "active medical staff" affiliated with the hospital at the end of the reporting period. [19] Active medical staff refer to hospital-based and non-hospital-based physicians that are voting members of and can hold office in the Medical Staff organization of the hospital (HCAI, 2003). Of the five categories of physicians who work in a hospital (attending, associate, house staff, courtesy, and consulting), only courtesy and consulting staff are excluded from the active medical staff category. The number of active medical staff therefore reasonably captures physician labor employed in patient care.

Given that the number of physicians is reported at the hospital level, I use the revenue share in acute care to allocate the number of physicians per patient to the unit level. I multiply the number of physicians for each hospital and year by the acute care share of total inpatient revenue, divide the resulting number by the number of acute care patient days, and then multiply by 1,000 days to obtain the physicians per 1,000 patient days. [20]

[16]RNs may be more useful than LVNs in hospitals due to a better match between the higher training their receive and the higher severity of the inpatient hospital setting relative to home health or nursing home settings which also employ licensed nurses. The fact that the highest staffing hospitals (also the highest patient severity hospitals) in Table A.1 have the highest RN share of total nursing staff is supportive of this match effect.

[17]Depending on the setting there may be a need to study the quality returns to the two types of nurses separately. For example, states where licensing restrictions are less stringent and/or lower-skilled nurses make up a larger part of the hospital workforce.

[18]The HCAI hospital financial accounting manual advises hospitals to use this assumption to convert hours to number of full-time equivalent employees.

[19]In California, regulation prohibits the majority of hospitals from directly employing physicians. The exceptions are county hospitals and teaching hospitals. Therefore most physicians are organized in physician practice groups that contract with hospitals and hospitals are not required to report these expenditures directly in the hospital reporting form. It is only specific types of physician contract arrangements (reported on Page 2 of the hospital reporting form) that require physician fees to be reported as expenditures to contractors. Given that these are not observed it is not known what proportion of physician fees are being reported by the hospital and therefore cannot be relied upon for measurement. The reported numbers of active medical staff on the other hand include all physicians regardless of whether they are employed directly by the hospital or not.

[20]The HCAI hospital financial accounting manual uses the revenue share of the hospital unit to allocate costs associated

The measure of the number of physicians affiliated with a hospital is not directly comparable in terms of time spent on patient care to the measure of the number of nurses employed by a hospital without making two adjustments. First, physicians do not necessarily work full-time in the hospitals with which they are affiliated as active medical staff but the number of nurses is constructed under the assumption of full-time equivalent hours.[21] My data report 124,542 active medical staff physicians in 2006 across the hospital-years of the unbalanced sample which still excludes Kaiser and federal hospital physicians. I compare this to aggregate data on the number of licensed physicians actively involved in patient care in California in the same year (49,753) from the California Health Care Foundation to obtain the number of full-time equivalent physicians represented by one affiliation (0.40). Second, while the nursing hours are comprised of only clinical hours not all physician time spent in the hospital is spent on direct patient care. I therefore apply the reported share of patient care from the external source (0.56) to the physician FTEs. [22] The resulting numbers are reported in Table 1.

The average number of FTE physicians in direct patient care per 1,000 patient days (hereafter "physicians per patient") in a hospital-year in the sample is 1.05 with an interquartile range of 0.67-1.58. The average and interquartile range for the number of active medical staff physicians assigned to acute care are 85 and 41-162.

The California financial reporting data are appealing because, as far as I am aware, there is a lack of data available to researchers on physician time at the hospital and year level. [23] We may be concerned, however, that hours worked by the average active medical staff physician varies across hospitals in systematic ways. This would lead to non-classical measurement error in the independent variable if the model used across-hospital variation for identification. I have two comments here: first, my model relies solely on within-hospital variation for identification and consequently the concern over retrieving consistent estimates depends only on the gap between the proxy and true physician time being correlated with time-varying unobservables within the hospital. Second, the financial reporting data have the granularity for robustness checks on this dimension. For each hospital and year, I can observe the number of physicians by hospital- vs. non hospital-based vs. resident vs. fellow, board certification status (board-certified, board-eligible, other), and each of 42 specializations which would allow me to link these figures with external data on physician hours worked at the board certification and specialization level (Leigh et al., 2010). The breakdowns by all categories other than specialization are shown in Appendix Table A.1.

with physicians from the hospital level to the unit level. Specifically, the costs listed in the columns associated with Page 18, line 255 are then allocated to the according to the statistic on Page 19, Column 11 (gross patient revenue).

[21]Several California hospital bylaws indicate that active medical staff must have "regular involvement in patient care" but these requirements are far below full-time.

[22]The HCAI hospital financial data has a section that requires the reporting of physicians' time spent on various activities for hospitals that employ salaried physicians. The average time spent on direct patient care as a function of the total is around 80 percent for salaried physicians.

[23]Survey data from the American Medical Association and proprietary data sources with physician hours are generally not linked to hospitals and the latter are furthermore available beginning only in the mid-2000s. Other papers that have used measures of physician time rely on the Veteran's Affairs hospitals for example Chan and Chen (2022).

Licensing restrictions – Licensing restrictions and the resulting "scope of practice" for nurses in the inpatient hospital setting drive the production relationships that I analyze in this paper. The degree of substitutability between nurses and physicians in the production of quality depends on two factors: (1) the degree of overlapping tasks between nurses and physicians; and (2) the quality returns to the tasks that nurses perform (whether overlapping or not). With respect to overlapping tasks, RNs were not allowed to practice independently of physicians in the inpatient hospital setting during my sample period.[24] The California Nursing Practice Act states that RNs require physicians' orders to perform dependent activities including administering medications and therapeutic agents and require written authorization at the provider-level to perform interdependent functions that overlap with medical practice. These interdependent functions which are termed "beyond the usual scope of nursing practice" include diagnosing disease, prescribing medication or treatment, and penetrating or severing tissue (California Nursing Practice Act, 2012). Even if licensing restrictions allow nurses to perform the same tasks as physicians there is evidence that nurses yield lower quality returns than physicians (Chan and Chen, 2022).

Patient health (inverse of the Case Mix Index) – As mentioned in Section 2.2.3, I model the relationship between quality and patient health using the inverse of the Case Mix Index (CMI) which is calculated by HCAI for each hospital and year beginning in 1996. I use the inverse of the index so that quality production is increasing in the input. I discuss details on the calculation of the CMI in the Appendix.

3 Descriptive Evidence on the California Mandate

In this section, I present reduced-form evidence of the treatment effects of nurse labor on hospital quality using quasi-experimental variation in nurse labor from the 1999 California nurse staffing mandate. The identifying variation that I highlight in this section will be used to identify the structural model in Section 4.

The 1999 California nurse staffing mandate (AB 394) imposed minimum nurse-to-patient ratios in the acute care units of GAC hospitals. As noted earlier, intensive care units were already required to maintain legislated ratios beginning in the 1970s (Spetz et al., 2000). AB 394 directed the California Department of Health to establish the ratios following a public comment period and the ratios were announced on January 2002 for initial implementation dates for the Medical/Surgical Acute Care unit of January 2004 for a ratio of 0.16 and January 2005 for the final ratio of 0.2 (Raja, 2023). The staggered implementation deadlines afforded hospitals extra time to reach the final ratio.

[24] State-level changes over the past decade including in California have loosened the practice restrictions on high-skilled Registered Nurses with a Nurse Practitioner (NP) license and allowed them to practice as independent practitioners. See Chan and Chen (2022). The proliferation of NPs is more widespread in the primary care setting than in the inpatient setting.

3.1 Average treatment effect on quality

I estimate the average treatment effects of the mandate on nurses per patient and risk-adjusted non-readmission. I follow Raja (2023) in defining hospitals as treated by the mandate if they have an average nurse-to-patient ratio of below 0.25 between 2000-2002. [25] I estimate the following event-study specification for y_{ht} as the log of nurses per 1,000 patient days and the log of the risk-adjusted non-readmission rate

$$y_{ht} = \beta_0 + \sum_{t \neq 2003} \beta_t \{YEAR_t = t\} * BELOW_h + \gamma_h + \xi_t + \epsilon_{ht} \tag{1}$$

where $BELOW_h$ is an indicator variable for the treatment that takes on a value of one if the hospital had an average nurse-to-patient ratio of below 0.25 in 2000-2002, and β_t are the treatment effects of interest for years following the excluded year 2003, and γ_h and ξ_t are hospital and year fixed effects.

In Appendix Figure A.2, I plot the raw means of nurses for the treated and control hospitals and the estimated treatment effects β_t from the estimation of Equation (1) with nurses per patient as the outcome variable. Appendix Figure A.2 shows that the mandate led to a 12 percent increase in acute care nurses per patient at treated hospitals within one year of implementation. In Figure 2, I plot the treatment effect for the risk-adjusted non-readmission rate. The corresponding event-study estimates are also presented in Appendix Table A.2. My findings indicate that the mandate had a statistically significant effect on the non-readmission rate at treated hospitals with a magnitude of 0.7 percent within one year of implementation. [26]

3.2 Heterogeneous treatment effect on quality

Notably, the mandate did not direct hospitals to hire other (potentially complementary) inputs and did not account for the heterogeneity in pre-existing levels of physicians or patient health at treated hospitals. I do not find a statistically significant increase in the number of physicians per patient due to the mandate. In Figure 3, I present histograms of the nurse and physician per 1,000 patient days across hospitals in 2000 and 2008 for treated hospitals. Figure 3 illustrates that the distribution of nurses is shifted significantly to the right but the distribution of physicians remains relatively stable. In Appendix Figure A.3, I produce the same graph for untreated hospitals.

However, nurses were added to hospitals that varied widely in terms of the levels of patient health and physicians per patient. In Appendix Figure A.4, I plot nurse staffing and physician staffing per patient in panel (a) or patient health in panel (b) in 2000

[25] I use 0.25 rather than the mandated 0.2 because I observe the nurse-to-patient ratio as an annual average and suppose that the incidence was broader given the need to abide by the 0.2 at all times rather than on average over the year. Raja (2023) finds that the effects of the mandate are estimated to be similar if research design uses a 0.2 threshold to assign treatment instead.

[26] To benchmark the magnitude of these findings to the literature, Gupta (2021) finds that the Hospital Readmissions Reductions Program, Medicare's value-based purchasing program which links reimbursements to readmission rates, led to a 5 percent decline in the readmission rate with approximately 40-50 percent of the decline in the probability of readmission explained by more stringent readmission policies for patients returning within 30 days.

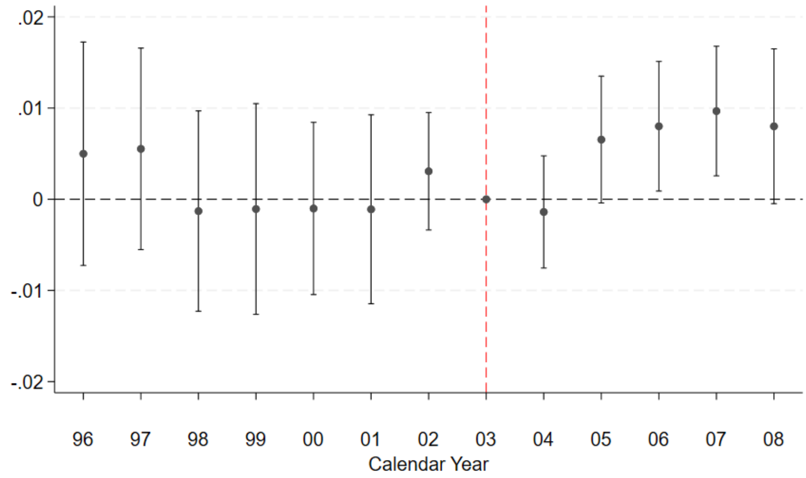


Figure 2: Event-Study Estimates of the Effect of the Mandate on Log Non-Readmission

Notes: This figure plots coefficients β_t and 95 percent confidence intervals from Equation (1) with the log of the risk-adjusted non-readmission rate as dependent variable. Standard errors are clustered at the hospital level. The mandate leads to an average quality effect of 0.7 to 1 percent between 2005 and 2008.

(in red) and in 2008 (in purple) for the treated hospitals. The fitted lines represent the estimated correlations between nurse and physician staffing in each year. In both panels, hospitals across the distribution received a positive shock to nurses. In panel (a), the comparison between the relatively flat slope in 2000 and the steeper one in 2008 indicates that the magnitude of the shock was differential across the distribution.

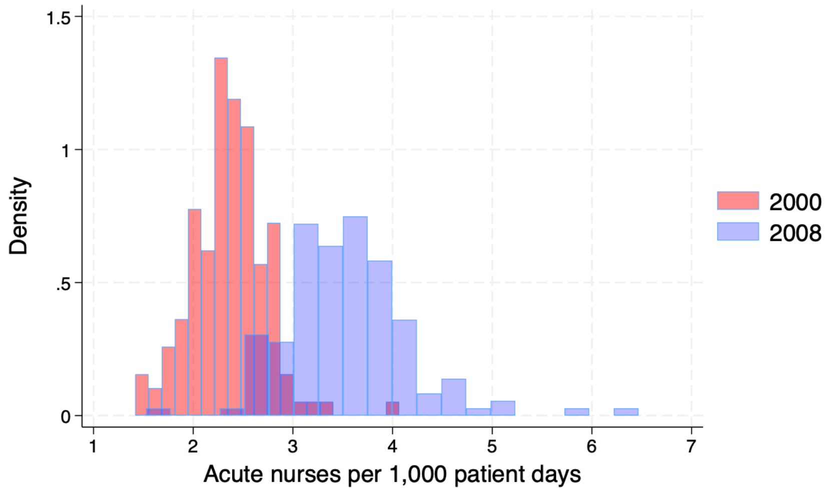
I use this identifying variation to estimate heterogeneous treatment effects of the mandate on quality and study the interaction between nurse labor and these other inputs. According to Seidman (1989)'s definition of "q-complements", two inputs are complements if the marginal product of one input increases in the level of the other and they are substitutes otherwise. [27] I estimate the difference-in-differences model shown in Equation (2) which allows for heterogeneous treatment effects based on the levels of other inputs

$$y_{ht} = \beta_0 + \beta_1 BELOW_h * POST_t + \beta_2 BELOW_h * POST_t * \ln(x^i_{ht}) + \beta_3 \ln(x^i_{ht}) + \gamma_h + \xi_t + \epsilon_{ht} \tag{2}$$

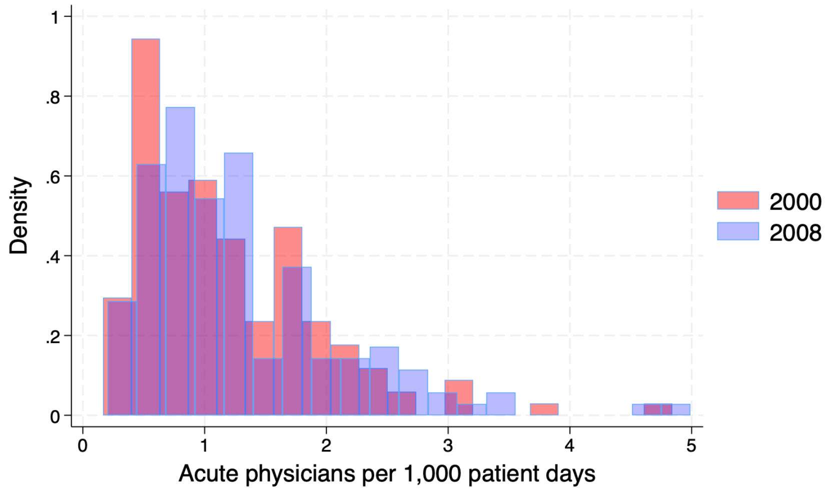
where x^i_{ht} for $i \in \{h, p\}$ represents patient health or physicians per patient in hospital h in year t, respectively. β_1 represents the average treatment effect of the mandate on hospitals with $\ln(x^i_{ht}) = 0$ or $x^i_{ht} = 1$. β_2 represents the heterogeneous treatment of the mandate based on the patient health or physician per patient level.

In Table 2, I present the results from the estimation of Equation (2). Column 1 indicates that the treatment effect increases in the number of physicians per patient. The average treatment effect for hospitals with $\ln(x^p_{ht}) = 0$ (one physician per 1,000

[27] The delineation of two inputs into "substitutes" vs. "complements" depends on which of several definitions we use. For example, if we use the "q-complements" definition in conjunction with a constant elasticity of substitution (CES) production function then we find that the Cobb-Douglas is the boundary between "substitutes" and "complements". Ferguson (1969)'s definition specific to the translog production function that I esitmate is that input i and j are substitutes if $e_i e_j \beta_{ij}$ (the product of their output elasticities and the coefficient on their interaction) is less than zero. Other definitions such as Seidman (1989)'s "p-complements" definition rely on behavioral assumptions which I do not make here.



(a) Nurses Per Patient in 2000 vs. 2008



(b) Physicians Per Patient in 2000 vs. 2008

Figure 3: Input Use in 2000 vs. 2008 for Hospitals Treated by the Mandate

Notes: In panel (a), this figure shows the histogram of hospitals treated by the mandate according to the number of nurses per 1,000 patient days prior to the mandate in 2000 (red) and after the mandate in 2008 (blue). In panel (b), I do the same for physicians per 1,000 patient days which represent the physician FTEs constructed using patient care time. Taken together, the figures indicate that the mandate led to a large shift in the ratio of nurses to physicians.

patient days) is a 0.003 log points reduction in the non-readmission rate (0.3 percent) and statistically insignificant. For the treated hospital with the average number of physicians, $\ln(x^p_{ht}) = 1.50$ and the treatment effect is a 0.006 log points increase (0.6 percent). For the treated hospital at the 90th percentile of the physician distribution, $\ln(x^p_{ht}) = 2.313$ and the treatment effect is a 0.011 log points increase (1.1 percent).

Column 2 indicates that the treatment effect decreases in patient health. The average treatment effect for hospitals with $\ln(x^h_{ht}) = 0$ (a Case Mix Index of one) is 0.004 log points (0.4 percent). For hospitals with the average patient health which is higher severity, $\ln(x^h_{ht}) =$ -0.060 and the treatment effect is 0.006 log points (0.6 percent). For the treated hospital at the 90th percentile of the patient health distribution which is lower severity, $\ln(x^h_{ht}) = 0.128$ and the treatment effect is 0.001 log points (0.1 percent).

Table 2: Diff-in-Diff Estimates of the Heterogeneous Effects of the Mandate

	Log Non-Readmission Rate	
	(1)	(2)
Treat x Post	-0.003	0.004
	(0.006)	(0.004)
Treat x Post x Log physicians per patient	0.006*	
	(0.003)	
Log physicians per patient	-0.000	
	(0.003)	
Treat x Post x Log patient health		-0.027**
		(0.011)
Log patient health		-0.009
		(0.019)
Observations	2,704	2,704
R^2	0.531	0.533
Mean	0.970	0.970
Hospital Fixed Effects	✓	✓
Year Fixed Effects	✓	✓

Notes: This table presents estimates of Equation (2) with the difference-in-differences effect of the mandate on treated hospitals. Column 1 indicates that the average treatment effect for hospitals with $\ln(x^p_{ht}) = 0$ is a 0.003 log points reduction (approximately 0.3 percent), with the average number of physicians $\ln(x^p_{ht}) = 1.50$ and the treatment effect is 0.6 percent. Column 2 indicates that the average treatment effect for hospitals with $\ln(x^h_{ht}) = 0$ is 0.4 percent, with the average patient health which is below $x^h_{ht} = 1$ $\ln(x^h_{ht}) =$ -0.060 and the treatment effect is 0.6 percent.

3.3 Discussion

In this section, I provided suggestive evidence of the interaction between inputs in production. However, these results are only suggestive because there are a number of potential correlates of physician and patient health levels that could be driving the heterogeneity in the returns to nursing across observations. For example, more productive hospitals may have higher physician staffing levels. In this case, adding nurses to a hospital with more physicians may lead to larger quality gains because it is a high productivity hospital rather than the fact that the returns to nurse labor increase in the number of physicians. The reduced-form results in Table 2 do not allow us to differentiate between those mechanisms.

To uncover the mechanisms behind these effects and to quantify the two dimensions of misallocation in Figure 1 we require a structural model of hospital quality production from which we can recover the underlying production primitives, including the elasticities of substitution and marginal product curves, and conduct counterfactual exercises.

4 Model of Hospital Quality Production

In this section, I present a structural model of hospital quality production. The assumptions that I make in this section are informed by the institutional features of the setting that I described in Section 2.

In each period t, hospital h produces quality Q_{ht} which I model as a function of three inputs: patient health (x^h) which is the inverse of the Case Mix Index observed in the data, physicians per patient (x^p), and nurses per patient (x^n). ω_{ht} is the hospital's unobserved productivity and ϵ_{ht} is measurement error.

$$\begin{aligned} Q_{ht} &= e^{\omega_{ht}+\epsilon_{ht}} F(x_{ht}^h, x_{ht}^n, x_{ht}^p) \\ &\equiv e^{\omega_{ht}+\epsilon_{ht}} \prod_{i\in\{h,n,p\}} (x_{ht}^i)^{\beta_i} \prod_{i\in\{h,n,p\}} (x_{ht}^i)^{\frac{1}{2}\left(\sum_{j\in\{h,n,p\}} \beta_{ij} \ln(x_{ht}^j)\right)}. \end{aligned} \tag{3}$$

I apply a translog parametric assumption on F in the second line of Equation (3). The translog parameterization is ideal in my setting because it allows me to estimate the elasticities of substitution between the inputs – one of the primary objects of interest in this paper – unlike the Leontief (which assumes zero elasticity of substitution) or Cobb-Douglas (which assumes unit elasticity of substitution) functional forms. The translog model additionally offers greater flexibility relative to a constant elasticity of substitution (CES) parameterization because it allows the elasticities of substitution to vary across input levels. This added flexibility may be important if the substitutability between nurses and physicians changes with the number of either input. For example, nurses and physicians may be more substitutable for one another at low levels of staffing compared to high levels if roles are more fluid when there are fewer staff.

Taking logs of both sides of (3) leads to the linear-in-parameters estimating equation

$$\ln(Q_{ht}) = \sum_{i\in\{h,n,p\}} \beta_i \ln(x_{ht}^i) + \frac{1}{2} \sum_{i\in\{h,n,p\}} \sum_{j\in\{h,n,p\}} \beta_{ij} \ln(x_{ht}^i) \ln(x_{ht}^j) + \omega_{ht} + \epsilon_{ht}. \tag{4}$$

4.1 Assumption of Value-Added Production

In specifying a "value-added" production function in three inputs, I make several non-trivial assumptions over production: first, that production is separable in acute care labor per patient and patient health[28]; second, that production is Leontief in the inputs excluded from the value-added model; and third, that these excluded inputs are always used in proportion to acute care labor per patient and patient health. Specifically, production should satsify the following properties where x_{ht}^m represent the inputs that are excluded from the value-added production model

$$\begin{aligned} Q_{ht} &= g(x_{ht}^n, x_{ht}^p, x_{ht}^h, x_{ht}^m)e^{\omega_{ht}+\epsilon_{ht}} \\ &= \min[F(x_{ht}^n, x_{ht}^p, x_{ht}^h), s(x_{ht}^m)]e^{\omega_{ht}+\epsilon_{ht}} \end{aligned}$$

where I assume that

$$F(x_{ht}^n, x_{ht}^p, x_{ht}^h) = s(x_{ht}^m) \tag{5}$$

These assumptions allow me to rewrite the gross output production function as the value-added production function shown in Equation (3), given by

$$Q_{ht} = F(x_{ht}^n, x_{ht}^p, x_{ht}^h)e^{\omega_{ht}+\epsilon_{ht}}$$

These assumptions which underlie the so-called "structural" value-added production function are discussed in Diewert (1978) and Gandhi et al. (2017). Given the presence of labor adjustment costs rendering the equality in Equation (5) unlikely to hold for each hospital and year without further assumptions, I make a fourth assumption that the function $s(.)$ is linear as discussed in Ackerberg et al. (2015) and Gandhi et al. (2017).

Imposing these assumptions on production allows me to focus on the inputs that interact with nurse labor (physicians, patient health) whereas the inclusion of all inputs used in the production of the non-readmission rate is challenging if not impossible. [29] Including all of these inputs separately in the production model is challenging from the perspective of identification and estimation of each additional production parameter. Including all of these inputs and aggregating some to reduce the number of parameters – for example, aggregating nurse labor across acute and intensive care units – is not trivial as makes equally restrictive assumptions on the substitutability between inputs (Berndt and Christensen, 1973).

In the hospital setting, the intermediate inputs x_m in Equation (5) are inputs used to treat the patient prior to their arrival in the acute care unit (for example, inputs used for their treatment in the Emergency Department or the intensive care unit) and non-labor inputs used to treat the patient in the acute care unit (capital and materials expenditures

[28] See discussions of separability in Leontief (1947) and in the introduction to Diewert (1978).

[29] Writing about a gross output production function for steel, Leontief (1947) states "the various material processes covered by this formula are so many and so different from each other that even a verbal description of such a vast technological complex would hardly be possible without reference to intermediate commodities."

per patient). The assumption of value-added production implies that these inputs are not substitutable with the modeled inputs.

With respect to the inputs used to treat the patient prior to their arrival in acute care, the implication is that the average patient requires a fixed proportion of resource use outside of acute care to produce quality. This is reasonable if the average patient is "stabilized" to a target health level prior to arrival in acute care and then transferred. The issue is when the "stabilized" health level is itself endogenized, for example if there is overcrowding in the Emergency Department or intensive care.

With respect to the capital and materials expenditures in acute care, Table 1 suggests that the use of non-labor inputs per patient rises with labor per patient and the patient severity index such that the proportions of non-labor to labor inputs remain relatively stable across the staffing distribution. Nurses and physicians play a diagnostic, prescriptive, and monitoring role in producing quality of care. Their role in preventing readmission requires the use of capital to test and diagnose patients and to communicate with one another and the use of materials including pharmaceuticals and medical instruments to administer treatment. [30] This insight is consistent with Gupta (2021)'s finding that the decline in readmission under the Hospital Readmissions Reductions Program is coincident with an increase in use of materials and diagnostic imaging as well as an increase in physician time: someone needs to assess the patient, administer the materials, request and perform the diagnostic imaging, and review the results.

4.2 Economic Assumption on ω

In the model specified in Equation (3), ω represents the elements of production that affect the quality of care and are known to the hospital in period t but unobserved to the econometrician.

There are two points to be made with respect to ω. The first is that variation in acute care labor and patient health explain little of the variation in quality as I show in the next section. I note that this is a finding of this paper. However, as a consequence "productivity" as defined explains a great deal of the variation in quality across and within hospitals. This is a limitation of this paper and may pose a problem depending on the application at hand – for example, if one were to identify and target the least productive hospitals for closure without an understanding of what "productivity" is. However, the research question here is a far simpler one: is nurse labor being allocated to hospitals which for either observed (patient health, physician labor) or unobserved (productivity) reasons obtain larger quality returns from its use? To answer this far simpler question, I argue that it is not necessary to explain the variation in quality than to explain whether the proposed regulation shifts quality and by how much.

[30] The assumption that production is Leontief in capital per patient contrasts with the use of capital expenditures on labor-replacing equipment – for example, artificial intelligence as a diagnostic tool in healthcare. Recent developments may therefore require that production function estimates based on current data make different assumptions on the relationships between labor and capital.

The second point is Equation (3) assumes that production is multiplicatively separable in productivity ("Hicks-neutral" in productivity). As shown by Leontief (1947) the separability assumption of two inputs from a third implies that the ratio of the marginal products of these inputs is independent of the level of the third. In this model, the implication is that the ratio of the marginal products of any two of the three inputs should be independent of the level of productivity. This assumption forecloses the possibility of factor-biased technological change that would differentially increase the returns to nurses, physicians, or patient health during during the sample period.[31]

5 Empirical Strategy

A long literature in industrial organization has discussed the identification challenges in estimating a production model (Marschak and Andrews, 1944; Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015). The framework in Olley and Pakes (1996) delineates these challenges into selection and simultaneity: the hospital's unobserved productivity determines whether and when the hospital enters and exits the market (selection) and its choice of inputs when it operates (simultaneity).[32]

I abstract from issues of entry and exit given that the majority of care is provided by long-lived hospitals. The 208 hospitals in my balanced panel comprise 77 percent of the patient days in acute care over the sample period. [33]

To address the endogeneity of inputs, prior work relies on a combination of timing assumptions over the factors of production, instrumental variables for endogenous factors, statistical restrictions on the productivity process, and the construction of a control function for unobserved productivity in order to identify the model (Ackerberg et al., 2015). The appropriateness of the approach depends on the empirical features of production in the industry.

I model the productivity ω_{ht} of a hospital h in year t as the sum of a hospital specific average (ω_h), a year specific average (γ_t), and a hospital-year specific shock (ξ_{ht})

$$\omega_{ht} = \omega_h + \gamma_t + \xi_{ht} \tag{6}$$

First, I impose a statistical restriction on the productivity process such that the hospital-year specific shock ξ_{ht} is serially uncorrelated. Imposing this assumption allows me to use lagged input variables as exogenous shifters for the three inputs. Second, I use instrumental variables for the endogenous factors by exploiting the two sources of

[31] Ideally, I would include separate productivity terms for labor per patient and patient health given that hospitals are differentiated along the patient heath dimension and a subset of hospitals (for example, teaching hospitals) may be differentially equipped to deal with the sickest patients. Other work including Gandhi et al. (2017) have proposed the use of first-order conditions to identify multiple dimensions of heterogeneity but this paper refrains from imposing behavioral assumptions on the hospital's objectives which forecloses the use of standard tools for identification.

[32] The simultaneity issue applies to patient health in addition to labor inputs because patients can observe the hospital's productivity in a given period and choose the hospital leading to endogenous patient selection. Roughly fifty percent of hospital admissions are elective.

[33] This is far larger coverage than prior work that has focused on selection bias. For example, Olley and Pakes (1996) find that restricting to a balanced panel in their setting uses only 35 percent of the full set of observations.

quasi-experimental variation in my setting and lagged input variables.

Importantly, I do not use timing assumptions over the factors of production or a control function for unobserved productivity. I argue that a timing assumption that any of my three factors of production are "fixed", or chosen before the realization of the productivity shock in a given period, is unlikely to hold true in my setting. The market for healthcare professionals is notably rigid relative to other sectors, however, the "time-to-hire" from the initial search to the contract date for nurses and physicians falls well below the one year mark indicating that nurses and physicians should be considered flexibly chosen at t. See the Appendix for details on the institutional features of the labor markets for nurses and physicians. Additionally, the construction of a control function for unobserved productivity relies on the assumption that there exists an input for which input use increases monotonically in unobserved productivity (Pakes, 1991) which I argue is also an unlikely assumption considering the documented market frictions that likely drive quality choice among hospitals including heterogeneity in hospital market structure (Gaynor and Town, 2011; Propper et al., 2004; Bloom et al., 2015) and heterogeneity in insurer market structure which affects insurer-hospital bargaining over provider rates (Ho and Lee, 2017).

5.1 Instruments for Flexible Inputs

The empirical strategy for estimation requires the use of hospital and year fixed effects in conjunction with instruments orthogonal to the hospital-year specific productivity shock ξ_{ht}. The moment condition for estimation for a vector of instruments z_{ht} would be

$$E[\xi_{ht} z_{ht}] = 0$$

I use lagged input variables as internal instruments for the endogenous inputs.[34] However, the exclusion restriction requires that the instrument affects quality only through the input variables and not through the unobserved productivity shock. Therefore I impose the assumption that ξ_{ht} are not serially correlated to ensure that $x^i_{ht-1} = f(\xi_{ht-1})$ does not imply $\xi_{ht} = g(x^i_{ht-1})$. Note that the adjustment costs model implies persistence in the input variables even if the productivity shocks themselves are uncorrelated (Bond and Soderbom, 2005).

This is a restrictive assumption on the productivity process but given that I have external instruments in addition to lagged input variables, the exclusion restriction can be tested using the Sargan test of overidentifying restrictions. If in fact there is significant serial correlation then the lagged input variables should not be excluded and we should expect to reject the null hypothesis of the Sargan test.

I utilize lagged input variables in addition to two sources of quasi-experimental variation as instruments. The estimation procedure is equivalent to an IV2SLS estimation of

[34] The existence of either labor market rigidities or serial correlation in input prices imply relevance of the instruments. I show that the instruments are relevant in the first-stage estimating equations and weak instrument tests.

Equation (4) with demeaned variables using demeaned instruments. [35]

Internal instruments – I utilize one-period lags of each of the three flexible inputs under the assumption that the lagged input variable is uncorrelated with the hospital-year specific productivity shock. These instruments are considered exogenous because hospitals are assumed to choose their $t-1$ inputs in period $t-1$ before the realization of the productivity shock in period t. I include the three one-period lags and their three interaction terms as instruments.

External instruments – I additionally use two sources of quasi-experimental variation as instruments. I use the 1999 California nurse staffing mandate as a shifter of nurse labor per patient at the treated hospitals. As shown in Section 3, the mandate led to an increase in nurse labor per patient at these hospitals relative to control. The treatment indicator that I use for the mandate is a kinked treatment variable which takes on a value equal to the difference between 0.25 and the average nurse-to-patient ratio of the hospital in 2000-2002 if the hospital had an average below 0.25 (a measure of the incidence of the mandate) and zero otherwise. This indicator is interacted with an indicator variable for whether the year is post-2003. Both the mandate and CAH conversion instruments require the inclusion of year fixed effects to capture the exogenous shift in inputs.

I use hospital enrollment in Medicare's CAH program as a shifter of patient health. Small and rural hospitals that enrolled in the CAH program, which was established in 1997 as a part of the Balanced Budget Act, became eligible to be reimbursed on the basis of cost rather than through the Prospective Payment System in efforts to support the financial health of rural hospitals. Participant hospitals are required to maintain fewer than 25 acute care beds and an average length of stay of less than 96 hours per patient. Prior work finds that as a consequence of enrollment in the CAH program, hospitals experienced reductions in length of stay and changes in the DRGs of admitted patients towards less severe patients (Schoenman and Sutton, 2008).

I corroborate the reduction in length of stay and the increase in patient health (decline in the CMI) in my data and find no effect on the numbers of nurses or physicians per patient suggesting that the reductions in personnel likely mirrored reductions in patient days. CAH conversion did not change the composition of inpatient services but participant hospitals were incentivized to engage in selective admissions or transfer severe patients to other hospitals leading to an increase in patient health. As stated by Schoenman and

[35] The estimation of an IV2SLS model with demeaned variables and demeaned lagged input variables as instruments introduces a mechanical correlation between the demeaned lagged input variables and the error term which is conceptually similar to the problems noted in Nickell (1981) over the inconsistency of a fixed effects model in which there is an underlying dynamic process. The process introduces due to the correlation between the productivity term and future values of the input variable which are captured by the mean of the input variable. The bias is significant in large N, small T panels and reduces as $T \to \infty$ with the bias of the estimates bounded to order T^{-1}. Here my sample consists of 12 years of data. The alternate strategy is to utilize a first differences model but fixed effects and first differences models operate under opposite extreme assumptions when it comes to the serial correlation of the error term: first differences is the efficient estimator when the error term follows a random walk and fixed effects is the efficient estimator when the error term is serially uncorrelated.

Furthermore any correlation between the demeaned instruments and the error term should be apparent from the Sargan test of overidentifying restrictions.

Sutton (2008), participant hospitals may "make more strategic admission decisions in order to ensure that they remain within the program limits on average length of stay (i.e., they would be less likely to admit a patient whose LOS is expected to be much longer than the average target LOS)."

The treatment indicator that I use for CAH conversion takes on a value of one if the hospital ever converted to a CAH hospital and zero otherwise. This indicator is interacted with an indicator variable for whether the year is post-CAH conversion for the given hospital. The staggered treatment dates require the inclusion of year fixed effects rather than a post-period indicator variable to isolate the exogenous shift in inputs.

Discussion of threats to identification – Both the internal and external instruments are valid under the exclusion restriction that impact hospital quality only through the observed inputs and not through the hospital-year specific productivity shock.

To investigate threats to identification, I do two things: first, I discuss threats to identification from the two external instruments and refer to the findings of Raja (2023) and Schoenman and Sutton (2008) which respectively highlight the effects of the nurse staffing mandate and CAH conversion events on hospitals to determine whether there were any plausible effects of these events on unobserved productivity. Second, I discuss threats to identification from the lagged input variables and I show first-stage regressions of the instruments on omitted variables of concern in Appendix Table A.6 to determine whether there is significant movement in these variables.

Raja (2023) documents the effects of the nurse staffing mandate on hospitals finding that it led to reduced capacity, increased bed utilization rates, increased share of lower-licensed nurses, and a reduction in length of stay. Raja (2023) does not find any effects on input use in the intensive care unit or on the number of admissions. It is unlikely that reduced capacity or increased bed utilization rates impact quality conditional on per patient resources and I assume that the returns to nursing being estimated are average returns given the skill-level of the nursing force that is added.

The assumption that length of stay does not affect quality is a restrictive one. Barring evidence of premature discharge, the literature considers length of stay to be an output rather than an input into production (Raja, 2023; Bartel et al., 2014). However, the endogeneity issue arises when hospitals engage in premature discharge – patients being discharged "quicker and sicker" to the detriment of their health – in which case length of stay is likely correlated with both per patient labor use or patient health and independently with the non-readmission rate. I return to this issue below.

Schoenman and Sutton (2008) document the effects of hospital enrollment in the CAH program finding that it led to reductions in the capacity, number of discharges, length of stay, and number of personnel in addition to changes in the DRGs of admitted patients. Reductions in capacity or number of personnel are unlikely to impact quality conditional on the per patient resource use. Length of stay is an issue for the same reason as under the nurse staffing mandate. The number of discharges may also be an issue if there are unmeasured quality returns to scale as found in Dingel et al. (2023).

To address the possibilities that length of stay or patient volumes are omitted variables, I estimate first-stage regressions with length of stay and number of discharges as dependent variables to determine whether they are shifted by the instrument vector. These are shown in Appendix Table A.6. The F-statistics indicate a weak first-stage but to follow-through with the possibility I estimate versions of the model that make average length of stay and number of discharges observable. I find that neither length of stay nor patient discharges are statistically significant at the ten percent level and that the coefficients on the other inputs remain statistically significant and similar in magnitude to the model without length of stay or patient discharges. [36]

6 Production Function Results

In this section, I present the results from the estimation of the production function in Equation (4) and I use the estimated production parameters to recover the elasticities of substitution between the inputs and the marginal product of nurse labor.

In Appendix Tables A.3, A.4, and A.5, I report the first-stage estimates and F-statistics for the joint significance of the instruments. Prior work has shown that the F-statistics from the first-stage regressions are not sufficient to dismiss weak or underidentification in models with multiple endogenous variables (Sanderson and Windmeijer, 2016). I therefore report the conditional F-statistics of Sanderson-Windmeijer. I compare these conditional F-statistics to the corresponding Stock-Yogo weak identification critical values and find that of the nine first-stage regressions that I estimate, I can reject that the variable is weakly identified in all but three cases. In these three cases, the F-statistic is larger than lowest critical value (representing 5 percent maximal IV relative bias).

I use the Sargan test of overidentifying restrictions to assess the validity of my instruments. For each of the models that I estimate using instrumental variables, I do not reject the null hypothesis of the Sargan test that the overidentifying restrictions are valid.

In Table 3, I report the production function estimates for the Cobb-Douglas and translog models estimated using either OLS, FE, or IVFE. The standard errors are not clustered given the assumption that I make over the serial correlation of the error term. In Column (7), I re-estimate the model in Column (6) without the two terms that were statistically insignificant to carry these estimates over to the calculation of the elasticities of substitution, marginal product of nurse labor, and counterfactual exercises. Column (7) displays my preferred estimates that address the endogeneity concerns with the OLS and FE estimation strategies by utilizing instruments.

[36] These findings are not inconsistent with Dingel et al. (2023): first, the instruments that I use are weak shifters of patient discharges in the first-stage because the focus of my empirical strategy is not in estimating returns to scale; second, there are differences in setting: I focus solely on acute inpatient care, which relative to outpatient services has a larger share of emergency rather than elective procedures and I aggregate across rare and common procedures. Returns to scale are larger for rare procedures which can be scheduled in advance.

Notably, my findings allow me to reject the Cobb-Douglas model which implicitly assumes that the coefficients on the squared and interaction terms are equal to zero. I employ an F-test of joint significance for the squared and interaction terms in Column (7) to assess this assumption formally and I reject the null hypothesis of this test at the one percent level. It is difficult to directly interpret the coefficient estimates but I assess the values of the R^2 in these models before moving to compute the elasticities of substitution and marginal product of nurse labor from these estimates. The low values of R^2 for the OLS estimates in Columns (1) and (4) indicate that nurse and physician labor per patient and the case mix explain very little of the variance in risk-adjusted non-readmission. The R^2 from the FE estimates indicate that nearly half of all variation in quality can be explained by hospital-specific characteristics that remain constant over the sample period. I do not include the R^2 for the instrumental variables models because they are not directly interpretable – see Sribney et al. for details.

6.1 Elasticities of Substitution

From the estimated production parameters in Table 3, I derive the structural objects of interest in this paper. The elasticities of substitution between the pairs of inputs determine the shape of the isoquant curve shown in Figure 1 and therefore determine the within-hospital misallocation due to minimum staffing ratios.

I algebraically derive the elasticity of substitution based on the definition of the "direct elasticity of substitution" from Sargan (1971) or Sato and Koizumi (1973). For clarity of notation, I have omitted the subscript from earlier notation that references the hospital-year (ht) and moved the superscript referencing the input ($i \in \{h, n, p\}$) to a subscript. [37]

$$\sigma_{np} = \frac{\mathrm{dln}\left(\dfrac{x_n}{x_p}\right)}{\mathrm{dln}\left(\dfrac{\frac{\partial Q}{\partial x_p}}{\frac{\partial Q}{\partial x_n}}\right)}$$

In the three-factor translog model, the elasticity of substitution between nurses and physicians, σ_{np}, is a complex function of the levels of the inputs (x_n, x_p, x_h) and the production parameters (β) given by[38]

$$\sigma_{np} = \frac{x_p e_n \left\{-\frac{1}{x_p}\frac{x_p e_n}{x_n e_p} - \frac{1}{x_n}\right\}}{\left\{(e_n + \beta_{np}) - \frac{x_p e_n}{x_n e_p}\frac{x_n \beta_{pp}}{x_p}\right\}\left\{-\frac{x_p e_n}{x_n e_p}\right\} + \left\{\frac{x_p \beta_{nn}}{x_n} - \frac{x_p e_n}{x_n e_p}(e_p + \beta_{np})\right\}} \tag{7}$$

[37] While the translog can be interpreted as a Taylor approximation of the CES production function, the original result in Kmenta (1967) is for two factors and limited work has been done to show the extension to the n factor case (Hoff, 2004) much less the extension to the nested n factor case. Most importantly, the Taylor approximation is a reasonable approximation of CES around the point of the approximation which is when the elasticity of substitution is close to unity (the Cobb-Douglas case) which I show is not the case in my setting.

[38] See Boisvert (1982), Appendix C for the derivation. Note that the coefficients on the squared terms in Table 3 should be multiplied by two to recover the corresponding parameters β_{nn}, β_{pp}, and β_{hh} in Equation (4).

Table 3: Production Function Estimates for Risk-Adjusted Non-Readmission

	Cobb-Douglas			Translog			
	(1) OLS	(2) FE	(3) IVFE	(4) OLS	(5) FE	(6) IVFE	(7) IVFE
Log nurses per patient	0.011***	0.014***	-0.002	0.026	0.035**	-0.028	
	(0.003)	(0.002)	(0.006)	(0.020)	(0.015)	(0.055)	
Log physicians per patient	0.004***	0.002	0.001	-0.004	-0.014**	0.110***	0.103***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.006)	(0.042)	(0.032)
Log patient health	0.006*	-0.024***	-0.018	0.054***	0.040*	0.227***	0.197***
	(0.003)	(0.008)	(0.012)	(0.020)	(0.021)	(0.065)	(0.053)
Log nurses squared				-0.014	-0.021***	-0.038*	-0.052***
				(0.010)	(0.007)	(0.022)	(0.011)
Log physicians squared				-0.000	0.001	-0.055***	-0.053***
				(0.001)	(0.002)	(0.018)	(0.013)
Log patient health squared				-0.012	-0.012	-0.104	
				(0.015)	(0.025)	(0.145)	
Log nurses x Log physicians				0.009**	0.013***	0.053***	0.056***
				(0.004)	(0.004)	(0.016)	(0.011)
Log nurses x Log patient health				-0.037**	-0.047***	-0.122**	-0.100***
				(0.017)	(0.014)	(0.049)	(0.027)
Log physicians x Log patient health				-0.006	-0.007	-0.060***	-0.047***
				(0.006)	(0.007)	(0.022)	(0.018)
Observations	2,704	2,704	2,496	2,704	2,704	2,496	2,496
R^2	0.013	0.511	–	0.018	0.518	–	–
Hospital Fixed Effects		✓	✓		✓	✓	✓
Year Fixed Effects			✓			✓	✓

Notes: This table reports the production function estimates for the Cobb-Douglas (Columns (1)-(3)) and translog (Columns (4)-(7)) production functions estimated using OLS, FE, or IVFE. Standard errors are not clustered.

where the output elasticities are equal to

$$e_n = \frac{\partial \ln(Q)}{\partial \ln(x_n)} = \beta_n + \beta_{nn}\ln(x_n) + \beta_{np}\ln(x_p) + \beta_{nh}\ln(x_h)$$
$$e_p = \frac{\partial \ln(Q)}{\partial \ln(x_p)} = \beta_p + \beta_{pp}\ln(x_p) + \beta_{np}\ln(x_n) + \beta_{ph}\ln(x_h)$$

In Table 4, I present the percentiles of the distribution of elasticities of substitution. For each hospital-year observation, I compute the elasticity of substitution implied at each of the levels of nurse staffing indicated in the first column of Table 4. The levels of physicians and patient health are data. I present the percentiles of the distribution of elasticities obtained from this exercise for each of the levels of nurse staffing indicated in the first column after excluding any hospital-year observations with negative marginal products for nurses or physicians at the specified nurse level.

The results indicate elasticities of substitution for this subset that range between zero ("perfect complements") and 0.2. The average elasticity of substitution is 0.05. The elasticity of substitution is near zero at high levels of staffing. When hospitals have a lot of nurses, a small reduction in physicians requires an almost infinite number of nurses to maintain the same level of quality. When hospitals have few nurses, nurses and physicians are more substitutable. Their roles may be more fluid when hospitals are understaffed. However, the substitutability does not change dramatically perhaps because the fluidity of roles is contingent on licensing restrictions which do not change according to the numbers of nurses or physicians. The tasks that a licensed nurse can perform remain constant regardless of how many nurses the hospital hires. The finding of strong complementarity suggests that some tasks important for preventing readmission require direction from a physician whether for skill or licensing reasons or both.

To investigate these channels further, I regress the elasticities of substitution on hospital fixed effects, number of physicians, and level of patient health, holding fixed the number of nurses. I find that the patient health is positively and statistically significantly correlated with the elasticity of substitution at each level of nurse staffing. Nurses and physicians are more substitutable when patients are healthy. This could either be because a higher proportion of tasks that need to be performed for healthy patients overlap between nurses and physicians or because nurses are better positioned in terms of their skill set to handle the overlapping tasks. These findings are consistent with Chan and Chen (2022)'s findings of a smaller quality gap between independently practicing nurses and physicians when the patients they treat are lower severity.

In Appendix Table A.7, I present the computed elasticities of substitution between nurses and patient health. As in Table 4, I compute the elasticity for each hospital-year observation taking physicians and patient health as data and productivity estimates and present the percentiles of the resulting distribution after omitting hospital-year observations with negative marginal products for either input. The average elasticity of

Table 4: Elasticities of Substitution - Nurses and Physicians

	Percentiles of Distribution				
Nurses per 1,000 Patient Days	10th	25th	50th	75th	90th
1.5	0.020	0.049	0.091	0.132	0.168
2	0.012	0.031	0.062	0.106	0.135
2.5	0.009	0.023	0.043	0.078	0.102
3	0.013	0.026	0.051	0.085	0.132
3.5	0.012	0.024	0.058	0.096	0.112
4	0.002	0.010	0.044	0.075	0.086

Notes: In this table, I present the percentiles of the distribution of elasticities of substitution derived using Equation (7) for each hospital-year in my sample with positive marginal products for nurses and physicians. The near zero elasticities of substitution indicate strong complementarities in quality production between nurses and physicians. Nurses and physicians are more substitutable at low levels of the two inputs i.e. when hospitals are relatively understaffed.

substitution for this subset of observations is 0.56 indicating substitutability between the factors. Appendix Table A.7 indicates that at low levels of staffing nurses and patient health are highly substitutable with an elasticity above seven for the top quantile but that the substitutability diminishes rapidly as staffing increases. There is a much broader range of values for the elasticity of substitution between patient health and nurses per patient compared to the elasticity between nurses and physicians.

6.2 The Marginal Product of Nurse Labor

In addition to deriving the elasticities of substitution, I derive the marginal product of nurse labor given by

$$\frac{\partial Q}{\partial (x_n)} = \left[\frac{\partial \ln(Q)}{\partial \ln(x_n)}\right]\left[\frac{Q}{x_n}\right] = (\beta_n + \beta_{nn}\ln(x_n) + \beta_{np}\ln(x_p) + \beta_{nh}\ln(x_h))(\frac{Q}{x_n}) \tag{8}$$

For each hospital-year observation, I compute the marginal product at each level of nurse staffing from a grid, physicians and patient health are data, and the productivity estimated from Equation (7). I plot the percentiles of the distribution in Figure 4 with the line denoting the threshold of the California nurse staffing mandate. Figure 4 suggests wide dispersion in the quality effects from minimum staffing ratios with negative effects for the bottom 25 percent of observations. These negative effects may be due to well-documented issues with nursing handoffs or an increase in the span of control for physicians and nursing team leaders that supervise and monitor nursing teams.

In Figures 5a and 5b, I compute the marginal products using fixed levels of physicians or patient health, respectively, rather than observed levels in the data. Given the limited shift in physician levels at treated hospitals following the mandate (Figure A.4a), the marginal product curves shown in Figure 5a imply that the quality gains were driven by hospitals that had inefficiently low nurse to physician ratios. Nearly all hospitals based on their patient health levels gain from adding nurses around the mandate threshold but

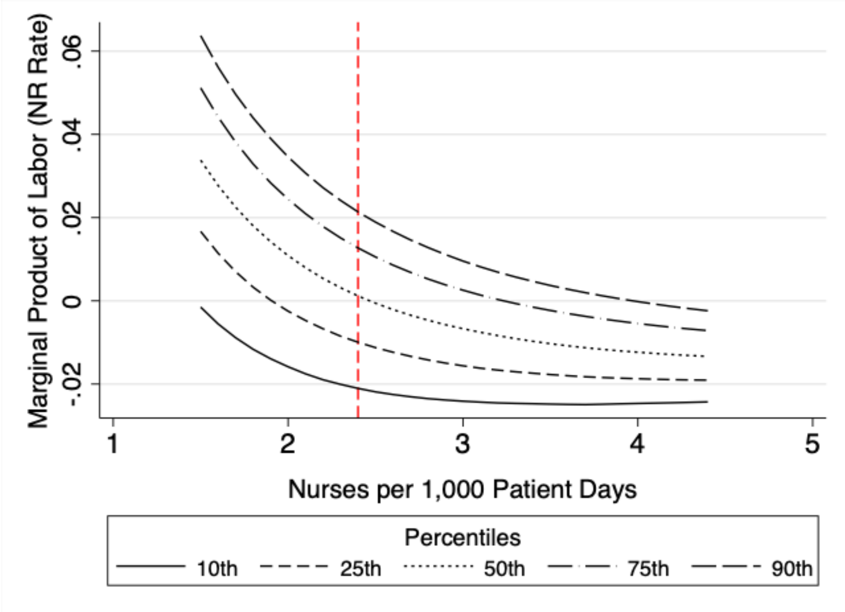


Figure 4: Distribution of the Marginal Product of Nurse Labor

Notes: In this figure, I plot the percentiles of the marginal product of nurse labor distribution. The marginal product is measured in units of the risk-adjusted non-readmission rate meaning a marginal product of 0.04 implies that a one unit increase in nurses per 1,000 patient days at the specified level would lead to a 0.04 point increase in the rate (the mean is 0.975 in the data so this corresponds to approximately 4 percent) if there is no change in physicians or patient health. I compute the marginal product of nurse labor for each point on the nurse grid taking the physicians, patient health, and productivity as data and plot the percentiles of the resulting distribution.

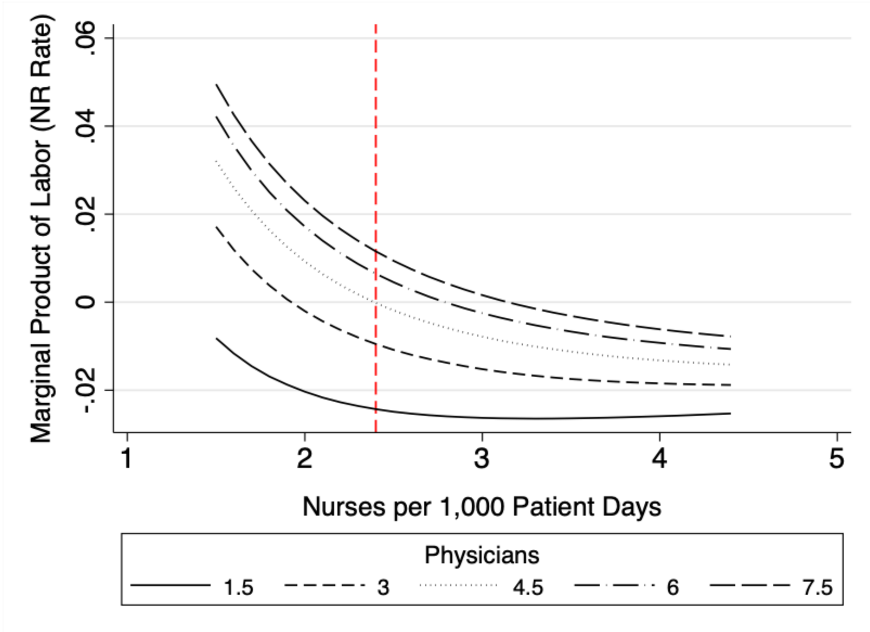
hospitals with sicker patients gain more. Hospitals allocate nurses across inpatient units in a manner that is consistent with this finding: staffing ratios are much higher in intensive care than in acute care because the hospital knows the quality returns to nursing are higher in intensive care.

Figures 5a and 5b show that marginal product of nurse labor is increasing in the level of physicians and decreasing in the level of patient health – by Seidman (1989)'s definition of "q-complements", nurses and physicians are complements and nurses and patient health are substitutes. Nurses are more valuable when they operate in an environment with physicians and with more severe patients.

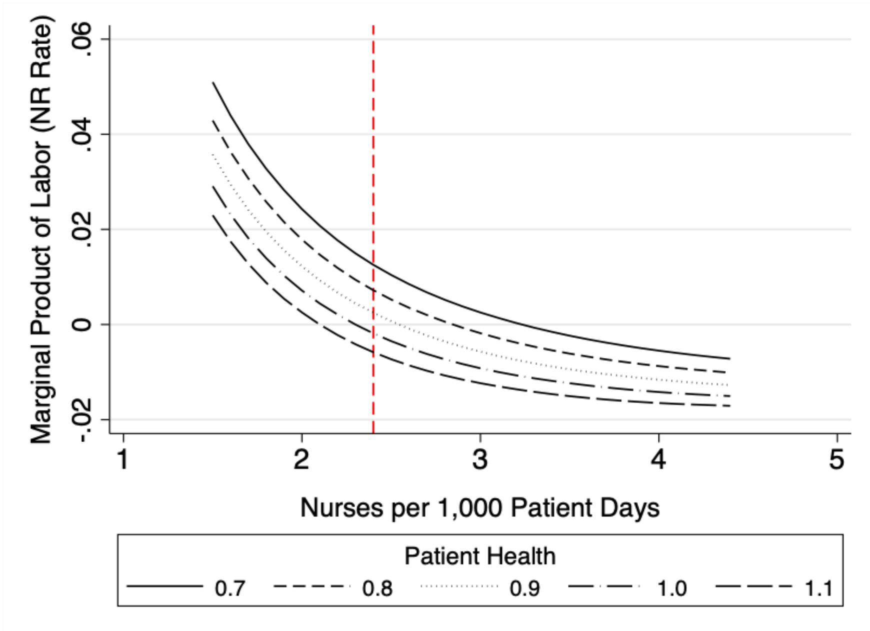
6.3 Model Validation Using Reduced-Form Effect of Mandate

Given that the estimated reduced-form treatment effects are not used in the estimation of the structural model, I use the reduced-form results to validate the model's results. The idea behind the validation exercise is replicate Table 2 using fitted values of quality from the structural model. I construct fitted values for the treatment and control groups at two periods (pre- and post-mandate) and estimate the difference-in-differences models from Table 2 using the fitted values of quality as the dependent variable.

For both treated and control hospitals in 2002 (pre-mandate), I compute the fitted



(a) Marginal Product by Physician Level



(b) Marginal Product by Patient Health Level

Figure 5: Heterogeneity in Marginal Product by Levels of Other Inputs

Notes: In panel (a), this figure shows the 50th percentile of the marginal product curve for each indicated level of physicians per 1,000 patient days. In panel (b), I do the same for patient health. The marginal product is measured in units of the risk-adjusted non-readmission rate meaning a marginal product of 0.04 implies that a one unit increase in nurses per 1,000 patient days at the specified level would lead to a 0.04 point increase in the rate (the mean is 0.975 in the data so this corresponds to approximately 4 percent) if there is no change in physicians or patient health. The marginal product of nurse labor is increasing in physicians and decreasing in patient health.

rate as follows where χ is the vector of parameters estimated from the structural model

$$Q_{h,pre} = Q(\chi, x^n_{h,2002}, x^p_{h,2002}, x^h_{h,2002}) \tag{9}$$

In 2006 (post-mandate), I compute the fitted rate using the observed values of the other inputs and the observed value of nurse staffing in 2002 (for control hospitals) or the observed value times the reduced-form estimate of the average treatment effect on nurse staffing, $\widehat{n_{DiD}}$, estimated from the specification in Equation (2) (for treated hospitals). The reduced-form estimate $\widehat{n_{DiD}}$ is roughly a 11 percent increase in nurse staffing which is consistent with the findings in Raja (2023) and earlier work. For control hospitals $\widehat{n_{DiD}}$ is set equal to zero.

$$Q_{h,post} = Q(\chi, x^n_{h,2002} + x^n_{h,2002} * \widehat{n_{DiD}}, x^p_{h,2006}, x^h_{h,2006}) \tag{10}$$

Given fitted values for each hospital in the treatment and control groups for two time periods, I estimate a version of Table 2 using the fitted values as the dependent variable. In Table 5, I present the results of this exercise. Table 5, Column (1) estimates a treatment effect of 0.6 percent compared to 0.5 percent in Table 2, Column (1). The heterogeneous treatment effect by physician level estimated in Column (2) implies that the treatment effect is decreasing in the number of physicians contrary to what we find in Table 2. The heterogeneous treatment effect by patient health estimated in Column (3) is consistent in magnitude and direction to Table 2.

The results in the two tables should not be expected to be identical for a few reasons: smaller sample size in Table 5 compared to Table 2; the fitted values are constructed by applying the average treatment effect on nurse staffing to all hospitals whereas in reality there was heterogeneous incidence of the mandate on staffing; and importantly for Columns (2) and (3) the relationship between the physician and patient health levels and fitted quality in Table 5 are identified whereas in Table 2 the relationships are correlations. Related to the last point, both physician nor patient health levels are estimated to be near zero and statistically insignificant in Table 2 whereas this is not the case in Table 5.

Table 5: Structural Model Replication of Table 2

	(1) Log NR Rate	(2) Log NR Rate	(3) Log NR Rate
Treat x Post	0.006*	0.009	0.004
	(0.003)	(0.006)	(0.004)
Post	0.002	0.003	0.003
	(0.003)	(0.003)	(0.003)
Treat x Post x Log physicians per patient		-0.002	
		(0.003)	
Log physicians per patient		0.007**	
		(0.004)	
Treat x Post x Log patient health			-0.019
			(0.012)
Log patient health			0.014
			(0.020)
Observations	416	416	416
R^2	0.846	0.849	0.848
Mean	0.979	0.979	0.979
Hospital Fixed Effects	✓	✓	✓
Year Fixed Effects			

Notes: This table presents the structural model replication of Table 2. The dependent variable is the log of the fitted risk-adjusted non-readmission rate from the structural model as described in Section 5.3. There are two periods in this structural replication (years 2002 and 2006) with the Post period indicator taking on a value of one for the 2006 observations.

7 Misallocation

7.1 Misallocation Within Hospitals

Motivated by the implications of the estimated elasticities of substitution and marginal product of labor, I use the estimated production parameters to evaluate the magnitude of within-hospital misallocation between nurses and physicians that arises from using a nurse staffing mandate relative to a direct quality mandate.

To quantify the magnitude of the within-hospital misallocation between nurses and physicians, I assume a static cost-minimization problem. For each of the 208 hospitals in my sample, I solve the hospital's cost-minimization problem below under three scenarios given its observed patient health, patient volumes, nurse wage, annual physician salary, and estimated productivity. In the first scenario ("pre-period cost minimizing"), I set the minimum quality constraint of the cost-minimization problem equal to the hospital's observed quality in the pre-mandate period. In the second scenario ("post-period cost minimizing"), I set the minimum quality constraint equal to a 0.5 percent increase from the observed quality in the pre-mandate period. This is intended to mimic the average quality effect of the mandate estimated by the reduced-form and structural models. However, in the "post-period cost minimizing" scenario hospitals are allowed to choose their allocation of nurses and physicians. The increase in quality constraint between the first to second scenarios should be interpreted as the imposition of a regulation that achieves the first-best outcome – assuming that a 0.5 percent increase in quality is the socially optimal amount for each hospital to produce. The hospital's problem for the first two scenarios is given by

$$\begin{aligned} &\min_{x^n_{ht}, x^p_{ht}} \{w^n_{ht} * x^n_{ht} + w^p_{ht} * x^p_{ht}\} \\ &\quad \text{s.t. } e^{\omega_{ht}+\epsilon_{ht}} F(x^h_{ht}, x^n_{ht}, x^p_{ht}) \geq Q_{ht} \end{aligned}$$

where w^n_{ht} and w^p_{ht} represent the average hourly nurse wage and annual physician salary, respectively and for ease of exposition the input variables have been transformed to equal the number of hours of clinical nursing time per year (x^n) and the number of active medical staff physicians per year (x^p) based on the patient volumes in the hospital-year.

In the third scenario ("post-period mandate"), I set the minimum quality constraint equal to a 0.5 percent increase in the observed quality in the pre-mandate period (same as in the second scenario) and impose a second constraint representing the minimum staffing requirement on nurses. The level of this constraint is set based on the incidence of the observed mandated threshold relative to the observed distribution of nurse-to-patient ratios prior to the mandate. I compare the input allocations and variable costs between the second and third scenarios to assess the within-hospital misallocation between nurses and physicians as a consequence of the mandate.

$$\begin{aligned} &\min_{x^n_{ht}, x^p_{ht}} \{w^n_{ht} * x^n_{ht} + w^p_{ht} * x^p_{ht}\} \\ &\quad \text{s.t. } e^{\omega_{ht}+\epsilon_{ht}} F(x^h_{ht}, x^n_{ht}, x^p_{ht}) \geq Q_{ht} \\ &\qquad x^n_{ht} \geq x^n_{min} \end{aligned}$$

Given that I do not observe physician wages in my data,[39] I use data from Gottlieb et al. (2023) who use IRS tax records to report summary statistics on physicians' individual total income, inclusive of business income, at the commuting zone level. The commuting zone level averages for 2017 wages are shown in Figure E.5 of the Online Appendix to Gottlieb et al. (2023). I adjust these 2017 wages to 2005 wages (denominated in 2017 USD) at the commuting zone level using the growth rate in physician salaries implied by Figure E.3(A). I then inflation-adjust the 2005 wages to be denominated in 2005 USD. Given the assumption that the average medical staff physician is 0.4 FTEs (discussed in Section 2), I multiply the wage by 0.4 to obtain the wage of an affiliated physician.

Prior to highlighting the misallocation results, I present in Appendix Table A.8 the results from the model and the data absent any regulation to illustrate the model fit

[39]Most physicians own or are employed by physician practice groups that contract with hospitals. Physicians' salaries are therefore not reported in the financial reporting data the way that direct employees' salaries are reported. Their financial contracts with hospitals and any payments from hospitals to physicians under these contracts may or may not need to be reported in the form. Furthermore, the nature of these contracts vary so greatly across and within hospitals (across hospital units) that one cannot be sure that reported payments, if they are observed, are made for physician labor as opposed to hospital reimbursements for expenditures made by physicians who in some cases supply their own inputs to the unit they staff.

with respect to the data. In the final two columns of Appendix Table A.8, I present the predicted levels of nurses and physicians from the cost-minimization model required to produce the observed quality at the hospital level in the pre-mandate period (top panel) and the observed levels of nurses and physicians in the pre-mandate period (bottom panel). Each cell represents an average value for the hospitals that fall into the staffing quartile reported in the first column. The final two columns in the table indicate that the cost-minimization model reports that hospitals can produce the observed level of quality with far fewer nurses and physicians than observed in the data.[40] At the same time, the model and the data show similar ordinal rankings of hospitals based on their characteristics. In both the model and the data, the high staffing hospitals have low patient health, high volume, and similar productivity levels.

In Table 6, I present the results from the cost minimization exercise. The model predicts that one-fifth of hospitals cannot increase quality by 0.5 percent using even an infinite amount of nurse and physician levels. For these hospitals, the regulation is highly misallocative as it leads to no quality returns. For the remainder of the hospitals, I present the results from "post-period cost minimizing" and "post-period mandate" scenarios in the first and second panels and the implied misallocation in the third panel of Table 6. The incidence quartile in the first column of the table reflects the incidence of the mandate relative to the hospital's staffing level absent any regulation (shown in the "% Incidence" column of the bottom panel). Table 6 indicates that the hospitals that were lowest staffing prior to the mandate and consequently in the highest incidence quantiles were a combination of high productivity, high patient health, and high nurse wage relative to physician wage. Notably, the model indicates that the incidence of the mandate fell on hospitals that were not the lowest quality hospitals which matches the stylized fact in Table 1 that there is little correlation between pre-mandate staffing and quality.

The "post-period mandate" scenario in the second panel imposes a minimum nurse staffing ratio set at the median of the pre-period cost minimizing level (1.41 nurses per 1,000 patient days). This placement of the threshold mirrors the incidence of the mandate which was set at roughly the median of the pre-period staffing distribution. The change in the nurse to physician ratios between the cost-minimizing and mandate scenarios illustrates the source of the misallocation: under the cost-minimizing scenarios hospitals prefer to stick to a ratio between 2.70 and 2.83 whereas the mandate requires them to deviate from this proportion and overutilize nurses. The hospitals with the largest incidence from the mandate have the largest deviations in the nurse to physician ratio.

If we exclude the untreated hospitals with zero incidence, the interquartile range is 2

[40] The mismatch between the cost-minimizing and observed allocations may reflect that hospitals' input choices have dynamic implications or that hospitals' input choices are made based on non-quality objectives. The staffing gap between the cost-minimizing and observed allocations is larger for low-volume hospitals, for example, which experience notably larger variance in day-to-day patient volumes and case mix and are therefore inefficient from a cost-minimization standpoint (Dalton et al., 2003). Additionally, from a model fit perspective this exercise requires the model to extrapolate away from the observed input levels into areas where it may have poor fit because, for example, I do not observe any large negative shocks to nursing in my data. This is why the exercise is not intended for use to prescribe a staffing level but to compare the sets of results within the model.

Table 6: Within-Hospital Misallocation Results

	Cost-Minimization Scenario				
Quartile	Nurses	Phys.	Nurse-Phys.	Costs p.p.d.	Quality
Top 25	2.28	0.81	2.83	508	0.987
50-75	1.56	0.58	2.70	359	0.980
25-50	1.18	0.44	2.71	248	0.976
Bottom 25	0.77	0.28	2.76	159	0.981
	Mandate Scenario				
Quartile	Nurses	Phys.	Nurse-Phys.	Costs p.p.d.	Quality
Top 25	2.28	0.81	2.83	508	0.987
50-75	1.57	0.58	2.71	359	0.980
25-50	1.41	0.42	3.31	258	0.976
Bottom 25	1.41	0.28	5.06	203	0.981
	Misallocation				
	% Incidence	% Diff. Costs	% Diff. Nurses	% Diff. Phys.	Diff. Ratio
Top 25	–26.53	0.00	0.00	–0.00	0.00
50-75	1.94	0.03	0.46	–0.15	0.02
25-50	31.97	4.57	19.15	–2.60	0.61
Bottom 25	106.48	32.83	83.69	0.38	2.29

Notes: In the top panel, this table shows the cost-minimizing allocations of nurses and physicians. The efficient ratio of nurses to physicians is between 2.70-2.83. In the second panel, I show the allocations under the mandate. The last panel shows the within-hospital misallocation between the two scenarios.

to 21 percent of variable costs. The misallocation for a hospital with an average incidence of the California staffing mandate is 1.4 percent of the total variable costs of nurses and physicians which amounts to roughly $180,000 USD for the average hospital and aggregates to $24 million across treated hospitals. This is an underestimate given the one-fifth of hospitals that are unable to make the required quality gains using any combination of nurses and physicians. The hospitals in the highest incidence quartiles have higher patient health, higher productivity, fewer patient days, and relatively high nurse salaries. Many of these attributes characterize rural hospitals. It should be noted that the California mandate allowed waivers for small and rural hospitals and roughly half of these hospitals received exemptions though I am unable to observe which ones (Raja, 2023). My model predicts that this type of exemption targeted at rural hospitals would mitigate the misallocation at the highest incidence hospitals.

Importantly, this exercise estimates the degree of misallocation relative to the efficient benchmark. If hospitals were operating inefficiently in terms of their nurse to physician ratios prior to the mandate, then the degree of observed misallocation would differ. The model implies that the efficient ratio of full-time equivalent nurses to full-time equivalent physicians engaged in patient care is around 2.70-2.90. Prior to the mandate, I estimate that the mean of this ratio among treated hospitals was 3.12 implying that on average the within-hospital misallocation that I calculate is an underestimate. Furthermore the distribution of the nurse to physician ratio was right-skewed with an interquartile range of 1.51 to 4.02 indicating a significant number of hospitals for whom the nurse to physician ratio was already inefficiently high. As discussed in Section 6, the average quality gains from the mandate were driven by hospitals where the nurse to physician ratio was inefficiently low and consequently these hospitals experience less misallocation due to the regulation.

7.2 Misallocation Across Hospitals

In addition to the within-hospital misallocation, I am interested in the across-hospital misallocation that arises due to the heterogeneity in marginal products across treated and untreated hospitals. If we hold fixed the number of nurses added due to the regulation and change their allocation across hospitals could we produce higher quality of care?

I consider the counterfactual that the nurses added to treated hospitals are instead added to the nearest untreated hospital within 10 miles. I find that 89 of the 134 treated hospitals have another hospital (treated or untreated) within 10 miles and only 38 have an untreated hospital within 10 miles. Several have more than one untreated neighbor. For the roughly one-third of treated hospitals with an untreated neighbor, I calculate the fitted quality in the pre- and post-mandate periods corresponding to Equations (9) and (10) in Section 6.3. I assume the treatment effect for both treated and untreated hospitals is equal to 10 percent of the number of nurses in the pre-mandate period at the treated hospital for each pair.

For the treated hospitals, my model implies an average 1.1 percent gain in quality from the mandate compared to an average 0.7 percent counterfactual gain at the untreated hospitals if they employed the number of nurses added to the treated hospital (0.4 percentage point differential). There are, however, allocative improvements to be made for specific pairs of hospitals located in densely populated counties where the subset of untreated hospitals admit higher severity patients. The untreated hospitals in the pairs where allocative improvements can be made have an average patient health level of 0.81 compared to 1.02 at untreated hospitals in the pairs where allocative improvements cannot be made. The nurse staffing levels at these untreated hospitals are similar.

The average quality differential can be due to differentials on one or more dimensions: the levels of physicians, patient health, or nurses or the hospital's productivity.[41] To determine the quantitative importance of each channel, I shut down one channel at a time and reassess the difference in the marginal products. The distribution of the marginal product differentials from conducting this exercise is shown in Table 7. Positive values indicate that there are quality gains to be made from adding nurses to substitute hospitals instead of treated hospitals. In the first row, I allow the marginal products to differ on all dimensions using the observed data on patient health, nurse, and physician levels and estimated productivities. In each of the subsequent rows, I set each of the variables for the untreated hospital equal to its value for the treated hospital that it is paired to. For

[41] As a placebo test, I estimated the difference-in-differences model from Section 3 on the productivities estimated from the production model and I do not find any treatment effect of the mandate on productivity.

example, in the second row I set the number of physicians at each untreated hospital to equal the number of physicians at the treated hospital that it is paired to. The average values of nurses, physicians, patient health, and productivity are displayed in the columns adjacent to the distribution and reflect these changes.

The hospital characteristics in the first row, which reflect the data without any changes, indicate that on average untreated hospitals have lower nurse levels and higher physician levels but similar levels of patient health and productivity. Looking at the distribution of the marginal product differentials across the rows, it is clear that the lower nurse levels and higher physician levels at treated hospitals relative to untreated drive the difference in quality gains from allocating nurses to the treated hospitals. When each of these channels is shut down in rows five and two, respectively, the differential is reduced substantially.

Importantly, I find a very weak correlation between staffing levels and productivity. In Table A.10, I present regressions of quality and estimated hospital productivity on hospital types including teaching, small and rural, and not-for-profit, government-owned, or investor-owned. In Table A.9, I present regressions of observed input use on estimated hospital productivity. In Column (1), show that a one percent increase in productivity is correlated with a 0.2 percent increase in nurse staffing and that variation in productivity explains less than one percent of the variation in nurse staffing across hospitals in a given year. The same applies to physician staffing in Column (2). The lack of correlation between staffing and productivity suggests that input choices are driven by pre-existing distortions in incentives to hire across hospitals rather than heterogeneous returns to input use. This suggests that regulatory design aimed at low staffing hospitals may correct pre-existing distortions in incentives to invest in quality.

7.3 Conclusion

Minimum nurse-to-patient ratios for hospitals are under legislative consideration in several states and at the federal level in the U.S. and have the potential to dramatically change the way patients receive inpatient medical care. Ratio regulation is in theory inefficient relative to direct quality regulation and may be inefficient along a second dimension if it allocates nurses to low productivity hospitals. "How" inefficient ratio regulation is along these two dimensions is an empirical question and depends on the interactions between nurses and other inputs and the productivity of low staffing hospitals.

We have a limited understanding of how nurses, physicians, and patients interact to produce hospital quality. I address this gap in the literature using methods from the industrial organization literature on production functions: I estimate a value-added production model in nurses per patient, physicians per patient, and patient health using administrative patient-level discharge data to construct hospital quality and detailed financial reporting data to measure nurse and physician labor. Importantly, I address the endogeneity of inputs to the hospital's productivity using quasi-experimental variation in nurse labor due to the 1999 California nurse staffing mandate for identification.

Table 7: Difference in Marginal Products of Nurse Labor - Treated Hospitals and Nearby Untreated Hospitals

	Distribution					Nurses		Physicians		Patient Health		Productivity	
	10th	25th	50th	75th	90th	Treat	Untrtd	Treat	Untrtd	Treat	Untrtd	Treat	Untrtd
As is in the data	-0.038	-0.027	-0.013	0.001	0.011	2.39	3.12	1.31	1.04	0.92	0.92	0.92	0.94
Fix physicians only	-0.031	-0.018	-0.007	0.002	0.008	2.39	3.12	1.31	1.31	0.92	0.92	0.92	0.94
Fix health only	-0.034	-0.022	-0.012	-0.001	0.006	2.39	3.12	1.31	1.04	0.92	0.92	0.92	0.94
Fix productivity only	-0.037	-0.027	-0.012	0.001	0.011	2.39	3.12	1.31	1.04	0.92	0.92	0.92	0.92
Fix nurse level only	-0.032	-0.014	-0.004	0.009	0.021	2.39	2.39	1.31	1.04	0.92	0.92	0.92	0.94

Notes: This table presents the results of the counterfactual exercise that considers the addition of nurses to untreated hospitals within 10 miles of a hospital treated by the mandate. It presents the distribution of the marginal product gaps between each of the 52 pairs of hospitals (treated and untreated) in the counterfactual exercise when nurse, physician, patient health, and productivity levels are allowed to vary (in the first row) and then when one variable for the untreated hospital is conformed to the treated hospital level (in rows two through five). Negative values of the quantiles indicate that the marginal product of nurse labor is higher at the treated hospital than its nearby counterpart. The table indicates that the higher marginal product at the treated hospitals is due largely to the lower nurse levels at these hospitals to begin with rather than productivity differences across the two sets of hospitals.

I find that nurses and physicians are highly complementary (near Leontief) in the production of quality and these complementarities imply inefficiencies in using minimum nurse-to-patient ratios to regulate quality on the order of $24 million across hospitals treated by the mandate. The efficient solution is to use nurses and physicians in proportion to one another which suggests increasing physician labor in a context where its aggregate supply is controlled. An alternative approach is the one that has been employed in the U.S. in the fifteen years since the end of my sample period – modifying the underlying production primitive (elasticity of substitution between nurses and physicians) through rollbacks on licensing restrictions for nurses and expansions in the supply of higher-skilled nurses such as Nurse Practitioners.

I show that the nurse's marginal product in quality units is larger when the patients they treat are higher severity and a mandate which does not account for differences in patient mix across hospitals leads to allocative inefficiency for specific pairs of hospitals where nurses are added to lower staffing hospitals with healthier patients. On average, however, I do not find evidence of productivity differences across low and high staffing hospitals which suggests that variation in hospitals' input choices are driven by heterogeneous incentives rather than heterogeneous returns to those inputs.

References

Fill rates and labor market activity : The Economics Daily: U.S. Bureau of Labor Statistics. Technical report, U.S. Bureau of Labor Statistics, Nov. 2019. URL https://www.bls.gov/opub/ted/2019/fill-rates-and-labor-market-activity.htm.

NSI National Health Care Retention and RN Staffing Report. Technical report, NSI Nursing Solutions, Inc., Mar. 2024. URL https://www.nsinursingsolutions.com/Documents/Library/NSI_National_Health_Care_Retention_Report.pdf.

D. A. Ackerberg, K. Caves, and G. Frazer. Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451, 2015. ISSN 1468-0262. DOI: 10.3982/ECTA13408. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13408. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13408.

J. Asker, A. Collard-Wexler, and J. De Loecker. (Mis)Allocation, Market Power, and Global Oil Extraction. *American Economic Review*, 109(4):1568–1615, Apr. 2019. ISSN 0002-8282. DOI: 10.1257/aer.20171438. URL https://www.aeaweb.org/articles?id=10.1257/aer.20171438.

A. P. Bartel, N. D. Beaulieu, C. S. Phibbs, and P. W. Stone. Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector. *American Economic Journal: Applied Economics*, 6(2):231–259, Apr. 2014. ISSN 1945-7782, 1945-7790. DOI: 10.1257/app.6.2.231. URL https://pubs.aeaweb.org/doi/10.1257/app.6.2.231.

E. R. Berndt and L. R. Christensen. The translog function and the substitution of equipment, structures, and labor in U.S. manufacturing 1929-68. *Journal of Econometrics*, 1(1):81–113, Mar. 1973. ISSN 0304-4076. DOI: 10.1016/0304-4076(73)90007-9. URL https://www.sciencedirect.com/science/article/pii/0304407673900079.

N. Bloom, C. Propper, S. Seiler, and J. Van Reenen. The Impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies*, 82(2):457–489, Apr. 2015. ISSN 0034-6527. DOI: 10.1093/restud/rdu045. URL https://doi.org/10.1093/restud/rdu045.

R. N. Boisvert, editor. *The Translog Production Function: Its Properties, Its Several Interpretations and Estimation Problems.* Research Bulletin. 1982. DOI: 10.22004/ag.econ.182035.

S. Bond and M. Soderbom. Adjustment costs and the identification of Cobb Douglas production functions, Jan. 2005. URL https://ifs.org.uk/publications/adjustment-costs-and-identification-cobb-douglas-production-functions.

J. Chan, David C. and Y. Chen. The Productivity of Professions: Evidence from the Emergency Department, Oct. 2022. URL https://www.nber.org/papers/w30608.

A. Chandra and J. Skinner. Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, 50(3):645–680, July 2012. ISSN 0022-0515. DOI: 10.1257/jel.50.3.645. URL https://www.aeaweb.org/articles?id=10.1257/jel.50.3.645.

A. Chandra and D. O. Staiger. Identifying Sources of Inefficiency in Healthcare*. *The Quarterly Journal of Economics*, 135(2):785–843, May 2020. ISSN 0033-5533, 1531-4650. DOI: 10.1093/qje/qjz040. URL https://academic.oup.com/qje/article/135/2/785/5698324.

A. Chandra, A. Finkelstein, A. Sacarny, and C. Syverson. Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector. *The American Economic Review*, 106(8):2110–2144, 2016a. ISSN 0002-8282. URL https://www.jstor.org/stable/43956908. Publisher: American Economic Association.

A. Chandra, A. Finkelstein, A. Sacarny, and C. Syverson. Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector. *American Economic Review*, 106(8):2110–2144, Aug. 2016b. ISSN 0002-8282. DOI: 10.1257/aer.20151080. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151080.

A. Chandra, A. Finkelstein, A. Sacarny, and C. Syverson. Productivity Dispersion in Medicine and Manufacturing. *American Economic Review*, 106(5):99–103, May 2016c. ISSN 0002-8282. DOI: 10.1257/aer.p20161024. URL https://www.aeaweb.org/articles?id=10.1257/aer.p20161024.

A. Chandra, C. Colla, and J. Skinner. Productivity Variation and Input Misallocation: Evidence from Hospitals. Technical Report w31569, National Bureau of Economic Research, Cambridge, MA, Aug. 2023a. URL http://www.nber.org/papers/w31569.pdf.

A. Chandra, M. Dalton, and D. Staiger. Are Hospital Quality Indicators Causal? Technical Report w31789, National Bureau of Economic Research, Cambridge, MA, Oct. 2023b. URL http://www.nber.org/papers/w31789.pdf.

A. Cook, M. Gaynor, M. Stephens, and L. Taylor. The effect of a hospital nurse staffing mandate on patient health outcomes: evidence from California's minimum staffing regulation. *Journal of Health Economics*, 31(2):340–348, Mar. 2012. ISSN 1879-1646. DOI: 10.1016/j.jhealeco.2012.01.005.

K. Dalton, M. Holmes, R. Slifkin, and C. Sheps. Unpredictable Demand and Low-Volume Hospitals. 2003. URL https://www.semanticscholar.org/paper/Unpredictable-Demand-and-Low-Volume-Hospitals-Dalton-Holmes/aaa50fc7dabcb51460b1135e7817301b10dc9bf6.

W. E. Diewert. Hicks' Aggregation Theorem and the Existence of a Real Value-Added Function*. In M. Fuss and D. McFADDEN, editors, *Contributions to Economic Analysis*, volume 2 of *Applications of the Theory of Production*, pages 17–51. Elsevier, Jan. 1978. DOI: 10.1016/B978-0-444-85013-3.50009-7. URL https://www.sciencedirect.com/science/article/pii/B9780444850133500097.

J. I. Dingel, J. D. Gottlieb, M. Lozinski, and P. Mourot. Market Size and Trade in Medical Services, Mar. 2023. URL https://www.nber.org/papers/w31030.

L. Einav, A. Finkelstein, and N. Mahoney. Producing Health: Measuring Value Added of Nursing Homes. Technical Report w30228, National Bureau of Economic Research, Cambridge, MA, Aug. 2022. URL http://www.nber.org/papers/w30228.pdf.

C. E. Ferguson. *The neoclassical theory of production and distribution / [by] C.E. Ferguson.* Cambridge U.P., London, 1969. ISBN 978-0-521-07453-7.

E. Fisher, D. Goodman, J. Skinner, and K. Bronner. Health Care Spending, Quality, and Outcomes. Technical report, Dartmouth Atlas Project, 2009.

B. U. Friedrich and M. B. Hackmann. The Returns to Nursing: Evidence from a Parental-Leave Program. *The Review of Economic Studies*, 88(5):2308–2343, Oct. 2021. ISSN 0034-6527. DOI: 10.1093/restud/rdaa082. URL https://doi.org/10.1093/restud/rdaa082.

A. Gandhi, S. Navarro, and D. Rivers. How Heterogeneous is Productivity? A Comparison of Gross Output and Value Added. *University of Western Ontario, Centre for Human Capital and Productivity (CHCP) Working Papers*, 2017. URL https://ideas.repec.org//p/uwo/hcuwoc/201727.html. Number: 201727 Publisher: University of Western Ontario, Centre for Human Capital and Productivity (CHCP).

A. M. Garber and J. Skinner. Is American Health Care Uniquely Inefficient? *Journal of Economic Perspectives*, 22(4):27–50, Dec. 2008. ISSN 0895-3309. DOI: 10.1257/jep.22.4.27. URL https://www.aeaweb.org/articles?id=10.1257/jep.22.4.27.

M. Gaynor and R. J. Town. Competition in Health Care Markets, July 2011. URL https://www.nber.org/papers/w17208.

P. J. Gertler and D. M. Waldman. Quality-adjusted Cost Functions and Policy Evaluation in the Nursing Home Industry. *Journal of Political Economy*, 100(6):1232–1256, 1992. ISSN 0022-3808. URL https://www.jstor.org/stable/2138832. Publisher: University of Chicago Press.

J. D. Gottlieb, M. Polyakova, K. Rinz, H. Shiplett, and V. Udalova. Who Values Human Capitalists' Human Capital? The Earnings and Labor Supply of U.S. Physicians, July 2023. URL https://www.nber.org/papers/w31469.

A. Gradney. AAPPR Report Shows Demand for Physicians Reaching a New High as Staffing Shortage Continues, Oct. 2023. URL https://aappr.org/2023/10/24/aappr-2023-benchmarking-report-press-release/.

P. L. E. Grieco and R. C. McDevitt. Productivity and Quality in Health Care: Evidence from the Dialysis Industry. *The Review of Economic Studies*, 84(3):1071–1105, July 2017. ISSN 0034-6527. DOI: 10.1093/restud/rdw042. URL https://doi.org/10.1093/restud/rdw042.

J. Gruber and S. A. Kleiner. Do Strikes Kill? Evidence from New York State. *American Economic Journal: Economic Policy*, 4(1):127–157, Feb. 2012. ISSN 1945-7731. DOI: 10.1257/pol.4.1.127. URL https://www.aeaweb.org/articles?id=10.1257/pol.4.1.127.

A. Gupta. Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program. *American Economic Review*, 111(4):1241–1283, Apr. 2021. ISSN 0002-8282. DOI: 10.1257/aer.20171825. URL https://www.aeaweb.org/articles?id=10.1257/aer.20171825.

A. Gupta, S. T. Howell, C. Yannelis, and A. Gupta. Does Private Equity Investment in Healthcare Benefit Patients? Evidence from Nursing Homes. Working Paper 28474, National Bureau of Economic Research, Feb. 2021. URL https://www.nber.org/papers/w28474. Series: Working Paper Series.

D. S. Havens, J. Vasey, J. H. Gitl, and W.-T. Lin. Relational coordination among nurses and other providers: impact on the quality of patient care. *Journal of Nursing Management*, 18(8):926–937, 2010. ISSN 1365-2834. DOI: 10.1111/j.1365-2834.2010.01138.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2834.2010.01138.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2834.2010.01138.x.

K. Ho and R. S. Lee. Insurer Competition in Health Care Markets. *Econometrica*, 85(2):379–417, 2017. ISSN 1468-0262. DOI: 10.3982/ECTA13570. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13570. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13570.

A. Hoff. The Linear Approximation of the CES Function with n Input Variables. *Marine Resource Economics*, 19(3):295–306, 2004. ISSN 0738-1360. URL https://www.jstor.org/stable/42629436. Publisher: [MRE Foundation, Inc., University of Chicago Press].

L. Horwitz, C. Partovian, Z. Lin, J. Herrin, J. Grady, M. Conover, J. Montague, C. Dillaway, K. Bartczak, L. Suter, J. Ross, S. Bernheim, H. Krumholz, and E. Drye. Hospital-Wide All-Cause Unplanned Readmission Measure: Final Technical Report. Technical report, Centers for Medicare and Medicaid Services, July 2012.

C.-T. Hsieh and P. J. Klenow. Misallocation and Manufacturing TFP in China and India*. *The Quarterly Journal of Economics*, 124(4):1403–1448, Nov. 2009. ISSN 0033-5533. DOI: 10.1162/qjec.2009.124.4.1403. URL https://doi.org/10.1162/qjec.2009.124.4.1403.

R. S. Huckman and G. P. Pisano. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, 52(4):473–488, 2006. ISSN 1526-5501. DOI: 10.1287/mnsc.1050.0464. Place: US Publisher: Institute for Operations Research & the Management Sciences (INFORMS).

S. F. Jencks, M. V. Williams, and E. A. Coleman. Rehospitalizations among patients in the Medicare fee-for-service program. *The New England Journal of Medicine*, 360(14): 1418–1428, Apr. 2009. ISSN 1533-4406. DOI: 10.1056/NEJMsa0803563.

J. Kmenta. On Estimation of the CES Production Function. *International Economic Review*, 8(2):180–189, 1967. ISSN 0020-6598. DOI: 10.2307/2525600. URL https://www.jstor.org/stable/2525600. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

J. Lee, J. S. McCullough, and R. J. Town. The impact of health information technology on hospital productivity. *The RAND Journal of Economics*, 44(3):545–568, 2013. ISSN 1756-2171. DOI: 10.1111/1756-2171.12030. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12030. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12030.

J. P. Leigh, D. Tancredi, A. Jerant, and R. L. Kravitz. Physician wages across specialties: informing the physician reimbursement debate. *Archives of Internal Medicine*, 170(19): 1728–1734, Oct. 2010. ISSN 1538-3679. DOI: 10.1001/archinternmed.2010.350.

W. Leontief. Introduction to a Theory of the Internal Structure of Functional Relationships. *Econometrica*, 15(4):361–373, 1947. ISSN 0012-9682. DOI: 10.2307/1905335. URL https://www.jstor.org/stable/1905335. Publisher: [Wiley, Econometric Society].

J. Levinsohn and A. Petrin. Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies*, 70(2):317–341, Apr. 2003. ISSN 0034-6527. DOI: 10.1111/1467-937X.00246. URL https://doi.org/10.1111/1467-937X.00246.

B. A. Mark, D. W. Harless, J. Spetz, K. L. Reiter, and G. H. Pink. California's Minimum Nurse Staffing Legislation: Results from a Natural Experiment. *Health Services Research*, 48(2pt1):435–454, 2013. ISSN 1475-6773. DOI: 10.1111/j.1475-6773.2012.01465.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6773.2012.01465.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-6773.2012.01465.x.

J. Marschak and W. H. Andrews. Random Simultaneous Equations and the Theory of Production. *Econometrica*, 12(3/4):143–205, 1944. ISSN 0012-9682. DOI: 10.2307/1905432. URL https://www.jstor.org/stable/1905432. Publisher: [Wiley, Econometric Society].

P. Mourot. Should Top Surgeons Practice at Top Hospitals?

E. L. Munnich. The Labor Market Effects of California's Minimum Nurse Staffing Law. *Health Economics*, 23(8):935–950, 2014. ISSN 1099-1050. DOI: 10.1002/hec.2966. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.2966. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.2966.

J. Needleman and S. Hassmiller. The role of nurses in improving hospital quality and efficiency: real-world results. *Health Affairs (Project Hope)*, 28(4):w625–633, 2009. ISSN 1544-5208. DOI: 10.1377/hlthaff.28.4.w625.

S. Nickell. Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426, 1981. ISSN 0012-9682. DOI: 10.2307/1911408. URL https://www.jstor.org/stable/1911408. Publisher: [Wiley, Econometric Society].

G. S. Olley and A. Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297, 1996. ISSN 0012-9682. DOI: 10.2307/2171831. URL https://www.jstor.org/stable/2171831. Publisher: [Wiley, Econometric Society].

A. Pakes. Dynamic Structural Models: Problems and Prospects. Mixed Continuous Discrete Controls and Market Interactions. *Cowles Foundation Discussion Papers*, July 1991. URL https://elischolar.library.yale.edu/cowles-discussion-paper-series/1227.

C. Propper and J. Van Reenen. Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance. *Journal of Political Economy*, 118 (2):222–273, Apr. 2010. ISSN 0022-3808. DOI: 10.1086/653137. URL https://www.journals.uchicago.edu/doi/full/10.1086/653137. Publisher: The University of Chicago Press.

C. Propper, S. Burgess, and K. Green. Does competition between hospitals improve the quality of care?: Hospital death rates and the NHS internal market. *Journal of Public Economics*, 88(7):1247–1272, July 2004. ISSN 0047-2727. DOI: 10.1016/S0047-2727(02)00216-5. URL https://www.sciencedirect.com/science/article/pii/S0047272702002165.

C. Raja. How do hospitals respond to input regulation? Evidence from the California nurse staffing mandate. *Journal of Health Economics*, 92:102826, Dec. 2023. ISSN 0167-6296. DOI: 10.1016/j.jhealeco.2023.102826. URL https://www.sciencedirect.com/science/article/pii/S0167629623001030.

D. Restuccia and R. Rogerson. Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10, Jan. 2013. ISSN 1094-2025. DOI: 10.1016/j.red.2012.11.003. URL https://www.sciencedirect.com/science/article/pii/S1094202512000725.

J. A. Romley and D. P. Goldman. How Costly is Hospital Quality? A Revealed-Preference Approach. *The Journal of industrial economics*, 59(4):578–608, 2011. ISSN 0022-1821. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3989349/.

B. L. Rosenberg, J. A. Kellar, A. Labno, D. H. M. Matheson, M. Ringel, P. VonAchen, R. I. Lesser, Y. Li, J. B. Dimick, A. A. Gawande, S. H. Larsson, and H. M. Iii. Quantifying Geographic Variation in Health Care Outcomes in the United States before and after Risk-Adjustment. *PLOS ONE*, 11(12):e0166762, Dec. 2016. ISSN 1932-6203. DOI: 10.1371/journal.pone.0166762. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166762. Publisher: Public Library of Science.

E. Sanderson and F. Windmeijer. A weak instrument F<math><mi is="true">F</mi></math>-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221, Feb. 2016. ISSN 0304-4076. DOI: 10.1016/j.jeconom.2015.06.004. URL https://www.sciencedirect.com/science/article/pii/S0304407615001736.

J. A. Schoenman and J. P. Sutton. Impact of CAH Conversion on Hospital Finances and Mix of Inpatient Services. Technical report, NORC, 2008.

L. S. Seidman. Complements and Substitutes: The Importance of Minding p's and q's. *Southern Economic Journal*, 56(1):183–190, 1989. ISSN 0038-4038. DOI: 10.2307/1059065. URL https://www.jstor.org/stable/1059065. Publisher: Southern Economic Association.

J. Skinner and D. Staiger. Technology Diffusion and Productivity Growth in Health Care. *The review of economics and statistics*, 97(5):951–964, Dec. 2015. ISSN 0034-6535. DOI: 10.1162/REST˙a˙00535. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792131/.

J. Spetz, J. A. Seago, J. Coffman, E. Rosenoff, and E. O'Neil. Minimum Nurse Staffing Ratios in California Acute Care Hospitals | Healthforce Center at UCSF. Technical report, 2000. URL https://healthforce.ucsf.edu/publications/minimum-nurse-staffing-ratios-california-acute-care-hospitals.

J. Spetz, D. W. Harless, C.-N. Herrera, and B. A. Mark. Using minimum nurse staffing regulations to measure the relationship between nursing and hospital quality of care. *Medical care research and review: MCRR*, 70(4):380–399, Aug. 2013. ISSN 1552-6801. DOI: 10.1177/1077558713475715.

W. Sribney, V. Wiggins, and D. Drukker. FAQ: Negative and missing R-squared for 2SLS/IV | Stata. URL https://www.stata.com/support/faqs/statistics/two-stage-least-squares/.

8 Appendix

8.1 Construction of the 30-day, hospital-wide non-readmission rate

I construct the readmission rate from the patient level discharge data by identifying index admissions according to the methodology report for 30-day all-cause readmission published by CMS (Horwitz et al., 2012). Consistent with the CMS exclusion criteria, I exclude patients who died during hospitalization, patients who were transferred to another acute care hospital upon discharge, and patients who were discharged against medical advice. Relative to CMS, I use the full sample of payors rather than restrict to Medicare patients again to reflect the entire patient population. I identify patients that were readmitted to any hospital within 30 days of their discharge date as readmitted and I exclude planned readmissions for diagnoses and clinical procedures that have been designated by CMS as likely to be planned readmissions (Horwitz et al., 2012).

8.2 Construction of the Case Mix Index by the California Department of Health

The CMI is calculated as follows. Patients are assigned to one of hundreds of diagnosis-related groups (DRGs) based on their primary and secondary diagnoses, comorbidities, procedures performed, and age and gender and each of these DRGs is assigned a weight according to CMS that reflects the average resource consumption of the DRG relative to the average resource consumption of all patients (HCAI, 2023). For example, a patient that undergoes knee replacement surgery would be assigned "DRG 469 – Major joint replacement or reattachment of lower extremity with MCC (Major Complication or Comorbidity)". The CMI is calculated as the average of the DRG weights across all patients discharged in the calendar year.[42] A higher CMI therefore reflects a case mix that requires greater resource intensity to improve the patient's health. [43] The link between resource use and case mix is acknowledged by CMS which uses an index known as the Case Mix Index (CMI) to adjust hospitals' Medicare reimbursement rates based on expected resource use at the diagnosis-related group (DRG) level for the patients admitted to the hospital. Hospitals with more severe patients are reimbursed at higher rates.

[42] CMS also calculates a version of the CMI that limits the sample to Medicare patients because their index is used to adjust Medicare reimbursement rates upwards for hospitals with more severe case loads.

[43] According to HCAI: "CMI is the average relative DRQ weight of a hospital's inpatient discharges, calculating by summing the Medicare Severity-Diagnosis Related Group (MS-DRG) weight for each discharge and dividing the total by the number of discharges. The CMI reflects the diversity, clinical complexity, and resource needs of all the patients in the hospital. A higher CMI indicates a more complex and resource-intensive case load."

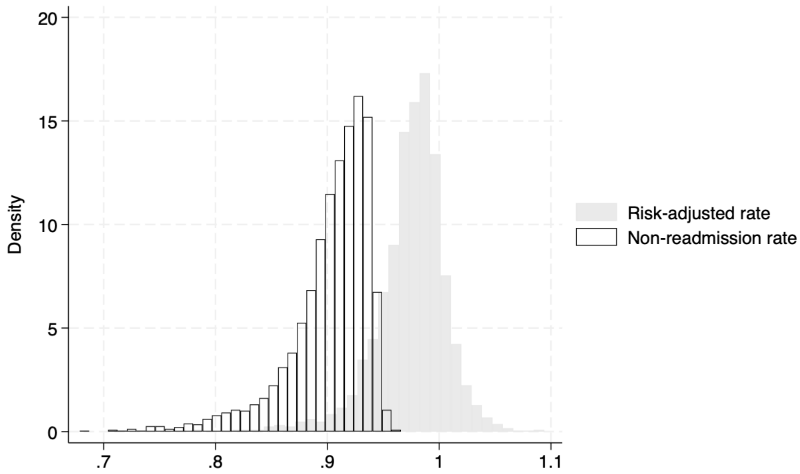


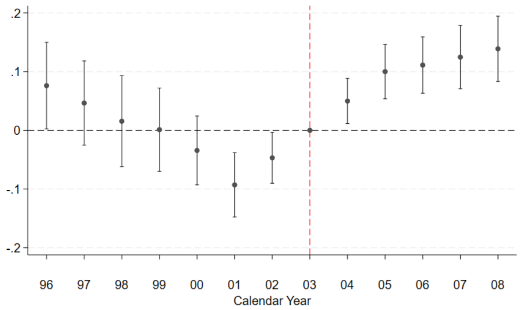
Figure A.1: 30-Day Non-Readmission and Risk-Adjusted Non-Readmission for Sample

Notes: This figure shows the histograms of the 30-day hospital-wide non-readmission rate and the risk-adjusted rate across hospital-years from 1996-2008. I follow CMS in the exclusion criteria for index admissions (Horwitz et al., 2012) and additionally exclude admissions in which the patient had an inpatient stay for any condition within the prior year. The risk-adjusted rate is the residualized rate after controlling for interacted age, gender, and race indicators.

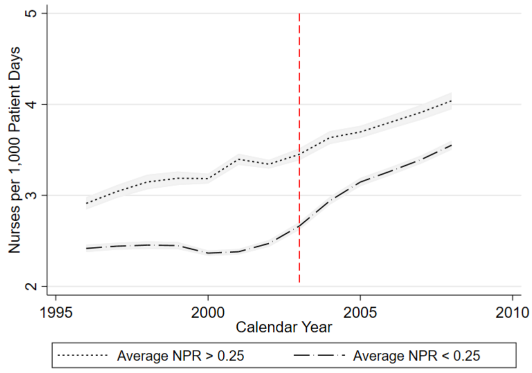
Table A.1: Physicians and Nurses at California Hospitals from 1996-2002

	Bottom 25	25-50	50-75	Top 25
Physicians (total, part-time)	261	309	401	376
Hospital-based, board-certified	0.15	0.15	0.24	0.33
Hospital-based, board-eligible	0.01	0.02	0.04	0.01
Hospital-based, other	0.01	0.02	0.01	0.02
Non-HB, board-certified	0.59	0.59	0.51	0.41
Non-HB, board-eligible	0.07	0.07	0.06	0.03
Non-HB, other	0.10	0.08	0.08	0.05
Residents	0.06	0.06	0.05	0.15
Fellows	0.01	0.01	0.01	0.01
Nurses (total, full-time)	142	156	208	197
Registered Nurses	0.85	0.88	0.89	0.90
Vocational Nurses	0.15	0.12	0.11	0.10
Nurses per 1,000 patient days	2.19	2.46	2.72	3.27
Physicians per 1,000 patient days	5.20	6.12	5.59	5.69
Nurse to physician ratio	0.42	0.40	0.49	0.57

Notes: This table includes the 208 hospitals in my balanced panel sample from 1996-2008. The number of physicians is reported as the number of active medical staff and delineated into hospital-based, non-hospital-based, and residents and fellows. Physicians hours are not limited to clinical hours. On the other hand, the number of nurses is based on the reported number of clinical nursing hours and therefore refers to full-time nursing personnel.



(a) Event-Study Estimates



(b) Raw Means

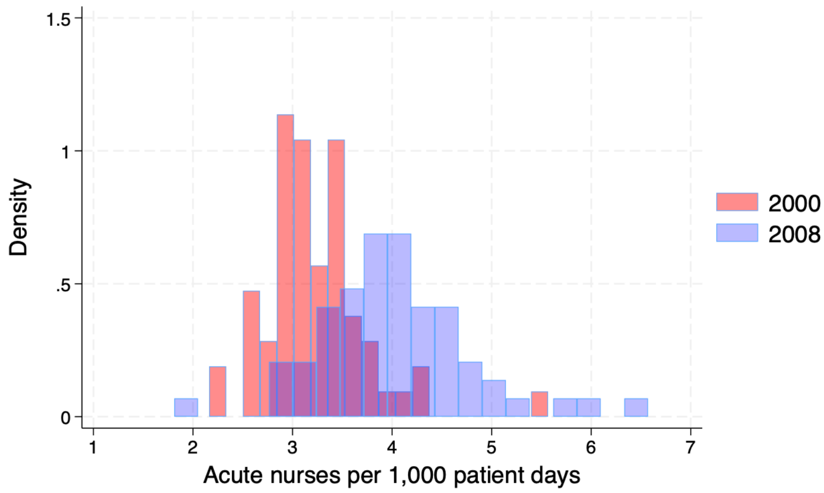
Figure A.2: Effect of Mandate on Nurses per 1,000 Patient Days

Notes: In panel (a), this figure plots coefficients β_t and 95 percent confidence intervals from Equation (1) with the log of nurses per 1,000 patient days as dependent variable. Standard errors are clustered at the hospital level. In Panel (b), this figure plots average values and standard error bands of the nurses per 1,000 patient days by group.

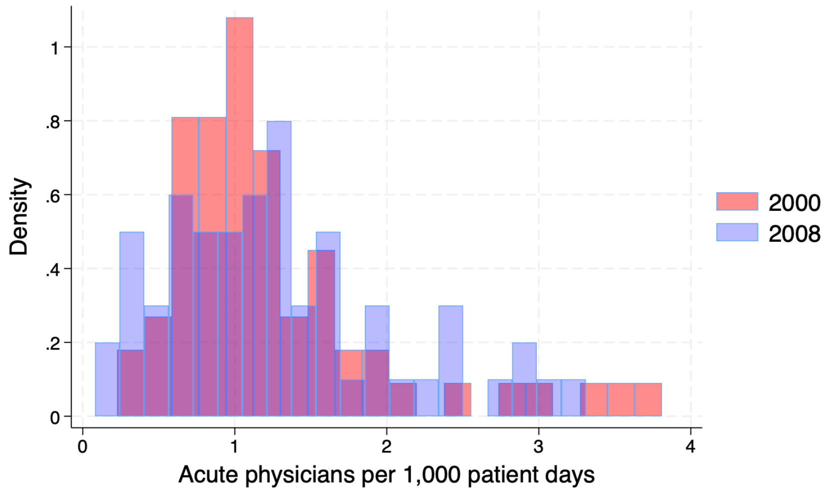
Table A.2: Event-Study Estimates for 30-Day Risk-Adjusted Non-Readmission

	(1) Log NR Rate
Treat x 1996	0.005 (0.006)
Treat x 1997	0.006 (0.006)
Treat x 1998	-0.001 (0.006)
Treat x 1999	-0.001 (0.006)
Treat x 2000	-0.001 (0.005)
Treat x 2001	-0.001 (0.005)
Treat x 2002	0.003 (0.003)
Treat x 2003	0.000 (.)
Treat x 2004	-0.001 (0.003)
Treat x 2005	0.007* (0.004)
Treat x 2006	0.008** (0.004)
Treat x 2007	0.010*** (0.004)
Treat x 2008	0.008* (0.004)
Observations	2,704
R^2	0.531
Mean	0.970
Hospital Fixed Effects	✓
Year Fixed Effects	✓
Weighted by Volume	

Notes: This table presents estimates of the coefficients β_t from Equation (1) with the log of the risk-adjusted non-readmission rate as dependent variable. Standard errors are clustered at the hospital level. The results indicate statistically significant increases in the non-readmission rate for treated hospitals in 2005-2008 after the mandate. The magnitude of the effect is roughly between 0.7 and 1 percent.



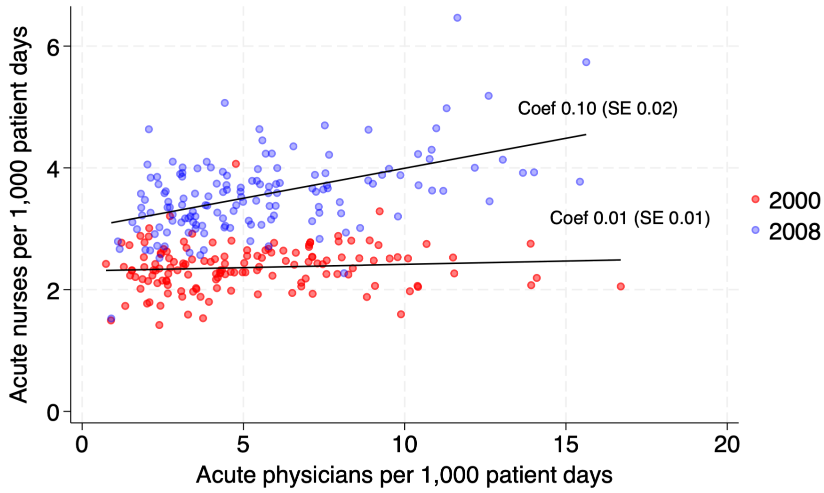
(a) Nurses Per Patient in 2000 vs. 2008



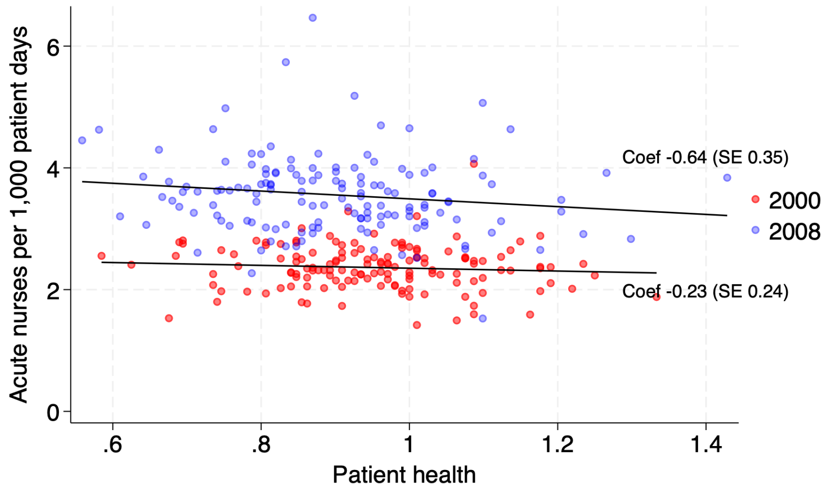
(b) Physicians Per Patient in 2000 vs. 2008

Figure A.3: Input Use in 2000 vs. 2008 for Control Hospitals

Notes: In panel (a), this figure shows the histogram of control hospitals according to the number of nurses per 1,000 patient days prior to the mandate in 2000 (red) and after the mandate in 2008 (blue). In panel (b), I do the same for physicians per 1,000 patient days.



(a) Nurses and Physicians Per Patient in 2000 vs. 2008



(b) Nurses and Patient Health in 2000 vs. 2008

Figure A.4: Change in Input Proportions for Hospitals Treated by the Mandate

Notes: This figure plots the correlations between nurses and physicians per patient across hospitals in 2000 (red) and 2008 (blue) with the estimated coefficients and standard errors indicated for each line. The lack of significant correlation in 2000 and the significant and positive correlation in 2008 indicate that the ratio of nurses per physician became more standardized across hospitals after the mandate.

8.3 Hospital's Problem

I model the production of quality Q_{ht} at a California general acute care hospital h in year t. At the beginning of year t, hospital h observes its productivity ω_{ht} and chooses its inputs to maximize a dynamic problem described by the Bellman equation in Equation (3). The inputs are: patient health (x^h) which is the inverse of the CMI observed in the data, physicians per patient (x^p), and nurses per patient (x^n). The problem is considered dynamic due to the presence of adjustment costs in the labor inputs which are denoted by $c(x^n_{ht}, x^n_{ht-1}, x^p_{ht}, x^p_{ht-1})$. I discuss the dynamic nature of the input choice further in Section 5.1. Quality is measured as the risk-adjusted, 30-day hospital-wide non-readmission rate.

ϵ_{ht} is measurement error.

The hospital's objective is shown as a function of profits, $\pi(.)$, and quality consistent with the literature which models hospital objectives as a function of profits and a non-pecuniary object such as quality or patient volumes (Gaynor and Town, 2011). The functional form is unimportant as recovering the incentives of hospitals is outside the scope of this paper. Profits are allowed to vary heterogeneously in quality improvements due to: heterogeneity in the costs of providing quality (ω_{ht}) and heterogeneity in the price and demand elasticities with respect to quality (γ_{ht}). The relative weight that the hospital places on profits (α) is also allowed to vary across hospitals and over time.

$$V(x^n_{ht-1}, x^p_{ht-1}, \omega_{ht}) = \max_{x^h_{ht}, x^n_{ht}, x^p_{ht}} E[\alpha_h * \{\pi(Q_{ht}, x^h_{ht}, \gamma_h) - c(x^n_{ht}, x^n_{ht-1}, x^p_{ht}, x^p_{ht-1})\} + (1-\alpha_h) * Q_{ht}] + \beta E[V(x^n_{ht}, x^p_{ht}, \omega_{ht}) | x^n_{ht}, x^p_{ht}, \omega_{ht}] \tag{11}$$

$$\text{where: } Q_{ht} = e^{\omega_{ht}+\epsilon_{ht}} F(x^h_{ht}, x^n_{ht}, x^p_{ht})$$

The hospital maximizes the expectation over its payoffs because the production of quality is subject to productivity shocks unanticipated by the hospital (represented by the measurement error term ϵ_{ht}). Patient health x^h_{ht} enters both the production function F and separately in the profit function $\pi(.)$ to denote the fact that higher severity patients incur higher costs to produce the same quality (by way of the production function) but also yield higher reimbursement rates.

8.4 Measurement of capital and materials expenditures

Table 1 shows measures of average capital equipment per patient which include expenditures on computers, testing and diagnostic equipment, beds, and sterilizers and average materials per patient which include expenditures on prosthetics, surgical and anesthetic materials, oxygen and medical gases, pharmaceuticals, food, cleaning supplies, and instruments. I measure capital expenditures per patient day as the reported expenditures on capital equipment which includes major movable equipment, minor equipment, and furniture (HCAI Chapter 3000, 1992). In this category of capital expenditures should fall the capital used in testing and diagnostics of patients. Major movable equipment usually have a minimum life of at least three years and are able to be moved. Examples include cars and trucks, desks, beds, chairs, computers, sterilizers, and oxygen tents (HCAI Chapter 3000, 1992). These are distinct from fixed equipment which are also large but immovable and include engines and boilers, generators, elevators, and large machinery (HCAI Chapter 3000, 1992). Minor equipment usually have a minimum life of less than three years and are relatively smaller and subject to storeroom control. Like physicians, capital expenditures are reported at the hospital level therefore I use revenue share to allocate capital expenditures to the acute care unit.

Expenditures on supplies per patient day include expenditures on prosthetics, surgical and anesthetic materials, oxygen and medical gases, pharmaceuticals, food, cleaning supplies, and instruments and minor medical equipment among others (HCAI Chart of Accounts, 2019).

8.5 Institutional features of labor markets for nurses and physicians

Many examples in the literature impose timing assumptions over the factors of production which eliminate the endogeneity problem for inputs for whom choice over the input is assumed to be made prior to the realization of the productivity shock in the period ("fixed" inputs). I argue that these timing assumptions are unreasonable in my setting and I consider all three inputs to be "flexible" and chosen in the same period in which they are used. Separately, I assume that nurse and physician labor are "dynamic" but patient severity is "nondynamic" implying that there are no fixed costs of changing the Case Mix Index but there can be fixed hiring or firing costs of labor.

The market for healthcare professionals is notably rigid relative to other sectors with the fill rate in healthcare and education services – the ratio of hires to job vacancies – remaining stable around 0.7 during the early 2000s compared to 1.3 over the same period for total nonfarm occupations (noa, 2019). Furthermore, significant rigidities may exist due to unionization – fifty percent of nurses employed in a California hospital during the sample period reported being unionized (Raja, 2023).

The low fill rate in the industry and documented search costs in the market for healthcare professionals suggest non-negligible fixed costs of hiring. Significant unionization among nurses and the complexity of hospital-physician group contracting suggests that there may be additional fixed costs associated with termination. These features of the input markets indicate the presence of adjustment costs. Prior work in this literature has also accounted for the adjustment costs in hiring healthcare workers when modeling production (Lee et al., 2013; Grieco and McDevitt, 2017).

Yet even in this relatively rigid labor market the "time-to-hire" from the initial search to the contract date falls well below the one year mark indicating that nurses and physicians should be considered flexibly chosen at t.[44] Recent survey evidence from organizations participating in physician search suggests that the upper bound on time from search to contract signing was slightly under one year for the longest specialties (urology, neurology) with the average time to fill a physician vacancy being four months (Gradney, 2023). The nursing labor market is less rigid than the physician market with time-to-hire periods for hospital nurses ranging from two-and-a-half months for the labor and delivery unit to just over three months for the medical/surgical unit (noa, 2024).

[44] While there are some exceptions among physicians – for example, teaching hospitals must determined the numbers of residents and fellows at least one year in advance – this is not true for the average physician.

Table A.3: First-Stage – Log Inputs

	(1) Log nurses	(2) Log physicians	(3) Log patient health
L1 Log nurses per patient	0.512***	0.199**	0.009
	(0.035)	(0.079)	(0.011)
L1 Log physicians per patient	0.034*	0.713***	0.007
	(0.020)	(0.044)	(0.006)
L1 Log patient health	-0.213*	0.106	0.781***
	(0.116)	(0.260)	(0.036)
L1 Log nurses x L1 Log patient health	0.061	-0.094	-0.018
	(0.077)	(0.172)	(0.024)
L1 Log nurses x L1 Log physicians	-0.008	-0.041***	-0.003
	(0.007)	(0.015)	(0.002)
L1 Log physicians x L1 Log patient health	0.037	0.058	-0.025**
	(0.039)	(0.086)	(0.012)
Log (0.25 - pre-mandate ratio) x Post-mandate	0.942***	0.883	0.216**
	(0.320)	(0.715)	(0.099)
L1 Log physicians x Mandate interact	0.353*	-0.712	0.007
	(0.194)	(0.434)	(0.060)
L1 Log patient health x Mandate interact	0.589	-1.840	0.250
	(0.786)	(1.757)	(0.244)
Critical Access Hospital x Post-enrollment	0.023	-0.072	0.016**
	(0.025)	(0.056)	(0.008)
Observations	2,496	2,496	2,496
R^2	0.778	0.829	0.959
Within R^2	0.286	0.355	0.555
Mean	2.916	4.613	0.921
F-Statistic	90.758	124.737	282.398
Sanderson-Windmeijer F-Statistic	31.91	14.77	105.85
Hospital Fixed Effects	✓	✓	✓
Year Fixed Effects	✓	✓	✓

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: First-Stage – Log Inputs Squared

	(1) Log nurses sq.	(2) Log physicians sq.	(3) Log health sq.
L1 Log nurses per patient	0.954***	-0.222	0.008
	(0.077)	(0.263)	(0.006)
L1 Log physicians per patient	-0.022	1.756***	-0.000
	(0.043)	(0.147)	(0.003)
L1 Log patient health	-0.305	1.338	0.020
	(0.254)	(0.863)	(0.018)
L1 Log nurses x L1 Log patient health	0.016	-0.523	-0.129***
	(0.168)	(0.569)	(0.012)
L1 Log nurses x L1 Log physicians	0.016	0.013	-0.002*
	(0.015)	(0.051)	(0.001)
L1 Log physicians x L1 Log patient health	0.072	-0.300	-0.035***
	(0.084)	(0.286)	(0.006)
Log (0.25 - pre-mandate ratio) x Post-mandate	0.308	1.729	0.036
	(0.698)	(2.367)	(0.051)
L1 Log physicians x Mandate interact	0.701*	-1.542	-0.040
	(0.424)	(1.438)	(0.031)
L1 Log patient health x Mandate interact	0.940	-3.361	0.578***
	(1.715)	(5.820)	(0.125)
Critical Access Hospital x Post-enrollment	0.115**	0.085	0.004
	(0.054)	(0.184)	(0.004)
Observations	2,496	2,496	2,496
R^2	0.770	0.806	0.916
Within R^2	0.259	0.313	0.245
Mean	3.330	15.035	1.037
F-Statistic	79.383	103.277	73.654
Sanderson-Windmeijer F-Statistic	179.11	13.90	34.54
Hospital Fixed Effects	✓	✓	✓
Year Fixed Effects	✓	✓	✓

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: First-Stage – Log Inputs Interacted

	(1) Nur. x Phys.	(2) Nur. x Patient health	(3) Phys. x Patient health
L1 Log nurses per patient	0.260**	0.023	0.055**
	(0.109)	(0.014)	(0.024)
L1 Log physicians per patient	0.311***	0.016**	0.040***
	(0.061)	(0.008)	(0.013)
L1 Log patient health	0.031	0.285***	0.342***
	(0.356)	(0.047)	(0.078)
L1 Log nurses x L1 Log patient health	-0.086	0.541***	-0.148***
	(0.235)	(0.031)	(0.051)
L1 Log nurses x L1 Log physicians	0.103***	-0.006**	-0.018***
	(0.021)	(0.003)	(0.005)
L1 Log physicians x L1 Log patient health	-0.032	-0.028*	0.592***
	(0.118)	(0.016)	(0.026)
Log (0.25 - pre-mandate ratio) x Post-mandate	-1.050	0.395***	0.110
	(0.978)	(0.130)	(0.214)
L1 Log physicians x Mandate interact	2.132***	-0.009	0.171
	(0.594)	(0.079)	(0.130)
L1 Log patient health x Mandate interact	-1.504	2.322***	0.963*
	(2.405)	(0.320)	(0.526)
Critical Access Hospital x Post-enrollment	-0.076	0.022**	0.027*
	(0.076)	(0.010)	(0.017)
Observations	2,496	2,496	2,496
R^2	0.810	0.949	0.940
Within R^2	0.322	0.580	0.483
Mean	5.252	0.908	0.865
F-Statistic	107.472	313.348	211.926
Sanderson-Windmeijer F-Statistic	19.35	51.06	88.22
Hospital Fixed Effects	✓	✓	✓
Year Fixed Effects	✓	✓	✓

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: First-Stage – Other Inputs

	(1) Log discharges	(2) Log length of stay
L1 Log nurses per patient	0.012	0.012
	(0.054)	(0.019)
L1 Log physicians per patient	-0.021	0.012
	(0.029)	(0.010)
L1 Log patient health	0.100	-0.380***
	(0.169)	(0.060)
L1 Log nurses x L1 Log patient health	-0.327***	0.094**
	(0.112)	(0.040)
L1 Log nurses x L1 Log physicians	-0.007	-0.002
	(0.010)	(0.004)
L1 Log physicians x L1 Log patient health	0.051	0.022
	(0.057)	(0.020)
Log (0.25 - pre-mandate ratio) x Post-mandate	-0.270	0.449***
	(0.491)	(0.174)
L1 Log physicians x Mandate interact	0.512*	-0.245**
	(0.291)	(0.103)
L1 Log patient health x Mandate interact	5.457***	1.357***
	(1.158)	(0.409)
Critical Access Hospital x Post-enrollment	-0.112***	-0.020
	(0.040)	(0.014)
Observations	2,431	2,431
R^2	0.958	0.833
Within R^2	0.023	0.053
Mean	4462.990	3.490
F-Statistic	5.138	12.440
Hospital Fixed Effects	✓	✓
Year Fixed Effects	✓	✓

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Elasticities of Substitution - Nurses and Patient Health

	Percentiles of Distribution				
Nurses per 1,000 Patient Days	10th	25th	50th	75th	90th
1.5	–3.609	0.277	1.024	2.929	7.474
2	–3.279	–1.517	0.038	0.493	1.992
2.5	–1.135	–0.580	–0.302	0.053	0.391
3	–0.391	–0.290	–0.158	–0.030	0.079
3.5	–0.082	–0.082	–0.082	–0.082	–0.082
4	0.002	0.010	0.044	0.075	0.086

Notes: In this table, I present the percentiles of the distribution of elasticities of substitution derived using Equation (7) for each hospital-year in my sample with positive marginal products for nurses and patient health. At low levels of nurses we see that there is significant substitutability between nurses and patient health but this diminishes quickly.

Table A.8: Hospital Characteristics by Staffing Quartile Absent Any Regulation

	Cost-Minimization Model						
Quartile	Health	Prod.	Patient days	Nurse wage '05	Phys. salary '05	Nurses	Phys.
Top 25	0.77	0.94	36,075	35	221,181	2.03	0.70
50-75	0.92	0.93	24,881	34	221,983	1.42	0.51
25-50	1.01	0.93	15,888	31	218,424	1.10	0.40
Bottom 25	1.09	0.96	16,084	34	212,554	0.73	0.26

	Data						
Quartile	Health	Prod.	Patient days	Nurse wage '05	Phys. salary '05	Nurses	Phys.
Top 25	0.90	0.93	25,760	33	215,901	3.41	1.18
50-75	0.93	0.93	27,729	35	221,503	2.90	1.32
25-50	0.94	0.94	26,295	32	224,298	2.55	1.37
Bottom 25	0.99	0.93	22,332	32	214,990	2.07	1.18

Notes: This table shows the average values of health, acute patient days, nurse and physician wages, and nurses and physicians per 1,000 patient days for each nurse staffing quartile. In the top panel, the average values are based on the nurse and physician staffing levels from the cost-minimization model and in the bottom panel, the values are based on the data for the observed incidence of the mandate.

Table A.9: Correlation Between Estimated Productivity and Input Use, Quality, and Revenues Per Day

	(1) Log nurses	(2) Log physicians	(3) Log patient health	(4) Log NR Rate	(5) Log revenues per day	(6) Log patient care costs per day
Estimated productivity, $\widehat{\omega_{ht}}$	0.214**	0.706**	0.105	0.607***	-0.202	0.455***
	(0.103)	(0.310)	(0.087)	(0.010)	(0.221)	(0.146)
Observations	2,496	2,496	2,496	2,496	2,496	2,479
R^2	0.292	0.004	0.014	0.608	0.425	0.461
Within R^2	0.002	0.002	0.001	0.598	0.000	0.004
Hospital Fixed Effects						
Year Fixed Effects	✓	✓	✓	✓	✓	✓

Notes: This table presents the correlations between estimated productivity and input use, quality, and costs and revenues per patient day.

Table A.10: Correlation Between Estimated Productivity and Hospital Type

	(1) Log NR Rate	(2) $\widehat{\omega_{ht}}$	(3) Log NR Rate	(4) $\widehat{\omega_{ht}}$	(5) Log NR Rate	(6) $\widehat{\omega_{ht}}$
Teaching hospital	-0.005***	-0.014***				
	(0.002)	(0.003)				
Small and rural hospital			-0.011***	-0.012***		
			(0.002)	(0.002)		
Not-for-profit owned					-0.004*	-0.012***
					(0.002)	(0.002)
Government owned					-0.009***	-0.015***
					(0.002)	(0.003)
Observations	2,704	2,496	2,704	2,496	2,704	2,496
R^2	0.029	0.031	0.041	0.032	0.033	0.032
Within R^2	0.003	0.012	0.015	0.013	0.007	0.013
Hospital Fixed Effects						
Year Fixed Effects	✓	✓	✓	✓	✓	✓

Notes: This table presents the correlations between estimated productivity and risk-adjusted quality and hospital type. Teaching hospitals and small and rural hospitals have lower quality and productivity relative to their counterparts. Not-for-profit and government-owned hospitals have lower quality and productivity relative to investor-owned hospitals.