

Medicare and Hospital Quality

Eric San Miguel*

October 14, 2026

Frequently updated. [Click here for the most recent version.](#)

Abstract

This paper studies how Medicare payment policy shapes hospital quality and patient welfare. I build a structural model of hospital quality choice and hospital–insurer bargaining in which quality increases demand from Medicare and commercial patients, affects negotiated commercial prices, and raises hospitals’ costs. I estimate the model using Massachusetts Medicare and commercial claims. I find that Medicare and commercial demand respond similarly to quality, but commercial margins are much larger and negotiated commercial prices rise with quality about 17 times more than Medicare’s quality adjustment. At the average hospital, commercial patients account for 81 percent of the marginal return to quality. Therefore, increases in Medicare’s share, through population aging or lowering the age of eligibility, decrease equilibrium quality and average patient welfare. I show that making Medicare rates more responsive to quality can prevent this decline, but payments and quality gains concentrate at the highest-quality, highest-margin hospitals.

Keywords: hospital quality, Medicare, population aging, hospital–insurer bargaining, payer mix, health care markets

JEL Codes: C78, I11, I13, I18, L13

*Department of Economics, Pennsylvania State University. Email: eric.sanmiguel@psu.edu. I thank Paul Grieco, Conor Ryan, Joris Pinkse, Bradley Setzler, Eric Yde, Minu Philip, Vijay Krishna, Marit Hinnosaar, Amrutha Manjunath, Ljubica Ristovska, and Kuan Wang for valuable comments and suggestions, as well as participants at brownbags and seminars at Penn State, the University of Nottingham, the ITAM Alumni Conference, and the 33rd European Workshop on Econometrics and Health Economics, and my discussant, Karine Lamiraud. Mohan Ramanujan, Carla Tokman, and Sara Schwarz at the NBER and Leah MacArthur at the Center for Health Information and Analysis (CHIA) provided invaluable help in accessing the data. I acknowledge financial support from the Ed Green Graduate Fund at Penn State. This paper uses data from the Massachusetts All-Payer Claims Database, provided by CHIA, and Medicare claims provided by the Centers for Medicare and Medicaid Services (CMS). The analysis and conclusions are my own and do not represent the views of CHIA or CMS.

1 Introduction

In markets where firms post prices, they have incentives to provide quality because it increases consumers’ willingness to pay and, therefore, allows them to charge higher prices. However, in the US, hospitals typically do not post prices and are instead reimbursed by insurers. Commercial insurers bilaterally bargain with hospitals over commercial rates, and higher-quality hospitals may be able to negotiate higher rates. In contrast, Medicare,¹ the largest payer of hospital stays (Liang, Moore, and Soni, 2020), reimburses hospitals at administered rates with small quality adjustments.² The quality a hospital provides therefore depends on the degree to which both commercial insurers and Medicare reward it. This paper studies how Medicare payment policy shapes hospital quality and patient welfare.

Medicare’s role in shaping hospital quality is of growing consequence, since population aging will increase the Medicare share of hospital demand. The US Census Bureau (2023) projects that the population aged 65 and older will rise from 58 million in 2022 to over 78 million by 2040. Moreover, there is substantial public support for expanding Medicare eligibility, for instance, by lowering the eligibility age or by extending Medicare to all US residents (“Medicare for All”). A 2026 poll of more than 19,000 voters found that at least two-thirds of respondents from both parties favored lowering the eligibility age to 55 (DeGroot, 2026).

To study how Medicare shapes hospital quality and patient welfare, I build and estimate a structural model of the hospital industry in which hospitals treat both Medicare and commercial patients, choose quality, and bargain over commercial prices with insurers. Interactions in this model unfold in three stages. First, all hospitals simultaneously choose a single quality level for all of their patients. Second, hospitals and insurers bilaterally bargain over commercial rates, taking the quality chosen in the first stage as given. Third, patients choose a hospital. When choosing quality, hospitals internalize how quality affects their demand from both Medicare and commercial patients, the commercial prices they negotiate with insurers and the Medicare rate they receive, and their cost of treating patients. While previous work has modeled commercial hospital prices as the outcome of bilateral bargaining between hospitals and insurers, taking hospital quality as given (Capps, Dranove, and Satterthwaite, 2003; Gowrisankaran, Nevo, and Town, 2015; Ho and Lee, 2017), I extend this framework by allowing hospitals to choose quality before bargaining with insurers over commercial rates, and show how commercial prices change with the quality a hospital chooses.

I find that, at the average hospital, commercially insured patients account for 81 percent of the marginal return to quality, despite making up only 34 percent of severity-weighted demand. As a result, increases in the Medicare share, either through population aging or lowering the eligibility

¹Medicare is the federal health insurance program for people 65 and older. It also covers certain younger people with long-term disabilities or end-stage renal disease.

²Medicare operates three quality-linked payment programs—Hospital Value-Based Purchasing, the Hospital Readmissions Reduction Program, and the Hospital-Acquired Condition Reduction Program; their combined variation across hospitals is less than one percent of base payments, far smaller than geographic wage adjustments or commercially negotiated price differences (Section 2).

age, lead to lower equilibrium quality and patient welfare. Finally, I show that Medicare could prevent average welfare from falling as the population ages by increasing its rates, at a fiscal cost, or by making them more responsive to quality, which requires no additional spending but concentrates payments and quality gains at the highest-quality, highest-margin hospitals.

I estimate the model using patient-level claims data covering both Medicare and commercially insured patients in Massachusetts over 2018–2022, which record, for each hospitalization, the hospital a patient chose, her residential ZIP code, and demographics such as age and gender. I measure quality with a composite index of spending per patient, patient experience, and clinical outcomes. Estimation proceeds in two steps. First, I use a multinomial logit model of hospital choice to measure how Medicare and commercial demand respond to quality. Second, using commercial prices observed in the claims and the model’s optimality conditions for bargaining and hospitals’ quality choice, I estimate how higher quality raises hospitals’ costs and how much insurers value enrollee welfare.

My demand estimates show that Medicare and commercial patients respond similarly to quality: a 0.1 standard-deviation increase in quality raises a hospital’s Medicare demand by about 8.2 percent and its commercial demand by about 7.4 percent. However, my supply-side estimates imply that commercial margins are much higher—the average margin for commercially insured patients is \$4,370 per patient, compared with \$90 for Medicare patients—and that commercial prices increase roughly 17 times more with quality than the Medicare rate does through its quality adjustments. These estimates imply that the return to quality comes predominantly from commercially insured patients.

I use the estimated model to study how a decrease in the share of commercial patients, through population aging or a lower Medicare eligibility age, would affect hospital quality and patient welfare. First, I project how population aging will change the payer mix in Massachusetts through 2040. Under this projection, the Medicare share rises from 65.3 percent in 2019 to 74.3 percent in 2040. In the 2040 equilibrium, hospitals lower their quality by 0.17 standard deviations on average. As a result, average consumer surplus per patient falls by \$118 for Medicare patients, \$200 for commercial patients, and \$152 across all patients. Commercial patients lose more welfare because quality declines are concentrated at hospitals with the highest commercial shares.

Second, I progressively lower the Medicare eligibility age from 65 to 50. In each counterfactual, all commercially insured patients at or above the new eligibility age become Medicare patients, and hospitals are now paid the Medicare rate for them. Medicare rates are lower and respond less to quality than negotiated commercial prices, which reduces the marginal return to quality and lowers equilibrium quality. With eligibility at 60, the decline in quality lowers average consumer surplus by \$218 per patient. With eligibility at 50, quality falls even further, lowering average consumer surplus by \$488 per patient.

I then use the model to study how Medicare payment policy can offset the welfare loss from aging. One option is to increase baseline Medicare rates. I estimate that preventing average per-patient welfare from falling by 2040 would require a 2.4 percent rate increase, at an additional

cost of approximately \$5.1 billion per year nationally. An alternative is to make Medicare rates more responsive to quality. I model this through a stylized version of Medicare’s Hospital Value-Based Purchasing program (HVBP), which withholds 2 percent of each hospital’s base payments and redistributes the pool based on quality scores. Preventing the same welfare loss would require withholding and redistributing 8.7 percent of each hospital’s base payments. Making Medicare rates more responsive to quality requires no additional spending, but it moves Medicare payments from low-margin hospitals toward the hospitals with the highest margins. Massachusetts General and Brigham and Women’s, the two flagship teaching hospitals of Harvard Medical School, together receive 38 percent of the redistributed pool. A larger withhold changes hospitals’ return to quality in two ways. First, it makes the Medicare rate a steeper function of quality. Second, a hospital with quality above the mean now receives a higher Medicare rate, which increases its return to attracting an additional Medicare patient, while a hospital below the mean receives a lower rate, which decreases it. As a result, quality rises most at the hospitals where it was already highest, and barely moves or falls where it was already lowest.

Together, these findings establish that the way Medicare reimburses hospitals—at an administered rate that responds weakly to quality—shapes the quality of care all patients receive, and that increases in Medicare’s share, through population aging or a lower eligibility age, lower quality by weakening hospitals’ incentive to provide it. The structural model provides a framework for evaluating Medicare reimbursement policies by their impact on hospital quality and patient welfare, and highlights that proposals to expand Medicare eligibility should account for these effects.

Related Literature: This paper is related to several strands of literature. First, this paper builds on the literature on hospital–insurer bargaining. [Capps et al. \(2003\)](#) develop the option-demand framework for measuring hospital market power, and [Gowrisankaran et al. \(2015\)](#) estimate a Nash-in-Nash bargaining model to evaluate hospital mergers, with [Collard-Wexler, Gowrisankaran, and Lee \(2019\)](#) providing the theoretical microfoundation for this framework. [Ho and Lee \(2017\)](#) and [Ho and Lee \(2019\)](#) extend the bargaining framework to incorporate insurer competition and endogenous network formation. [Setzler \(2025\)](#) develops a unified merger framework that integrates endogenous quality with product and labor market power, applied to U.S. hospital consolidation. [Balan and Sandford \(2026\)](#) derive theoretical conditions under which mergers reduce quality in a bilateral bargaining setting with a single buyer. In these papers, quality is either taken as given, studied outside a bargaining framework, or analyzed theoretically in a simpler bilateral setting. This paper extends this literature by making quality an endogenous choice that hospitals make before bargaining, and showing how the price they negotiate changes with it.

Second, this paper is closely related to work on how a provider’s payer mix shapes the quality it provides. [Garthwaite, Ody, and Starc \(2022\)](#) show that hospitals with a larger predicted share of commercially insured patients have higher quality across a range of measures. [Hackmann \(2019\)](#) estimates a model of the nursing home industry in which the administered Medicaid rate shapes the quality provided to all residents. Relative to these papers, my structural model allows me to estimate how the return to quality from commercial patients compares to the return from Medicare

patients, including through the response of bargained prices to quality. It also allows me to evaluate how changes in Medicare’s share, through population aging or a lower eligibility age, would change hospital quality and patient welfare.

Third, this paper contributes to work studying how financial incentives affect hospital quality. [Cutler \(1995\)](#) finds that the shift from cost-based to prospective payment led to changes in mortality timing and readmissions. More recently, [Gupta \(2021\)](#) shows that financial penalties under Medicare’s Hospital Readmissions Reduction Program reduced readmissions and heart attack mortality, and [Gross, Sacarny, Shi, and Silver \(2024\)](#) document that hospitals receiving larger Medicare payment increases from the MS-DRG reform increased patient volume and hired more nurses. [Cooper, Doyle, Graves, and Gruber \(2022\)](#) use quasi-random ambulance assignment to show that higher commercially negotiated prices are associated with lower mortality in competitive hospital markets, while the price–quality link breaks down in concentrated markets. These papers study the effects of direct payment changes or payment reforms on hospital behavior. This paper instead studies how Medicare’s payment policy shapes the quality of care all patients receive, and provides a structural framework to evaluate counterfactual payment policies.

The paper proceeds as follows. Section 2 describes the institutional setting. Section 3 provides motivating empirical evidence. Section 4 presents the model. Section 5 estimates demand. Section 6 estimates costs and bargaining parameters. Section 7 presents counterfactual exercises. Section 8 concludes.

2 Institutional Background

Medicare is a federal health insurance program that covers virtually all adults aged 65 and older in the United States.³ Over 99 percent of adults aged 65 and older are enrolled ([Centers for Medicare & Medicaid Services, 2025](#); [U.S. Census Bureau, 2023](#)). The program accounts for approximately 48 percent of all inpatient days at general acute-care hospitals, and Medicare spent approximately \$111 billion on inpatient hospital services under its prospective payment system in fiscal year 2022 ([Medicare Payment Advisory Commission, 2024a](#)). Medicare enrollees can receive coverage through either traditional Medicare (fee-for-service), in which the government directly reimburses hospitals for every stay, or Medicare Advantage (MA), in which the government provides capitated payments to private insurers, which enroll beneficiaries and pay hospitals for their care. However, hospitals receive nearly the same rates for traditional Medicare and MA patients: [Maeda and Nelson \(2018\)](#) estimate a median MA-to-FFS price ratio of 1.01 across metropolitan areas, and [Berenson, Sunshine, Helms, and Lawton \(2015\)](#) document that MA plans nominally pay 100–105 percent of traditional Medicare rates.⁴

³Medicare also covers individuals under 65 who qualify through Social Security Disability Insurance, end-stage renal disease, or amyotrophic lateral sclerosis. These groups account for approximately 12 percent of all Medicare beneficiaries nationally ([Centers for Medicare & Medicaid Services, 2025](#)). The paper’s identifying variation comes from population aging among those reaching the standard eligibility age of 65.

⁴Federal law requires hospitals to accept fee-for-service rates for out-of-network MA patients (Section 1866 of the Social Security Act; 42 CFR 422.214), which limits the scope for MA plans to negotiate substantially different

Medicare reimburses hospitals for inpatient stays through the Inpatient Prospective Payment System (IPPS). Under the IPPS, each hospital stay is classified into a Diagnosis-Related Group (DRG) based on the patient’s diagnoses and procedures, and the hospital receives a predetermined payment for that DRG regardless of the actual cost of treatment. Each DRG has a numerical weight that reflects the expected resource use of treating a patient in that DRG. DRG weights are normalized so that a hospitalization of average intensity has a weight of 1. The Centers for Medicare and Medicaid Services (CMS) sets a base payment per DRG weight following a national formula, adjusted for local variation in wages. For instance, a major joint replacement has a DRG weight of approximately 2, so a hospital receives roughly twice the base payment for it.

Medicare also operates three quality-linked payment adjustment programs: the Hospital Value-Based Purchasing (HVBP) program, which withholds and redistributes 2 percent of base DRG payments based on clinical and patient experience measures; the Hospital Readmissions Reduction Program (HRRP), which penalizes hospitals with excess 30-day readmissions by up to 3 percent of base payments; and the Hospital-Acquired Condition Reduction Program (HACRP), which imposes a 1 percent payment reduction on hospitals in the worst-performing quartile. These adjustments are small relative to other sources of variation in hospital reimbursement. In 2019, the base operating payment rate was approximately \$5,800 per DRG-weight. The combined standard deviation of the three quality adjustments across IPPS hospitals was just \$55 per DRG-weight—less than one percent of the base payment—compared with a standard deviation of \$832 for the geographic wage-index adjustment and \$6,655 for commercially negotiated prices. Nearly all cross-hospital variation in Medicare reimbursement reflects local wage differences, not quality incentives.

Commercially insured patients—primarily working-age adults and their dependents covered through employer-sponsored health insurance—are the other major source of hospital inpatient revenue. Commercial insurers negotiate reimbursement rates bilaterally with individual hospitals, and these negotiated rates vary substantially across hospitals and markets. Commercial insurers maintain provider networks: a hospital that reaches a price agreement with an insurer is included in the insurer’s network, and the insurer’s enrollees can seek in-network care at that hospital. Commercially negotiated rates average roughly 2.5 times Medicare reimbursement nationally ([White and Whaley, 2024](#)), with substantial variation across hospitals (Appendix Figure 14). Unlike Medicare rates, which are determined by a national formula and vary primarily with local wages, commercial prices are the outcome of bilateral negotiations and differ substantially even among hospitals in the same market. There is empirical evidence that these price differences reflect, at least in part, differences in the care that hospitals provide: [Cooper et al. \(2022\)](#) use quasi-random ambulance assignment to show that hospitals negotiating higher commercial prices have lower patient mortality in competitive markets. Medicare reimbursement, beyond the limited quality adjustments described above, does not vary with these hospital-level differences in care.

The gap between commercial and Medicare reimbursement has direct consequences for hospi-

rates. The Medicare share of hospital demand used throughout this paper includes both traditional Medicare and MA patients.

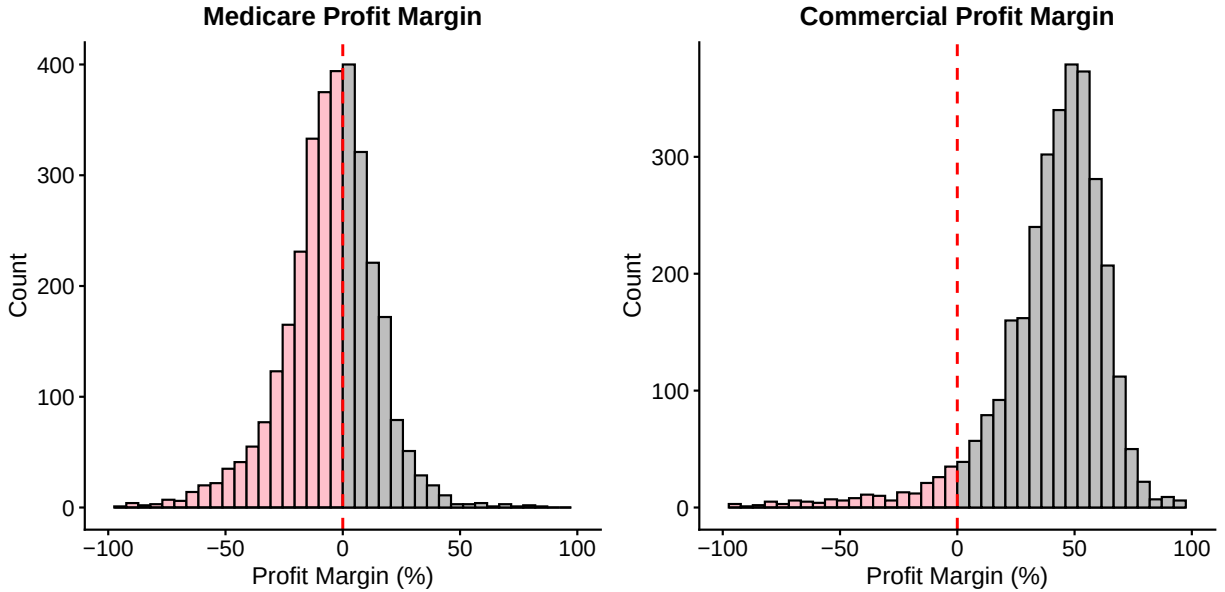


Figure 1: Distribution of hospital profit margins by payer type. Data are from the Healthcare Cost Report Information System (HCRIS) for general acute-care hospitals. The dashed red line marks zero. Medicare margins are centered near zero with a majority of hospitals reporting negative margins, while commercial margins are concentrated well above zero.

tal finances. Figure 1 plots the distribution of hospital accounting margins by payer type from the Healthcare Cost Report Information System (HCRIS), the annual cost reports that Medicare-certified hospitals file with CMS. The Medicare margin distribution is centered near zero, and approximately 59 percent of U.S. hospitals report negative Medicare accounting margins. In contrast, the distribution of commercial margins is concentrated well above zero, and most hospitals report positive margins on commercially insured patients.

The population of the United States is aging rapidly, driven by the aging of the baby boom generation. Census projections indicate that the number of adults aged 65 and older will grow from approximately 58 million in 2022 to over 78 million by 2040, raising the elderly share of the population from 17 to over 22 percent (U.S. Census Bureau, 2023). Because virtually all adults become eligible for Medicare at age 65, this demographic shift is increasing the share of the population enrolled in Medicare, which rose from roughly 16 percent in 2011 to over 18 percent by 2019. The trend is broad-based: between 2010 and 2020, over 98.6 percent of U.S. counties experienced an increase in their elderly population share, with a median increase of 4 percentage points and increases exceeding 6.6 percentage points at the 90th percentile. As this demographic shift continues, hospitals will face a patient mix increasingly composed of Medicare patients relative to commercially insured patients.

3 Motivating Empirical Evidence

3.1 Data

I construct a hospital-year panel spanning 2010–2019 by linking financial and operational data from the CMS Hospital Cost Reports Information System (HCRIS), patient satisfaction scores from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, and clinical quality measures from the CMS Hospital Compare program. The sample includes all Medicare-certified general acute-care hospitals—those that participate in the Medicare program and file annual cost reports with CMS.⁵ The final panel contains approximately 4,500 hospitals per year and up to 45,000 hospital-year observations, with coverage varying across outcome measures. I end the sample in 2019 to avoid the confounding effects of the COVID-19 pandemic on hospital outcomes and operations.

Payer composition and financial outcomes. I measure exposure to Medicare at the hospital level using the Medicare share of inpatient days, constructed from the HCRIS (RAND Corporation, 2025). The share includes both traditional Medicare (fee-for-service) and Medicare Advantage patients.⁶ From the same source, I measure average non-Medicare inpatient revenue per discharge⁷—a widely used proxy for commercially negotiated prices (Dafny, 2009; Dafny, Ho, and Lee, 2019; Setzler, 2025)—and the Medicare operating profit margin.

Expenditures and staffing. I measure Medicare inpatient costs per discharge, which captures spending on Medicare patients only, and non-Medicare inpatient costs per discharge, which captures spending on all other patients.⁸ I also measure medical full-time equivalents (FTEs) per discharge and total FTEs per discharge from the HCRIS. I classify hospital workers into medical and non-medical categories following the National Academy for State Health Policy Hospital Cost Tool (National Academy for State Health Policy, 2024): medical workers include nurses, nursing aides, and the hospital’s directly employed physicians, while non-medical workers include workers in categories such as administrative, maintenance, housekeeping, and support staff. I also measure the share of registered nurses among all nurses from the American Hospital Association Annual Survey, a proxy for the skill composition of the nursing workforce.

Patient satisfaction. I measure patient experience using HCAHPS, a government-mandated survey administered to a random sample of recently discharged patients at every Medicare-certified hospital. CMS publicly reports hospital-level scores for the full period of my sample. I use the top-box score—the share of patients giving the most favorable response—for six measures: overall hospital rating, willingness to recommend, responsiveness of hospital staff, cleanliness, communica-

⁵Virtually all US hospitals are Medicare-certified.

⁶As discussed in Section 2, Medicare Advantage plans pay hospitals rates close to traditional Medicare fee-for-service rates, so Medicare share includes both traditional and MA patients throughout.

⁷Following Dafny (2009), non-Medicare revenue per discharge is total inpatient charges net of contractual discounts and Medicare payments, divided by non-Medicare discharges. All components are reported in the HCRIS.

⁸I construct the non-Medicare costs per discharge as a residual using Medicare cost reports, subtracting Medicare inpatient costs from total allowable inpatient costs and dividing by non-Medicare discharges. I trim this measure at the 1st and 99th percentiles.

tion about medicines, and quietness of rooms at night. I also construct an HCAHPS composite as the mean of the six top-box scores, each standardized to zero mean and unit variance within year.

Clinical quality. I measure clinical outcomes from three CMS Hospital Compare reporting programs. I construct a mortality and readmission composite from six risk-adjusted 30-day rates: mortality and readmission for acute myocardial infarction, heart failure, and pneumonia. I standardize each component to zero mean and unit variance and take their mean, with higher values indicating worse outcomes. I similarly construct an emergency department waiting time composite from three ED waiting time measures: median time from arrival to departure for admitted patients, admit decision time to departure for admitted patients, and median time from arrival to departure for discharged patients, with higher values indicating worse outcomes (longer wait times). I construct a standard-of-care composite from four indicators of routine clinical care: central line-associated bloodstream infections, catheter-associated urinary tract infections, hospital-acquired pressure ulcers, and in-hospital falls resulting in hip fractures. I standardize each to zero mean and unit variance and take their mean, with higher values indicating worse outcomes.⁹

Quality indexes. I construct two aggregate quality indexes. Both combine four components: all-payer inpatient costs per discharge,¹⁰ the HCAHPS overall hospital rating, the share of patients who would recommend the hospital, and the mortality-readmission composite (reversed in sign so that higher values indicate better quality). The components I use to construct the indexes combine three different ways of measuring hospital quality—patient experience, resource use, and clinical outcomes—that have been validated in the literature. Using an ambulance-referral-pattern instrument for hospital assignment, [Doyle, Graves, Gruber, and Kleiner \(2015\)](#) find that a one-standard-deviation increase in hospital spending causally reduces patient mortality by approximately 10 percent, though the authors caution that high-spending hospitals could in principle achieve the same outcomes at lower cost if they became more efficient. [Doyle, Graves, and Gruber \(2019\)](#) use the same identification strategy to validate the remaining two component types directly: a two-standard-deviation improvement in patient experience causally reduces 30-day readmission and one-year mortality, and a hospital’s mortality and readmission scores causally predict the mortality and readmission its patients actually experience. Each component is standardized to zero mean and unit variance across the panel. The primary index is the mean of the four standardized components, re-standardized to a z-score. The second index is the first principal component of the same four standardized measures, as a robustness check on this equal weighting.

Demographics and local economic conditions. I link the hospital panel to county-level data on population, economic conditions, and demographic structure. Total population by age comes from the Surveillance, Epidemiology, and End Results (SEER) program’s county-level single-year-of-age population file, spanning 1969–2022. County income per capita comes from the Bureau

⁹Not all quality measures are available for every hospital-year. When a component of a composite is missing, I predict it from the remaining observed components using a linear regression fitted on hospital-years where all components are reported.

¹⁰Measured as total allowable costs per discharge equivalent, which adjusts the inpatient discharge count for outpatient volume.

of Economic Analysis regional accounts and county unemployment rates from the Bureau of Labor Statistics Local Area Unemployment Statistics. Section 3.2 describes how I use the county age distributions to construct a predetermined aging instrument for hospital Medicare share, following Maestas, Mullen, and Powell (2023).

3.2 Empirical Strategy

One potential concern when estimating how payer mix affects hospital quality is that the share of Medicare patients in a geographic region may be related to local economic conditions. For instance, areas experiencing economic decline may see their working-age population out-migrate, which would raise the share of people over the age of 65. At the same time, local economic conditions may independently affect hospital quality through channels such as funding, the availability of skilled workers, and the composition of patient demand. More broadly, areas with large elderly populations differ systematically from those with younger populations: they are disproportionately rural, have lower population density, and have fewer healthcare providers per capita (Glasgow and Brown, 2012).

I address this concern using an instrumental variables strategy based on the predetermined component of population aging, following Maestas et al. (2023). The instrument predicts what each county's share of population over 65 would be in year t in the absence of migration, based on lagged demographics and national survival rates. Specifically, I construct the predicted elderly share

$$Z_{ct} \equiv \frac{\hat{A}_{ct}}{\hat{N}_{ct}}, \quad (1)$$

where

$$\hat{A}_{ct} = \sum_{j \geq 65-(t-1990)} N_{jc,1990} \times \frac{N_{j+(t-1990),t}}{N_{j,1990}} \quad (2)$$

is the predicted number of people aged 65 and older in county c in year t , $N_{jc,1990}$ is the number of people of age j in county c in 1990, $N_{j+(t-1990),t}/N_{j,1990}$ is the national survival rate from age j to age $j + (t - 1990)$, and

$$\hat{N}_{ct} = \sum_{j \geq 20-(t-1990)} N_{jc,1990} \times \frac{N_{j+(t-1990),t}}{N_{j,1990}} \quad (3)$$

is the predicted total adult population in county c in year t . The identifying assumption is that because Z_{ct} is constructed entirely from the 1990 county age distribution and national survival rates, it is orthogonal to the county-specific economic shocks that drive selective migration and that may independently affect hospital quality.

I also control for three time-varying county-level measures of local economic conditions: log labor force size, the unemployment rate, and log per-capita personal income. I include hospital and year fixed effects, which absorb all time-invariant hospital characteristics and common time trends. Identification therefore relies on within-hospital variation over time in Medicare share driven by

the predetermined aging of the hospital’s local population. I estimate

$$\log(Y_{jt}) = \alpha_j + \eta_t + \gamma \text{Med. percent}_{jt} + X'_{jt}\beta + \varepsilon_{jt}, \quad (4)$$

where Y_{jt} is the outcome of interest for hospital j in year t , Med. percent_{jt} is the Medicare share of inpatient days, α_j and η_t are hospital and year fixed effects, and X_{jt} contains the county-level economic controls. I instrument Med. percent_{jt} with the predetermined component of population aging Z_{ct} . I cluster standard errors at the hospital and county levels in all regressions.

The first-stage relationship is strong: the first-stage F -statistic is 49.4, and a one-percentage-point increase in the predicted elderly share is associated with a 0.75 percentage-point increase in hospital Medicare share ($p < 0.001$). Table 3 in Appendix A.2 reports the full first-stage regression.

3.3 Results

I report IV estimates of the effect of Medicare share on hospital outcomes, grouping results by outcome type. All specifications include hospital and year fixed effects, the county-level economic controls described in Section 3.2, and instrument Medicare share with the predetermined aging instrument. Appendix A.3 reports the full regression tables underlying each figure.

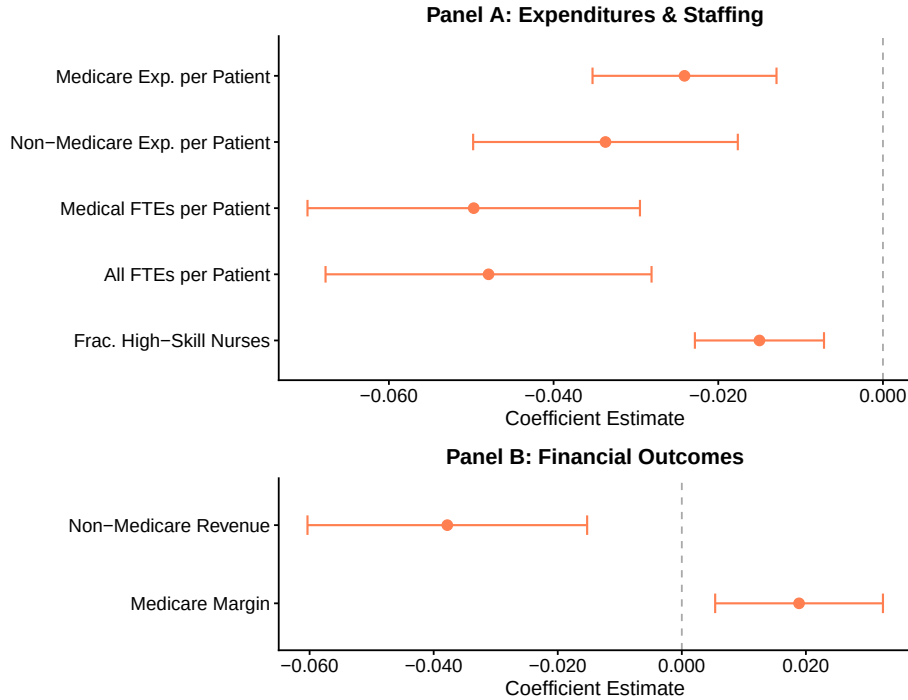


Figure 2: Effect of Medicare share on expenditures, staffing, and financial outcomes. Each point is an IV estimate; bars show 95% confidence intervals. See Section 3.3 for specification details.

Expenditures, staffing, and financial outcomes. Figure 2 presents IV estimates for hospital spending, staffing, and financial outcomes. A one-percentage-point increase in Medicare share

caused by population aging reduces inpatient spending per Medicare patient by approximately 2.4% (-0.024 , $SE = 0.006$) and per non-Medicare patient by approximately 3.4% (-0.034 , $SE = 0.008$). Measures of staffing per patient also decline: the number of medical workers per patient falls by approximately 5% (-0.050 , $SE = 0.010$) and the total number of workers per patient by approximately 4.8% (-0.048 , $SE = 0.010$). The fraction of high-skill nurses also declines, indicating that hospitals shift toward a less expensive staffing mix as Medicare share rises. These reductions suggest a deterioration in the quality of care, as both nurse staffing levels and nurse skill composition are significant determinants of patient outcomes (Needleman, Buerhaus, Mattke, Stewart, and Zelevinsky, 2002; Aiken, Clarke, Sloane, Sochalski, and Silber, 2002; Gandhi, Grabowski, McGarry, and Ristovska, 2025). Non-Medicare revenue per discharge—a proxy for commercially negotiated prices—falls by approximately 3.8% (-0.038 , $SE = 0.011$). In the model developed in Section 4, equilibrium commercial prices rise with hospital quality; the observed decline in commercial prices is therefore consistent with the quality reductions documented above. The Medicare operating profit margin, by contrast, increases by 1.9 percentage points (0.019 , $SE = 0.007$). Because Medicare reimbursement rates are administratively fixed and respond minimally to quality, the cost savings from lower quality translate into higher Medicare margins. This is consistent with the cross-sectional pattern documented by Garthwaite et al. (2022), in which hospitals located in areas with more commercially insured residents earn lower Medicare margins.

Patient satisfaction. Population aging also affects patient experience. Figure 3 presents IV estimates for the six HCAHPS measures of patient-reported care experience. All six point estimates are negative and statistically significant. A one-percentage-point increase in Medicare share reduces willingness to recommend the hospital by approximately 0.8% (-0.0084 , $SE = 0.0020$), responsiveness of staff by approximately 0.6% (-0.0061 , $SE = 0.0018$), communication about medicines by approximately 0.6% (-0.0059 , $SE = 0.0017$), quietness of the hospital at night by approximately 0.6% (-0.0058 , $SE = 0.0023$), overall hospital rating by approximately 0.5% (-0.0054 , $SE = 0.0019$), and cleanliness of rooms and bathrooms by approximately 0.3% (-0.0032 , $SE = 0.0016$). These measures capture dimensions of hospital care that patients themselves value, and hospital-level HCAHPS scores have been shown to predict patient outcomes (Doyle et al., 2019). The uniformity of these results—every dimension of the patient experience declines—indicates that the quality deterioration associated with higher Medicare share extends broadly across aspects of care valued by patients.

Quality composites and aggregate indexes. Figure 4 presents IV estimates for quality composites and aggregate quality indexes. The mortality and readmission composite shows a small and statistically insignificant point estimate (-0.019 , $SE = 0.018$), though the confidence interval is wide enough that we cannot rule out effects of meaningful magnitude. This imprecision likely reflects the substantial year-to-year noise in hospital-level mortality and readmission rates, which limits the precision of within-hospital IV estimates. Moreover, 30-day mortality and readmission may require a larger deterioration in care to move than the process and experience measures documented above. By contrast, the ED waiting time composite increases by 0.040 standard deviations ($SE = 0.015$),

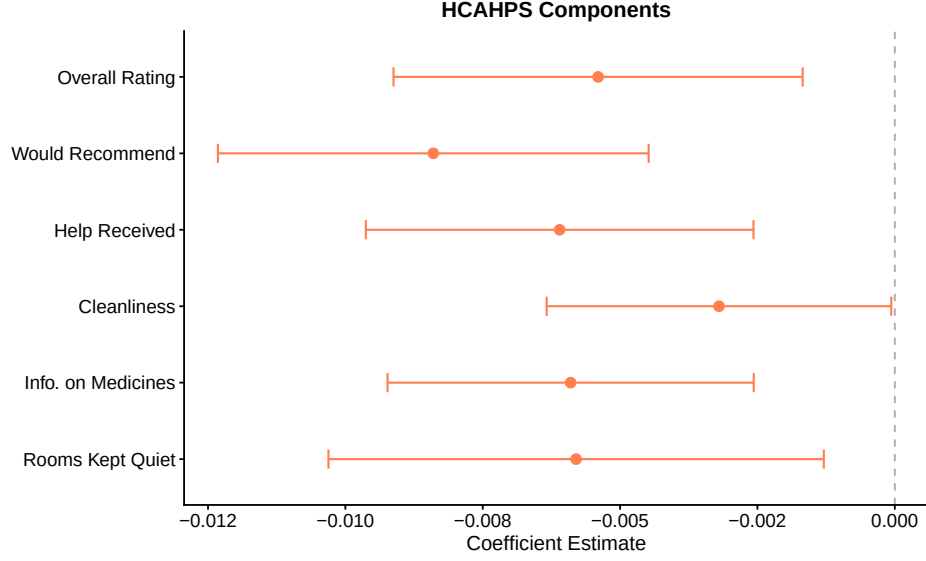


Figure 3: Effect of Medicare share on HCAHPS patient satisfaction components. Each point is an IV estimate; bars show 95% confidence intervals. See Section 3.3 for specification details.

indicating substantially longer wait times in hospitals with higher Medicare share. The standard-of-care composite—which captures nursing-sensitive outcomes such as hospital-acquired infections, pressure ulcers, and in-hospital falls—increases by 0.110 standard deviations ($SE = 0.050$), pointing to a deterioration in routine clinical processes. The HCAHPS composite declines by -0.057 standard deviations ($SE = 0.014$), indicating a broad-based decline in patient experience. The two aggregate quality indexes, which combine spending, satisfaction, and mortality-readmission into a single measure, both decline: the z-score index by -0.033 standard deviations per percentage point of Medicare share ($SE = 0.016$) and the PCA index by -0.044 ($SE = 0.021$). Appendix A.4 presents the individual components of the standard-of-care composite.

3.4 Additional Robustness Checks

The estimates above show that spending, staffing, patient experience, and clinical outcomes deteriorate as a hospital’s Medicare share rises. I now discuss other possible explanations for these findings and additional robustness checks, and report the full results in Appendix A.5.

Patient composition. The instrument identifies the effect of the Medicare share from predetermined population aging. Hospitals located in areas where the elderly share rises therefore face not only a change in the payer covering their patients, but also a change in patient composition. Older patients are sicker and require more care during a hospital stay, so measured outcomes could deteriorate as a hospital’s caseload ages even if the care it provides is unchanged. Patient composition, however, is unlikely to account for the estimates, for three reasons. First, CMS adjusts the patient-experience and clinical measures for patient characteristics: HCAHPS scores for patient mix and survey mode (Elliott et al., 2009), and mortality and readmission rates for age, sex, and

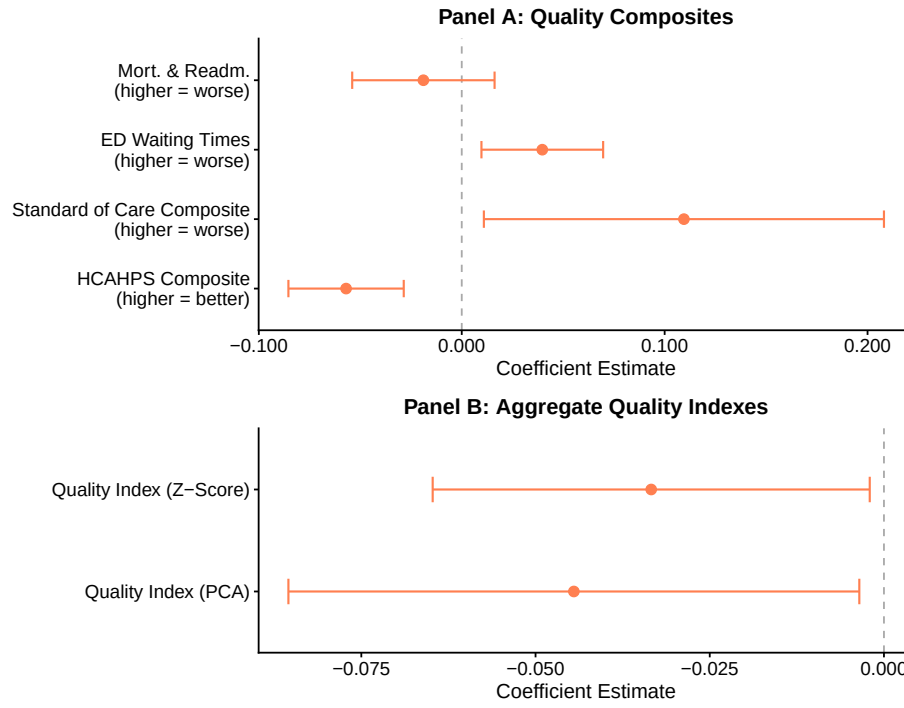


Figure 4: Effect of Medicare share on quality composites and aggregate quality indexes. Each point is an IV estimate; bars show 95% confidence intervals. Composites and indexes are standardized to zero mean and unit variance; coefficients are in standard-deviation units. See Section 3.3 for specification details.

comorbidities (Krumholz et al., 2006). Second, we would expect an older and sicker caseload to require higher resource utilization and staffing, which would push spending and staffing per patient upward. I find the opposite. Third, the estimates are essentially unchanged when I control directly for the hospital’s Medicare case-mix index, the average DRG weight of its Medicare discharges (Appendix A.5.1).

Hospital congestion. Aging raises hospital use as well as the Medicare share: older adults are hospitalized more often, so hospitals in aging areas may treat more patients overall. If beds and staffing adjust slowly, spending and staffing per patient would fall because the hospital is fuller, and not because it provides less to each patient. Controlling for the hospital’s total discharges leaves the estimates essentially unchanged (Panel B of Table 10 in Appendix A.5.1).

Differential trends. Over the sample period, the cross-county variation in the instrument comes from the share of county residents aged 36–45 in 1990, the cohorts that cross 65 between 2010 and 2019. While the hospital fixed effects absorb any time-invariant hospital and market characteristics, the remaining concern is that counties with a high share of these residents in 1990 were on different quality trends for reasons unrelated to the share of Medicare patients. Following Goldsmith-Pinkham, Sorkin, and Swift (2020), I test whether predicted 2010–2019 aging predicts changes in hospital outcomes over 2000–2009, before these cohorts reach the Medicare eligibility age of 65. Because the satisfaction and clinical measures are not reported far enough before 2010 to provide a pre-period, I run the test on the spending and staffing outcomes, which the cost reports cover back to 2000. I find that the staffing outcomes show no pre-period relationship with the predicted increase in the Medicare share. Hospitals located in counties with a larger predicted increase, however, had growing spending per patient over 2000–2009, such that a continuing trend would bias the estimates upward, in the opposite direction of my findings. Finally, I re-estimate the 2010–2019 regressions adding each hospital’s 2000–2009 outcome change as a control, so that the comparison is between hospitals that were on the same trajectory before 2010. The estimates are similar in magnitude (Appendix A.5.2).

4 Model

The reduced-form evidence in Section 3 establishes that a rising Medicare share lowers hospital quality, but cannot quantify welfare consequences or evaluate policy responses. To address these questions, I model a hospital market in which hospitals serve patients from two distinct payer types—Medicare and commercially insured. Hospitals first choose quality, after which they bilaterally negotiate commercial prices with insurers. Medicare reimburses hospitals at administratively determined rates, which include Medicare’s quality adjustment. Commercial prices, by contrast, are negotiated bilaterally between hospitals and insurers, who maintain provider networks on behalf of their enrollees. When choosing quality, hospitals internalize how quality affects (i) their demand from both Medicare and commercial patients, (ii) the commercial prices they negotiate with insurers and the Medicare rate they receive, and (iii) their cost of treating patients.

The interaction unfolds in three stages. In the first stage, each hospital simultaneously chooses a quality level y_{jt} , anticipating the Stage 2 bargaining outcome. In the second stage, given quality, each in-network hospital-insurer pair (j, g) bilaterally negotiates the commercial price, taking all other pairs' agreements as given. In the third stage, patients choose a hospital, given the quality levels $\{y_{jt}\}$.

4.1 Demand

Consider a market t with J_t hospitals indexed $j = 1, \dots, J_t$ and G_t commercial insurers indexed $g = 1, \dots, G_t$. Each insurer g maintains a provider network $\mathcal{N}_{gt} \subseteq \{1, \dots, J_t\}$. Patient i has payer type $p \in \{M, C\}$: she is either a Medicare beneficiary ($p = M$) or a commercial enrollee of one of the G_t insurers ($p = C$). For commercial patients, let $g(i) \in \{1, \dots, G_t\}$ denote the insurer covering patient i . I write $\mathcal{I}_t^M$ for the set of Medicare patients in market t and $\mathcal{I}_{gt}^C$ for the set of commercial enrollees of insurer g in market t . When a patient requires hospitalization, she chooses a hospital from her available choice set to maximize her utility.¹¹ The choice set differs by payer type: a Medicare patient can choose any of the J_t hospitals in the market, while a commercial enrollee of insurer g is restricted to the hospitals in g 's network $\mathcal{N}_{gt}$. Both types of patients also have an outside option, indexed $j = 0$, representing care received outside the market. The indirect utility i receives from treatment at hospital j depends on the patient's distance to the hospital (d_{ij}), the hospital's chosen quality (y_{jt}), a predetermined quality component (ξ_{jt}), and an idiosyncratic taste shock (ε_{ij}). I assume commercial enrollees face no out-of-pocket cost at the point of care. This reflects the institutional setting of employer-sponsored insurance, where copayments for inpatient care are typically small relative to total charges.¹² Prices therefore do not enter the utility function.

Formally, the utility that patient i derives from choosing hospital j is

$$U_{ij}^p = \alpha_i^p d_{ij} + \theta_i^p y_{jt} + \xi_{jt}^p + \varepsilon_{ij}, \quad (5)$$

where $V_{ij}^p \equiv \alpha_i^p d_{ij} + \theta_i^p y_{jt} + \xi_{jt}^p$ denotes the deterministic component of utility.

I assume that the idiosyncratic shock ε_{ij} is distributed i.i.d. Type-I extreme value. I allow patients' cost of traveling (α_i^p) and preference for quality (θ_i^p) to depend on their demographic characteristics. The utility of the outside option (seeking care at an out-of-market hospital) is

$$U_{i0}^p = \eta_t^p + \varepsilon_{i0},$$

where η_t^p is a year-specific intercept that allows the mean attractiveness of the outside option to vary over time. The deterministic component of the outside option is $V_{i0}^p \equiv \eta_t^p$.

Given the Type-I extreme value distribution of ε_{ij} , the probability that patient i chooses hospital

¹¹Patients may have multiple hospitalizations per year. For notational simplicity I suppress the hospitalization subscript and let i index a patient-hospitalization pair.

¹²Gowrisankaran et al. (2015) report that patients pay an average of 2–3 percent of the hospital bill as coinsurance in their sample. Duffy, Ly, and Trish (2026) find that cost-sharing obligations average 4–6 percent of facility allowed amounts for inpatient services nationally.

j takes the standard multinomial logit form:

$$\sigma_{ij}^p = \frac{\exp(V_{ij}^p)}{\sum_{k \in \mathcal{C}_i^p \cup \{0\}} \exp(V_{ik}^p)}, \quad (6)$$

where $\mathcal{C}_i^p$ is patient i 's choice set. For Medicare patients ($p = M$), their choice set is composed of all hospitals in the market ($\mathcal{C}_i^M = \{1, \dots, J_t\}$). The choice set of commercially insured patients is restricted to their insurer's network ($\mathcal{C}_i^C = \mathcal{N}_{g(i)}$).

Patients also differ by the severity of their illness, which I denote by w_i , and which captures the expected resource intensity of the hospitalization. Both Medicare reimbursement and hospital costs scale with w_i , so demand and prices throughout the model are measured in w_i -weighted units: hospital demand from each payer type is a w_i -weighted sum of patient-level choice probabilities, and prices are per unit of w_i . Following [Gowrisankaran et al. \(2015\)](#), the empirical counterpart of w_i is the DRG weight described in Section 2. Because commercial insurers commonly incorporate DRG weights into hospital contracts, measuring both Medicare and commercial prices in DRG-weighted units is consistent with the institutional setting. The severity-weighted demand at hospital j for Medicare patients and commercial patients insured by g is therefore given by:

$$Q_{jt}^M = \sum_{i \in \mathcal{I}_t^M} w_i \sigma_{ij}^M, \quad Q_{jgt}^C = \sum_{i \in \mathcal{I}_{gt}^C} w_i \sigma_{ij}^C. \quad (7)$$

4.2 Insurers

Each insurer g manages a provider network $\mathcal{N}_{gt}$ on behalf of its enrollees. I take each insurer's enrollee pool as given. I follow [Gowrisankaran et al. \(2015\)](#) by assuming that insurers maximize enrollee welfare minus total payments to hospitals. The value that insurer g derives from network $\mathcal{N}_{gt}$ is therefore given by:

$$V_g(\mathcal{N}_{gt}) = \tau \sum_{i \in \mathcal{I}_{gt}^C} W_i(\mathcal{N}_{gt}) - \sum_{j \in \mathcal{N}_{gt} \cup \{0\}} Q_{jgt}^C \cdot P_{jgt}^C, \quad (8)$$

where $W_i(\mathcal{N}_{gt}) = \log(\exp(\eta_t^C) + \sum_{k \in \mathcal{N}_{gt}} \exp(V_{ik}^C))$ is enrollee i 's expected utility from the network (the McFadden surplus) and the second term is total payments to hospitals, including payments for care received through the outside option ($j = 0$). The parameter $\tau > 0$ is the dollar value the insurer places on one utility unit of enrollee welfare.¹³

The insurer's surplus from including hospital j in its network is the change in the value of its

¹³In [Gowrisankaran et al. \(2015\)](#), the analogous parameter is the ratio τ/α , where α is the coefficient associated with out-of-pocket spending (utils per dollar) and τ is a welfare weight measuring how insurers value enrollee welfare relative to spending. Under zero coinsurance, the utility function has no out-of-pocket cost coefficient, so unit conversion and welfare weighting cannot be separately identified. I therefore treat τ here as a single composite parameter—corresponding to τ/α in their setup—and estimate it directly.

network, relative to the case in which j is excluded:

$$\Delta V_{jg} \equiv V_g(\mathcal{N}_{gt}) - V_g(\mathcal{N}_{gt} \setminus j). \quad (9)$$

Expanding (9):

$$\Delta V_{jg} = \underbrace{\tau \cdot GW_{jg}}_{\text{enrollee welfare}} - \underbrace{Q_{jgt}^C \cdot P_{jgt}^C}_{\text{payment to } j} + \underbrace{\sum_{k \in (\mathcal{N}_{gt} \setminus \{j\}) \cup \{0\}} \Delta_g[j, k] \cdot P_{kgt}^C}_{\text{payment reallocation}}. \quad (10)$$

The first term, $\tau \cdot GW_{jg}$, is the dollar value of the welfare enrollees gain when hospital j is added to their choice set. Under the distributional assumptions above, including hospital j increases enrollee i 's expected utility by $-\log(1 - \sigma_{ij}^C)$, so the total welfare gain from including j is $GW_{jg} = -\sum_{i \in \mathcal{I}_{gt}^C} \log(1 - \sigma_{ij}^C)$, scaled to dollars by τ . The second term, $Q_{jgt}^C \cdot P_{jgt}^C$, is the direct payment from insurer g to hospital j . The third term captures payment reallocation: when j is excluded, its commercial patients redistribute to other in-network hospitals and the outside option, generating offsetting payments at those alternatives. The term $\Delta_g[j, k]$ is the additional DRG-weighted demand at alternative k when j is removed from the network, and P_{kgt}^C is the corresponding price (P_{0gt}^C denotes the average price of care received through the outside option).

Intuitively, when the insurer considers dropping hospital j , it loses $\tau \cdot GW_{jg}$ in enrollee welfare, but saves the direct payment. However, the insurer does not save the full amount because patients displaced from j seek care at other hospitals or through the outside option, at potentially different prices. The insurer surplus ΔV_{jg} is the net of these effects. Formal definitions of GW_{jg} and $\Delta_g[j, k]$ are in Appendix B.

4.3 Hospital Costs

Producing a discharge requires inputs whose cost depends on the hospital's chosen quality. Following Grieco and McDevitt (2017), I model this as a quality-quantity tradeoff: quality enters the production technology by scaling the effective resource requirement per discharge, so a hospital choosing higher quality must employ more inputs per case. Each discharge combines two complementary inputs in fixed proportions: a *medical service*, the cost of which scales with quality; and an *administrative service* whose per-unit cost varies by payer but does not scale with quality.

I assume hospitals produce medical service by combining multiple factor inputs using a constant-returns-to-scale Cobb-Douglas production technology. Since output is measured in DRG-weighted units, a discharge of severity w_i contributes w_i to Q_{jt} and requires w_i units of each input, so treating a more resource-intensive case is proportionally more costly. Cost minimization over the factor inputs delivers the following per-DRG-weight medical cost (derived in Appendix C):

$$f(y_{jt}) = \frac{e^{\alpha_Y y_{jt}}}{A_t}, \quad (11)$$

where A_t is the market-time effective productivity absorbing both physical TFP and the cost of the variable-factor bundle in market t , and $\alpha_Y > 0$ governs the elasticity of medical cost with respect to quality. Differentiating yields:

$$\frac{\partial f}{\partial y_{jt}} = \alpha_Y \cdot f(y_{jt}), \quad (12)$$

so raising quality by one unit increases the per-DRG-weight medical cost by a fraction α_Y of its current value.

The medical component $f(y_{jt})$ is common across payers because medical service is produced through a single technology that does not distinguish between Medicare and commercial cases. Marginal cost also includes a structural residual that captures how each hospital's administrative costs, such as pre-authorization requirements, paperwork burden, or reimbursement speed, deviate from the average. Per-DRG-weight marginal cost of treating a Medicare patient is then given by

$$MC_{jt}^M = f(y_{jt}) + \nu_{jt}^M, \quad (13)$$

and for a patient enrolled in commercial insurer g ,

$$MC_{jgt}^C = f(y_{jt}) + \nu_{jgt}^C, \quad (14)$$

where ν_{jt}^M is the Medicare residual, common to all of hospital j 's Medicare discharges in year t , and ν_{jgt}^C is the commercial residual, specific to each hospital-insurer pair. Appendix C provides a Leontief microfoundation for this additive cost structure.

Three assumptions—(i) constant returns to scale in the medical technology, (ii) competitive factor markets so that factor prices do not respond to hospital volume, and (iii) the absence of capacity constraints—together imply that marginal cost is independent of the number of discharges the hospital handles, which is what allows the bargaining problem to separate across insurers in Section 4.4.

4.4 Stage 2: Bargaining over Commercial Prices

I solve the model by backward induction. Given quality levels $\{y_{jt}\}$ chosen in Stage 1, both hospitals and insurers take the demand and cost objects from Sections 4.1–4.3 as given. Following Gowrisankaran et al. (2015), each hospital-insurer pair (j, g) with $j \in \mathcal{N}_{gt}$ bilaterally maximizes a Nash product over the per-DRG-weight commercial price P_{jgt}^C , taking all other pairs' agreed prices as given. Assuming equal bargaining weights, the Nash product for pair (j, g) is¹⁴

$$\max_{P_{jgt}^C} [Q_{jgt}^C \cdot (P_{jgt}^C - MC_{jgt}^C)]^{1/2} \cdot [\Delta V_{jg}(P_{jgt}^C, \mathbf{P}_{-j,gt}^C)]^{1/2}, \quad (15)$$

¹⁴Gowrisankaran et al. (2015) argue that, interpreting bargaining weights as relative discount factors in an alternating-offers game (Rubinstein, 1982), equal weights are the most natural assumption in models where agents share a common discount factor.

where $\mathbf{P}_{-j,gt}^C$ denotes the vector of prices that insurer g has agreed with all other in-network hospitals. I take each insurer’s enrollee pool and network as given during the bargaining stage. Since enrollees cannot switch insurers in response to bargaining outcomes, disagreement means hospital j is excluded from insurer g ’s network and loses all commercial demand from g ’s enrollees. The hospital’s gain from agreement is therefore its commercial profit from insurer g , $Q_{jgt}^C(P_{jgt}^C - MC_{jgt}^C)$, while disagreement yields zero. The insurer’s gain from agreement is ΔV_{jg} , the surplus from including hospital j in its network (Section 4.2). Under zero coinsurance, demand does not respond to prices. The insurer surplus ΔV_{jg} , however, depends on the own price and on other in-network hospitals’ prices through the payment terms in (10).

The equilibrium of the bargaining stage is a price vector

$$\{P_{jgt}^{C*} : j \in \mathcal{N}_{gt}, g = 1, \dots, G_t\}$$

such that for every hospital–insurer pair (j, g) with $j \in \mathcal{N}_{gt}$, the price P_{jgt}^{C*} maximizes the bilateral Nash product (15) taking all other pairs’ equilibrium prices as given.

Under zero coinsurance, the first-order condition of (15) simplifies to:

$$P_{jgt}^C = MC_{jgt}^C + \frac{\Delta V_{jg}(\mathbf{P}_{gt}^C)}{Q_{jgt}^C}. \quad (16)$$

Price equals marginal cost plus a markup equal to the insurer’s surplus per DRG-weighted discharge.¹⁵ Multiplying both sides by Q_{jgt}^C yields

$$Q_{jgt}^C(P_{jgt}^C - MC_{jgt}^C) = \Delta V_{jg}, \quad (17)$$

so that hospital j ’s commercial profit from insurer g equals the insurer’s surplus from including j in its network. That is, under zero coinsurance and equal bargaining weights, each side’s gain from reaching an agreement is equal.¹⁶ The markup is hospital–insurer specific, reflecting the value that each hospital adds to a particular insurer’s network. A hospital that is the primary option for many of an insurer’s enrollees—for instance, because it is the only in-network hospital in their area—generates a large surplus per discharge if excluded, and under equal bargaining weights this translates into a higher markup. The same hospital may command a lower markup from a different insurer whose enrollees have more nearby alternatives.

Since ΔV_{jg} depends on the prices of all other in-network hospitals through the payment reallocation term in (10), equation (16) defines a system of $|\mathcal{N}_{gt}|$ simultaneous equations that determines the equilibrium price vector for each insurer. The price interdependence is entirely within insurer:

¹⁵The derivation of (16) from the first-order condition of (15) is in Appendix D.

¹⁶Under zero coinsurance, a price increase is a transfer from the insurer to the hospital that does not change the total surplus being divided, so the Nash bargaining solution splits the gains from trade in proportion to the bargaining weights. With coinsurance, prices also reallocate patients across hospitals, and the solution instead equates each side’s marginal gain relative to its total gain, weighted by its bargaining weight (Gowrisankaran et al., 2015, equations 12 and 17).

insurer g 's demand objects $(Q_{jgt}^C, GW_{jg}, \Delta_g[j, k])$ depend only on its own enrollee pool and network, and marginal cost is independent of volume. Each insurer's pricing system can therefore be solved independently, reducing the problem to G_t separate systems per market.

Substituting the definition of ΔV_{jg} into the bargaining FOC, we get:

$$P_{jgt}^C = \frac{1}{2} \left(MC_{jgt}^C + \underbrace{\tau \frac{GW_{jg}}{Q_{jgt}^C} + \sum_{k \in (\mathcal{N}_{gt} \setminus \{j\}) \cup \{0\}} \frac{\Delta_g[j, k]}{Q_{jgt}^C} P_{kgt}^C}_{\equiv WTP_{jgt}} \right). \quad (18)$$

WTP_{jgt} is the most that insurer g would pay per DRG-weighted patient to keep hospital j in its network. The negotiated price is therefore the midpoint of the segment from the hospital's marginal cost to the insurer's willingness to pay.

Because hospitals choose quality anticipating the outcome of bargaining, their quality choice depends on how negotiated prices respond to quality. To derive the price response dP_{jgt}^{C*}/dy_{jt} , consider first the simplest network: an insurer with a single in-network hospital, so that its commercial enrollees choose between hospital j and an outside option with an exogenously determined price P_{0gt}^C . The pricing equation (18) is then a single equation in P_{jgt}^C , with $WTP_{jgt} = P_{0gt}^C + \tau GW_{jg}/Q_{jgt}^C$, and differentiating it with respect to y_{jt} gives

$$\frac{dP_{jgt}^{C*}}{dy_{jt}} = \frac{1}{2} \left[\underbrace{\alpha_Y \cdot f(y_{jt})}_{\text{cost of quality (+)}} + \underbrace{\tau \frac{d}{dy_{jt}} \left(\frac{GW_{jg}}{Q_{jgt}^C} \right)}_{\text{welfare per patient } (\pm)} \right]. \quad (19)$$

Higher quality moves the negotiated price through two forces. The first is cost pass-through: quality raises marginal cost by $\alpha_Y f$, and bargaining passes half of it into the negotiated price. The second is the change in the insurer's willingness to pay per patient. On the one hand, higher quality raises the welfare enrollees gain from adding j to the network. On the other hand, higher quality attracts more patients, which raises Q_{jgt}^C and spreads that welfare gain over a larger volume. The welfare-per-patient term is therefore positive only when $(\partial GW_{jg}/\partial y_{jt})/GW_{jg} > (\partial Q_{jgt}^C/\partial y_{jt})/Q_{jgt}^C$, that is, when the welfare gain rises proportionally faster than DRG-weighted demand. Appendix E shows that with logit demand and identical enrollees this condition always holds, so negotiated commercial prices increase with quality. More generally, with asymmetric enrollees, this need not be the case. Figure 5 illustrates how quality moves the endpoints of the bargaining range and, therefore, the equilibrium price.

Two features change with more than one in-network hospital. Dropping j now redistributes its patients across rival hospitals as well as the outside option, at potentially different prices, and the prices of all in-network hospitals are jointly determined. Differentiating (18) with respect to y_{jt}

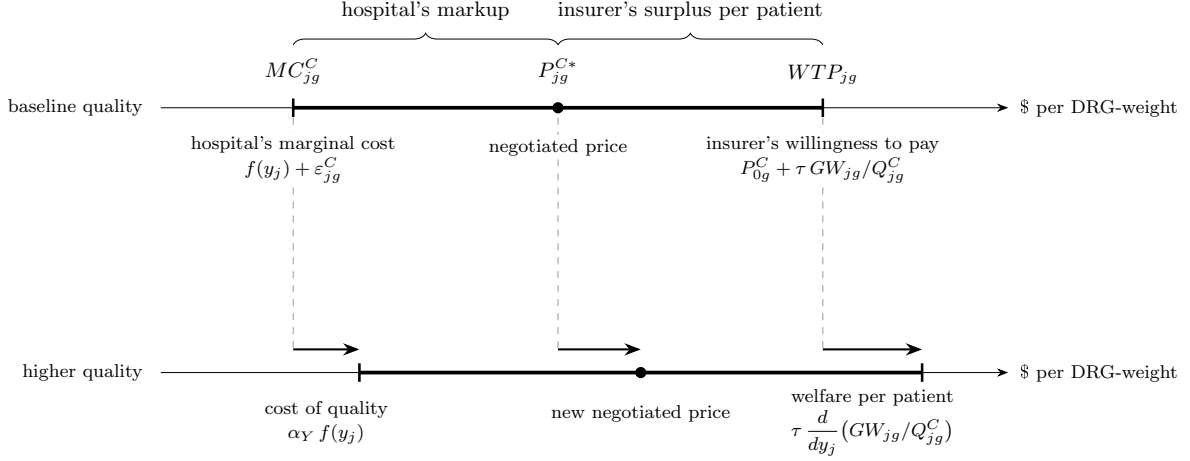


Figure 5: The negotiated price and its response to quality in the single-hospital case. In the top row, the negotiated price is the midpoint of the segment from the hospital's marginal cost to the insurer's willingness to pay per DRG-weighted patient. In the bottom row, higher quality moves the lower endpoint by $\alpha_Y f(y_{jt})$ and the upper endpoint by $\tau d(GW_{jg}/Q_{jgt}^C)/dy_{jt}$, and the negotiated price moves by one half of the sum of the two movements (19).

gives

$$\begin{aligned}
 \frac{dP_{jgt}^{C*}}{dy_{jt}} = & \frac{1}{2} \left[\underbrace{\alpha_Y \cdot f(y_{jt})}_{\text{cost of quality (+)}} + \underbrace{\tau \frac{d}{dy_{jt}} \left(\frac{GW_{jg}}{Q_{jgt}^C} \right)}_{\text{welfare per patient (\pm)}} + \underbrace{\sum_{k \in (\mathcal{N}_{gt} \setminus \{j\}) \cup \{0\}} P_{kgt}^C \frac{d}{dy_{jt}} \left(\frac{\Delta_g[j, k]}{Q_{jgt}^C} \right)}_{\text{diversion (\pm)}} \right. \\
 & \left. + \underbrace{\sum_{k \in \mathcal{N}_{gt} \setminus \{j\}} \frac{\Delta_g[j, k]}{Q_{jgt}^C} \frac{dP_{kgt}^{C*}}{dy_{jt}}}_{\text{rivals' prices (\pm)}} \right]. \tag{20}
 \end{aligned}$$

Higher quality attracts patients unevenly across the market, changing the share of hospital j 's patients that would reallocate to each alternative if j were excluded. The diversion term is the resulting change in the average price the insurer would pay for hospital j 's patients if j were excluded, with each alternative's price weighted by the share of patients that would reallocate to it. It is positive when the reallocation shifts toward more expensive alternatives. The rivals'-price term captures how changes in its competitors' prices affect the price the insurer would pay in j 's absence. If j 's rivals respond by lowering their prices, excluding j becomes cheaper, since the insurer would pay less at the alternatives patients would reallocate to, and the insurer's willingness to pay for j falls.

A hospital's price also changes with the quality of its competitors. Differentiating hospital k 's pricing equation (18) with respect to y_{jt} shows how its equilibrium price changes when one of its competitors increases its quality. The expression, written out in Appendix E, differs from the own-quality response (20) in two ways. First, the cost-of-quality term drops out, since hospital k 's marginal cost does not depend on hospital j 's quality. Second, higher quality at j lowers

both the welfare hospital k adds to the network and its demand. The welfare-per-patient term is therefore negative when the welfare hospital k adds falls proportionally faster than its demand. The diversion term is again the change in the average price the insurer would pay for k 's patients if k were excluded. Higher quality at j raises the share of k 's patients that would reallocate to j , so the term is positive when j is expensive relative to k 's other alternatives. The rivals'-price term now includes hospital j 's own price response. If j 's price rises with its quality, excluding k becomes more expensive, since the insurer would pay more at the alternatives k 's patients would reallocate to, and the insurer's willingness to pay for k rises.

The derivative of each hospital's pricing equation (18) with respect to y_{jt} contains the other hospitals' price responses through the rivals'-price term, so the price responses of all in-network hospitals are jointly determined. Stacking the equations gives

$$\frac{d\mathbf{P}_g^C}{dy_{jt}} = \frac{1}{2} \left(\mathbf{v}_g + \mathbf{D}^g \frac{d\mathbf{P}_g^C}{dy_{jt}} \right), \quad (21)$$

where $d\mathbf{P}_g^C/dy_{jt}$ is the vector of price responses, $\mathbf{v}_g$ collects the cost-of-quality, welfare-per-patient, and diversion terms of each equation, and $\mathbf{D}^g$ is a $|\mathcal{N}_{gt}| \times |\mathcal{N}_{gt}|$ matrix with elements $D_{kl}^g = \Delta_g[k, l]/Q_{kgt}^C$ for $k \neq l$, the share of hospital k 's patients that would reallocate to hospital l if k were excluded, and $D_{kk}^g = 0$ (zeros on the diagonal). Solving for the price responses yields

$$\frac{d\mathbf{P}_g^C}{dy_{jt}} = (2\mathbf{I} - \mathbf{D}^g)^{-1} \mathbf{v}_g, \quad (22)$$

where $2\mathbf{I} - \mathbf{D}^g$ is invertible, since its diagonal elements equal two and each row of $\mathbf{D}^g$ sums to at most one. Appendix E derives each row of the system.

4.5 Medicare Payments

Medicare pays each hospital an administered base rate per DRG-weight, $\bar{P}_{jt}^M$, set by a national formula and adjusted for local wages. Of Medicare's three quality programs (see Section 2), the largest is the Hospital Value-Based Purchasing program (HVBP), which withholds 2 percent of each hospital's base DRG payments and redistributes the pool based on quality scores. In fiscal year 2019, HVBP redistributed about \$1.9 billion (Centers for Medicare & Medicaid Services, 2018). Because HVBP only redistributes payments across hospitals, it is budget neutral and it does not change total Medicare spending. I model a stylized program based on HVBP that redistributes the withhold according to the quality y_{jt} chosen by each hospital and is budget neutral within the market.¹⁷

The program withholds a share $\omega_t \in [0, 1]$ of every hospital's base payments (2 percent under HVBP) and returns the pool in proportion to quality, so that the realized rate depends on the

¹⁷The actual program withholds and redistributes over the national pool of IPPS hospitals. In the program I model, transfers therefore net to zero across the market's hospitals instead of across the entire country.

hospital's own quality:

$$P_{jt}^M(y_{jt}) = \underbrace{\bar{P}_{jt}^M (1 - \omega_t)}_{\text{post-withhold base}} + \underbrace{b_t \bar{P}_{jt}^M (y_{jt} - y_{\min})}_{\text{quality payment}}, \quad (23)$$

where b_t is the payment per unit of quality, expressed as a fraction of the hospital's own base rate, and $y_{\min}$ is the lowest quality across hospitals in the market. The worst a hospital can do under the program is earn none of its withhold back and keep $(1 - \omega_t) \bar{P}_{jt}^M$.

Given the withhold ω_t , budget neutrality determines the slope b_t . Total quality payments must equal the total dollars withheld,

$$b_t \sum_{j=1}^{J_t} Q_{jt}^M \bar{P}_{jt}^M (y_{jt} - y_{\min}) = \omega_t \sum_{j=1}^{J_t} Q_{jt}^M \bar{P}_{jt}^M, \quad (24)$$

and solving for b_t yields

$$b_t = \frac{\omega_t}{\bar{y}_t - y_{\min}}, \quad (25)$$

where $\bar{y}_t$ is the Medicare-revenue-weighted mean of equilibrium quality. Figure 6 illustrates the design.

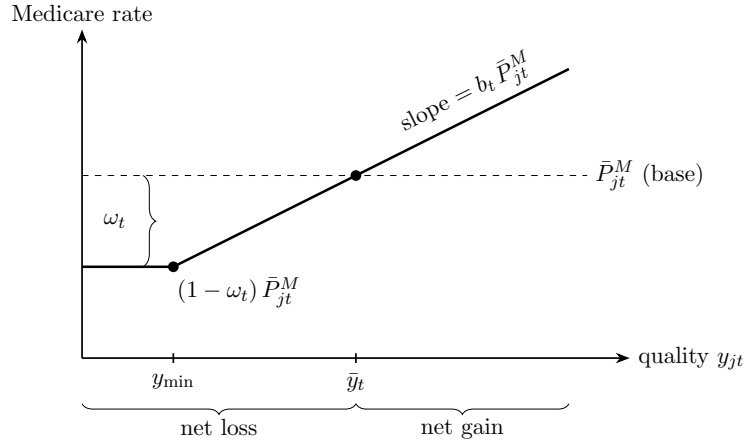


Figure 6: The realized Medicare rate as a function of hospital quality under the stylized version of Value-Based Purchasing. The lowest-quality hospital earns none of its withhold back and keeps $(1 - \omega_t) \bar{P}_{jt}^M$, the rate rises with quality at slope $b_t \bar{P}_{jt}^M$, and a hospital at the revenue-weighted mean quality $\bar{y}_t$ is paid its base rate. Below $y_{\min}$ the quality payment is zero, so the rate is flat at $(1 - \omega_t) \bar{P}_{jt}^M$.

Substituting $\omega_t = b_t(\bar{y}_t - y_{\min})$ back into (23) expresses the realized rate as a deviation from the mean,

$$P_{jt}^M(y_{jt}) = \bar{P}_{jt}^M + b_t \bar{P}_{jt}^M (y_{jt} - \bar{y}_t), \quad (26)$$

so a hospital at the mean quality is paid exactly its base rate, hospitals above the mean are paid more, and hospitals below it are paid less. I assume hospitals take the slope b_t and the minimum

$y_{\min}$ as given when choosing quality. A hospital could move the slope only through the mean $\bar{y}_t$, and in the actual program that mean is computed over the thousands of hospitals paid under the IPPS nationwide, where any single hospital carries negligible weight.

The Medicare rate therefore responds to quality:

$$\frac{\partial P_{jt}^M}{\partial y_{jt}} = b_t \bar{P}_{jt}^M. \quad (27)$$

4.6 Stage 1: Hospital Quality Choice

In the first stage, hospitals simultaneously choose quality levels. Each hospital j takes other hospitals' quality choices $\mathbf{y}_{-j,t}$ as given and solves the following problem:

$$\max_{y_{jt}} \pi_{jt} = P_{jt}^M \cdot Q_{jt}^M + \sum_{g=1}^{G_t} P_{jgt}^C \cdot Q_{jgt}^C - MC_{jt}^M \cdot Q_{jt}^M - \sum_{g=1}^{G_t} MC_{jgt}^C \cdot Q_{jgt}^C, \quad (28)$$

subject to the demand functions (7), the cost functions (13)–(14), the bargaining first-order condition (16) that determines commercial prices, and the stylized HVBP program (23), through which the Medicare rate P_{jt}^M depends on quality. Both demand and equilibrium commercial prices depend on the full quality vector, so each hospital's optimal quality depends on its rivals' choices.

Setting $\partial \pi_{jt} / \partial y_{jt} = 0$ and using $\partial MC / \partial y_{jt} = \alpha_Y f$ from (12), the quality first-order condition is:

$$\begin{aligned} & \underbrace{(P_{jt}^M - MC_{jt}^M) \frac{\partial Q_{jt}^M}{\partial y_{jt}}}_{\text{Medicare demand}} + \underbrace{Q_{jt}^M \frac{\partial P_{jt}^M}{\partial y_{jt}}}_{\text{Medicare price}} + \underbrace{\sum_{g=1}^{G_t} (P_{jgt}^C - MC_{jgt}^C) \frac{\partial Q_{jgt}^C}{\partial y_{jt}}}_{\text{commercial demand}} + \underbrace{\sum_{g=1}^{G_t} Q_{jgt}^C \frac{dP_{jgt}^{C*}}{dy_{jt}}}_{\text{commercial price}} \\ & = \underbrace{\alpha_Y \cdot f \left(Q_{jt}^M + \sum_{g=1}^{G_t} Q_{jgt}^C \right)}_{\text{marginal cost}}. \end{aligned} \quad (29)$$

The demand terms are the marginal revenue from attracting additional patients: higher quality draws patients from both payer types, each contributing their respective margin $P - MC$. The commercial price term captures the equilibrium response of negotiated prices to quality, derived in Section 4.4. For each insurer g , the commercial price response dP_{jgt}^{C*} / dy_{jt} is the j -th element of expression (22). The Medicare price term captures how Medicare rates change with quality under the stylized HVBP program (27). The marginal cost term captures how increasing quality raises total costs of treating both Medicare and commercial patients. An equilibrium of the quality game is a vector $\mathbf{y}_t^* = (y_{1t}^*, \dots, y_{J_t}^*)$ such that every hospital's first-order condition (29) is satisfied simultaneously, given the Stage 2 pricing equilibrium induced by $\mathbf{y}_t^*$.

5 Demand Estimation

5.1 Data

I estimate demand using patient-level claims and demographic data from Massachusetts, covering the years 2018–2022. The Medicare and commercial claims are held on separate secure servers, so I estimate demand separately for each payer. Both samples are restricted to non-emergency inpatient stays at general acute-care hospitals. I exclude emergency admissions because patients requiring emergency care are typically transported to the nearest hospital and do not exercise meaningful choice over where to receive treatment.

Medicare. I construct the Medicare sample from two CMS administrative datasets. The Medicare Provider Analysis and Review (MedPAR) file contains a record for each inpatient hospital stay, reporting the admitting hospital, admission type, and the assigned Diagnosis-Related Group (DRG) code. I link each stay to the Master Beneficiary Summary File (MBSF), which records beneficiary demographics such as age, sex, race, Medicaid dual-eligibility status—an indicator of low income—and the beneficiary’s residential ZIP code. Both files cover the universe of Medicare fee-for-service beneficiaries, so the resulting dataset includes all inpatient hospitalizations for Medicare beneficiaries residing in Massachusetts over the sample period.

I restrict the sample to beneficiaries aged 65 and older¹⁸ admitted for at least one non-emergency inpatient stay during my sample period. The final sample contains approximately 262,000 unique beneficiaries and 303,000 hospitalizations across the five years, at approximately 55 Massachusetts acute-care hospitals per year. Beneficiaries with multiple hospitalizations in a year contribute one observation per stay.

For each hospitalization, I observe the hospital chosen by the patient, the DRG weight—a CMS-assigned measure of expected resource intensity relative to a reference case—and the distance from the patient’s residential ZIP code to each hospital.¹⁹ I also link each hospital-year to the quality index described in Section 3.1. I group hospitalizations at out-of-state hospitals, and at Massachusetts hospitals with fewer than 11 distinct patients in a year, into an outside option.

Commercial. I construct the commercial sample from the Massachusetts All-Payer Claims Database (MA-APCD), maintained by the Center for Health Information and Analysis (CHIA), the Massachusetts state agency that collects claims filed by insurers operating in the state ([Center for Health Information and Analysis, 2025](#)). I use three files: the Medical Claims file, which contains a record for each claim line billed for medical services; the Member Eligibility file, which records enrollee demographics and coverage periods; and the Product file, which identifies each enrollee’s insurance product type and line of business. Patient residential ZIP codes come from a separate geographic file.

I construct inpatient hospital stays from the Medical Claims file by filtering to hospital inpatient facility claims, applying carrier-specific versioning to deduplicate amended and replaced claims, and

¹⁸See footnote 3.

¹⁹I aggregate patient locations to three-digit ZIP codes, computing population-weighted average distances to each hospital within each three-digit zone. Massachusetts contains 18 three-digit ZIP zones.

grouping claim lines into stays.²⁰ Because the APCD does not include a Diagnosis-Related Group field, I assign DRG codes by running the CMS MS-DRG grouper software on the diagnosis and procedure codes reported in each stay, then map each code to its CMS DRG weight.

I restrict the sample to enrollees with commercial fully-insured coverage admitted for non-emergency inpatient stays at general acute-care hospitals, spanning 2018–2022.²¹ For each hospitalization, I observe the hospital chosen by the patient, the DRG weight, the distance from the patient’s residential ZIP code to each hospital, patient age and sex, and the identity of the patient’s insurer. Unlike the Medicare sample, the APCD does not report patient race or Medicaid dual-eligibility status. I link each hospital-year to the same quality index used in the Medicare sample. I group hospitalizations at out-of-state hospitals, and at hospitals with fewer than 11 distinct patients in a year, into the outside option, as in the Medicare sample.

I also use the APCD to construct insurer-specific hospital networks. I classify an insurer–hospital–year cell as in-network if at least 50 percent of that insurer’s non-emergency stays at that hospital are coded as in-network. Figure 19 in Appendix F.2 shows that the distribution of in-network fractions is concentrated at zero and one, so the classification is not sensitive to the choice of threshold.²²

5.2 Demand Model Specification

I estimate demand separately for Medicare ($p = M$) and commercial ($p = C$) patients. As described in Section 4.1, patient i chooses the hospital j in her choice set that maximizes

$$U_{ijt}^p = \alpha_i^p d_{ij} + \theta_i^p y_{jt} + \xi_{jt}^p + \varepsilon_{ijt}, \quad \varepsilon_{ijt} \sim \text{i.i.d. Type-I extreme value}, \quad (30)$$

where d_{ij} is the distance from patient i to hospital j (in miles), y_{jt} is the quality index described in Section 3.1, ξ_{jt}^p is an unobserved, predetermined component of hospital quality, and α_i^p and θ_i^p are patient-specific coefficients on distance and quality. The utility from the outside option—an out-of-state hospital—is $U_{i0t}^p = \eta_t^p + \varepsilon_{i0t}$, where η_t^p allows the mean attractiveness of the outside option to vary by year.

The choice set of a patient also depends on their insurer. Because Medicare beneficiaries face no network restrictions, Medicare patients may choose any Massachusetts acute-care hospital ($\mathcal{C}_i^M = \{1, \dots, J_t\}$), while commercial enrollees of insurer g are restricted to hospitals in g ’s network ($\mathcal{C}_i^C = \mathcal{N}_{g(i)t}$).

²⁰The MA-APCD reports claims at the claim-line level, with each line representing a distinct service or charge within a claim. I aggregate lines to the claim level and then group claims into hospitalization stays by matching on patient, facility, and admission month, combining interim-billed claims for long stays.

²¹Self-insured employer plans are largely absent from the APCD after the Supreme Court’s 2016 decision in *Gobeille v. Liberty Mutual Insurance Co.*

²²I exclude emergency admissions from the network classification because the Affordable Care Act requires insurers to cover emergency visits at any hospital regardless of network status.

I parameterize the heterogeneous coefficients as

$$\alpha_i^p = \alpha_0^p + \boldsymbol{\alpha}_{\text{dem}}^p \cdot \mathbf{X}_i + \alpha_{\text{sev}}^p w_i, \quad (31)$$

$$\theta_i^p = \theta_0^p + \theta_{\text{sev}}^p w_i, \quad (32)$$

where $\mathbf{X}_i$ is a vector of patient demographics and w_i is the patient’s severity (DRG weight). For Medicare patients, $\mathbf{X}_i$ includes age, sex, race, and Medicaid dual-eligibility status (a proxy for low income); for commercial patients, $\mathbf{X}_i$ includes age and sex only, as the APCD does not report race or Medicaid status. The severity interactions α_{sev}^p and θ_{sev}^p allow sicker patients—those with higher DRG weights—to respond differently to distance and quality.

I estimate the model in two stages. In the first stage, I define hospital-year fixed effects

$$\delta_{jt}^p \equiv \theta_0^p y_{jt} + \xi_{jt}^p, \quad (33)$$

which absorb both the mean quality effect and unobserved hospital quality. Substituting into the choice probability (6), the first-stage parameters $(\boldsymbol{\alpha}^p, \alpha_{\text{sev}}^p, \theta_{\text{sev}}^p, \boldsymbol{\delta}^p)$ are estimated by maximum likelihood.

In the second stage, I recover the mean quality coefficient θ_0^p by regressing the estimated fixed effects on the quality index:

$$\hat{\delta}_{jt}^p = \theta_0^p y_{jt} - \eta_t^p + \xi_{jt}^p, \quad (34)$$

where η_t^p are year fixed effects that absorb the time-varying outside-option normalization. Because y_{jt} is a choice variable in the model (Section 4.6) and may be correlated with the predetermined component ξ_{jt}^p , I instrument y_{jt} with the predicted Medicare share within 15, 30, and 50-mile radii of each hospital and the surrounding county, constructed from the predetermined aging instrument described in Section 3.2.

5.3 Demand Model Estimates

Table 1 reports demand estimates for both types of payers. My demand estimation results show that patients have a strong preference for hospital quality. The mean quality coefficient θ_0^p , recovered from the instrumented second-stage regression (34), is positive and statistically significant in both samples: 0.623 (SE = 0.177, first-stage $F = 15.9$) for Medicare and 0.751 (SE = 0.280, first-stage $F = 15.1$) for commercial patients (Panel B). Patients therefore choose hospitals that score more highly on the index, indicating that its components capture aspects of a hospital stay that matter to them. I also find that sicker patients value quality even more: the quality–severity interaction θ_{sev}^p is positive and significant for both payers (0.109, SE = 0.001 for Medicare; 0.091, SE = 0.003 for commercial), implying that patients with higher DRG weights—those requiring more intensive treatment—place greater weight on hospital quality when choosing where to seek care. Similarly, the distance–severity interaction α_{sev}^p shows that sicker patients are less distance-sensitive (0.003, SE = 0.000 for Medicare; 0.008, SE = 0.000 for commercial), consistent with a willingness to travel

Table 1: Demand Model Estimates

	Medicare ($p = M$)	Commercial ($p = C$)
<i>Panel A: First Stage (Maximum Likelihood)</i>		
θ_{sev} (quality $\times$ severity)	0.109*** (0.001)	0.091*** (0.003)
α_0 (distance)	-0.103*** (0.001)	-0.208*** (0.001)
α_{sev} (distance $\times$ severity)	0.003*** (0.000)	0.008*** (0.000)
α_{male} (distance $\times$ male)	0.003*** (0.000)	0.014*** (0.001)
α_{age} (distance $\times$ age)	-0.001*** (0.000)	0.001*** (0.000)
α_{white} (distance $\times$ white)	0.007*** (0.001)	—
$\alpha_{\text{low inc.}}$ (distance $\times$ low income)	-0.012*** (0.000)	—
<i>Panel B: Second Stage</i>		
θ_0 (quality, 2SLS)	0.623*** (0.177)	0.751** (0.280)
First-stage F	15.9	15.1
θ_0 (quality, OLS)	0.276*** (0.077)	0.706*** (0.118)

Notes: Standard errors in parentheses. Panel A reports maximum likelihood estimates of the heterogeneous logit demand model described in Section 5.2. Hospital-year fixed effects absorb the mean quality effect and unobserved hospital quality. Panel B reports the second-stage estimate of the mean quality coefficient, obtained by regressing estimated hospital-year fixed effects on the quality index. The 2SLS rows instrument the quality index with the predicted Medicare share within 15-, 30-, and 50-mile radii of each hospital and in the surrounding county (Section 3.2); the OLS rows are reported for comparison. Year fixed effects included in the second stage. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

farther to reach higher-quality hospitals.

To translate these results into an economically interpretable quantity, I compute hospital-level own-quality semi-elasticities, defined as the percentage change in a hospital’s expected patient volume from a one-standard-deviation improvement in quality. The average semi-elasticity is 0.817 for Medicare and 0.741 for commercial patients: a 0.1 standard-deviation quality improvement raises a hospital’s expected Medicare volume by roughly 8.2 percent and its commercial volume by roughly 7.4 percent.

Consistent with prior work on hospital demand (Capps et al., 2003; Ho, 2006; Gowrisankaran et al., 2015; Prager, 2020), distance strongly deters hospital choice. The baseline distance coefficient α_0^p is -0.103 (SE = 0.001) for Medicare and -0.208 (SE = 0.001) for commercial patients (Panel A), with commercial patients roughly twice as distance-sensitive. The demographic interactions in the Medicare sample reveal disparities in effective travel costs: non-white patients are more distance-sensitive (the positive coefficient on α_{white} implies that white patients face lower travel costs), and Medicaid-eligible patients—a proxy for low income—are significantly more distance-averse ($\alpha_{\text{low inc.}} = -0.012$, SE = 0.000). These patterns are consistent with transportation barriers faced by lower-income and minority populations.

6 Cost and Bargaining Estimation

The demand estimates from Section 5 characterize how patients value hospital quality and how patient demand responds to changes in quality. Two supply-side parameters remain to be estimated. The first, α_Y , governs how the cost of producing care rises with quality: since $\partial f / \partial y = \alpha_Y \cdot f$ (12), a 0.1 standard-deviation increase in quality raises per-discharge medical cost by $100 \times 0.1 \times \alpha_Y$ percent. The second, τ , measures the dollar value that insurers place on enrollee welfare when bargaining with hospitals. A larger τ means that hospitals which improve quality—and thereby raise the value of belonging to an insurer’s network—can negotiate higher commercial reimbursement, strengthening the return to choosing a higher y_{jt} .

I estimate these parameters using the first-order conditions from the bargaining problem (16) and from the hospital’s optimal quality choice (29). The price FOC determines the relationship between commercial prices, marginal costs, and bargaining markups for each hospital–insurer pair. The quality FOC determines how each hospital balances the marginal return and marginal cost of quality. I invert these two conditions to recover structural residuals, and estimate $\theta = (\alpha_Y, \tau, \{A_t\}_{t=1}^T)$ by GMM, where $\{A_t\}$ are market-time effective productivities that combine physical TFP and factor prices into a single cost-level shifter for each period (see Appendix C for full details).

6.1 Price Data

The estimation requires data on prices for both types of payers. I construct commercially negotiated prices from the MA-APCD as the ratio of total allowed amounts to total DRG weights for

each insurer–hospital–year pair, excluding claims paid under capitation or global payment arrangements.²³ I compute hospital Medicare reimbursement rates from the CMS Inpatient Prospective Payment System (IPPS) Final Rule Impact Files, which report a national standardized amount and hospital-specific geographic adjustment factors that reflect local wage indices. The Medicare price per DRG-weight for each hospital-year is the product of the standardized amount and the hospital’s geographic adjustment factor.

In the estimation sample, the average commercial price is \$11,430 per DRG-weight, roughly 1.65 times the average Medicare rate of \$6,930. Appendix Figure 20 plots the distributions of both prices. Commercial prices exhibit substantially more cross-hospital variation than Medicare rates (SD \$2,630 vs. \$240), reflecting the fact that commercial prices are negotiated bilaterally while Medicare rates are set administratively with limited geographic adjustment. Commercial prices are positively and significantly correlated with the quality index ($\text{cor} = 0.41$, $p < 0.001$), consistent with the finding in Cooper, Craig, Gaynor, and Van Reenen (2019) that higher-priced hospitals tend to deliver higher-quality care.

6.2 Estimation Strategy

I estimate the parameter vector $\theta = (\alpha_Y, \tau, \{A_t\}_{t=1}^T)$ by inverting the model’s two equilibrium conditions—the price FOC (16) and the quality first-order condition (29)—to obtain structural residuals, and then imposing that these residuals are orthogonal to a set of instruments. Recall from Section 4.3 that marginal cost decomposes into a medical component that depends on quality and a structural residual:

$$MC_{jt}^M = f(y_{jt}) + \nu_{jt}^M, \quad MC_{jgt}^C = f(y_{jt}) + \nu_{jgt}^C, \quad (35)$$

where $f(y_{jt}) = e^{\alpha_Y y_{jt}} / A_t$. The estimation strategy inverts the equilibrium conditions to express each structural residual as a function of observables and the parameter vector θ .

Commercial residual. Rearranging the price FOC (16) and substituting (35):

$$\nu_{jgt}^C(\theta) = P_{jgt}^C - \frac{\Delta V_{jg}(\tau)}{Q_{jgt}^C} - f(y_{jt}; \alpha_Y, A_t), \quad (36)$$

The observed commercial price equals the sum of the bargaining markup $\Delta V_{jg} / Q_{jgt}^C$, the medical cost f , and the structural residual ν_{jgt}^C , which captures administrative cost differences across hospital–insurer pairs. The insurer surplus ΔV_{jg} includes the term $\tau \cdot GW_{jg}$ (10), the dollar value of the welfare that insurer g ’s enrollees gain from having hospital j in the network. Because the demand specification includes no out-of-pocket price coefficient, τ is a composite parameter that both converts enrollee welfare from utils to dollars and captures the weight the insurer places on that welfare. The residual is defined only for in-network pairs ($j \in \mathcal{N}_{gt}$).

²³Under capitation or global payment contracts, insurers pay hospitals a fixed amount per enrollee or per episode rather than a per-service fee. The amounts recorded on these claims do not reflect negotiated prices for individual hospitalizations.

Medicare residual. Since equal bargaining weights imply that each side’s gains from agreement are equal (17), the quality first-order condition (29) can be written as

$$(P_{jt}^M - MC_{jt}^M) \frac{\partial Q_{jt}^M}{\partial y_{jt}} + Q_{jt}^M \frac{\partial P_{jt}^M}{\partial y_{jt}} + \sum_{g=1}^{G_t} \frac{d\Delta V_{jg}}{dy_{jt}} = \alpha_Y \cdot f(y_{jt}) \cdot Q_{jt}^M, \quad (37)$$

where the left-hand side is the marginal return to quality—the Medicare demand and price terms plus the total commercial surplus response across all insurers—and the right-hand side is the marginal cost of quality applied to Medicare discharges.²⁴ Substituting $MC_{jt}^M = f + \nu_{jt}^M$ and solving for the Medicare residual:

$$\nu_{jt}^M(\theta) = (P_{jt}^M - f) - \frac{\alpha_Y \cdot f \cdot Q_{jt}^M - Q_{jt}^M \frac{\partial P_{jt}^M}{\partial y_{jt}} - \sum_{g=1}^{G_t} \frac{d\Delta V_{jg}}{dy_{jt}}}{\frac{\partial Q_{jt}^M}{\partial y_{jt}}}. \quad (38)$$

The total derivative $d\Delta V_{jg}/dy_{jt}$ accounts for how other hospitals’ equilibrium prices respond to hospital j ’s quality change, and is computed by solving a linear system for each insurer (See Appendix E for full details). In estimation, $\partial P_{jt}^M/\partial y_{jt}$ follows the stylized HVB program (Section 4.5) with a 2 percent withhold.

GMM objective. At the true parameter values, the structural residuals are orthogonal to instruments $\mathbf{Z}_{jt}$:

$$E[\nu_{jgt}^C(\theta_0) \cdot \mathbf{Z}_{jt}] = 0, \quad E[\nu_{jt}^M(\theta_0) \cdot \mathbf{Z}_{jt}] = 0. \quad (39)$$

I stack the commercial moment conditions—one block per insurer, so that each insurer contributes its own set of orthogonality restrictions—together with the Medicare moment conditions into a single moment vector, and estimate θ by two-step efficient GMM.²⁵ The demand objects entering the estimation are computed from the second-stage estimates that instrument quality (Section 5.3). Because the APCD does not capture the full universe of commercial hospitalizations, I rescale commercial demand in each market-year so that the implied Medicare-to-commercial payer mix matches the aggregate shares observed in HCRIS (Appendix G.1).

6.3 Identification

Quality–cost elasticity (α_Y). The parameter α_Y governs how marginal cost rises with quality through the cost function $f(y_{jt}) = e^{\alpha_Y y_{jt}}/A_t$. It is identified from cross-hospital variation in quality within markets (years). Hospitals with higher quality have higher implied marginal costs, and this cost difference is reflected in both the implied commercial residual—through the level of f —and

²⁴The commercial cost-of-quality terms in (29) are absorbed into $d\Delta V_{jg}/dy_{jt}$ through the equal-split identity (17), which is why only Medicare volume Q_{jt}^M appears on the right-hand side.

²⁵The first step uses a block-diagonal $(\mathbf{Z}'\mathbf{Z}/N)^{-1}$ weighting matrix with one block per insurer and one block for Medicare. The second step uses the inverse of the heteroskedasticity-robust variance of the moment conditions evaluated at the first-step estimates.

the implied Medicare residual—through the cost-of-quality term $\alpha_Y f Q_{jt}^M$ in the quality FOC (37). Because hospitals choose quality after observing both ν_{jgt}^C and ν_{jt}^M , y_{jt} is potentially endogenous. I instrument using the predicted Medicare share from the aging instrument described in Section 3.2, which shifts the marginal return to quality through demand composition (Section 6.5) without entering the cost function.

Insurer welfare weight (τ). The parameter τ enters the bargaining markup through the welfare component of insurer surplus: ΔV_{jg} includes the term $\tau \cdot GW_{jg}$, the dollar value of enrollee welfare from having hospital j in insurer g 's network (10). Higher τ raises markups at hospitals that contribute more welfare per discharge, so τ is identified from the relationship between markups and network value added. This variation operates both across hospitals and across insurers at the same hospital. For instance, a hospital that is the only nearby in-network option for many of an insurer's enrollees would generate a large surplus from inclusion and negotiate a higher markup. The same hospital may be less valuable to a different insurer whose enrollees have closer alternatives. I construct two instruments that capture each hospital's value to each insurer's network: Total Value Added ($TV A_{jgt}$) and Average Willingness to Pay ($AWTP_{jgt}$).²⁶ $TV A_{jgt}$ measures the enrollee welfare gain from having hospital j in insurer g 's network, and $AWTP_{jgt}$ normalizes by DRG-weighted demand. I evaluate both instruments at mean quality rather than each hospital's observed quality, so that the instruments reflect patient geography and network structure—which are predetermined—rather than the hospital's endogenous quality choice.

Market-time productivities (A_t). Each A_t absorbs the market-time cost level in the cost function. I include market-time dummies in the instrument set, so that each A_t is identified from the requirement that both the Medicare and commercial residuals be mean-zero within each market-time period.

6.4 Estimates

Table 2: GMM Estimates of Cost and Bargaining Parameters

	Estimate
α_Y (quality–cost elasticity)	0.253*** (0.024)
τ (insurer welfare weight)	3.22*** (0.21)

Notes: Standard errors in parentheses. Two-step efficient GMM. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2 reports GMM estimates of the structural parameters. The quality–cost elasticity α_Y is estimated at 0.253 (SE = 0.024), implying that a 0.1 standard-deviation increase in hospital

²⁶ $TV A_{jgt} = -\sum_{i \in g} \log(1 - \bar{\sigma}_{ij})$ and $AWTP_{jgt} = TV A_{jgt} / \bar{Q}_{jgt}^C$, where $\bar{\sigma}_{ij}$ and $\bar{Q}_{jgt}^C$ are the choice probabilities and DRG-weighted demand evaluated at the sample mean quality level $\bar{y}_t$.

quality raises per-DRG-weight marginal cost by approximately 2.5%. The insurer welfare weight is estimated at $\tau = 3.22$ (SE = 0.21). The implied marginal cost averages \$6,930 per DRG-weight across all hospital-payer pairs (Appendix Figure 21, left panel). The parameter τ scales the enrollee welfare component of insurer surplus (10) and, through the bargaining markup (16), determines how much of patients' willingness to pay for quality is reflected in commercial prices. The positive and significant estimate confirms that insurers place substantial value on enrollee welfare when bargaining with hospitals, consistent with the prediction that equilibrium commercial prices rise with quality (Section 4.6).

6.5 Return to Quality by Payer

Once we have estimated all the structural parameters of the model, we can compare how the returns from quality from commercially insured patients compare to those from Medicare patients. Recall that the hospital's optimal quality choice satisfies

$$\begin{aligned}
 & \underbrace{(P_{jt}^M - MC_{jt}^M) \frac{\partial Q_{jt}^M}{\partial y_{jt}} + Q_{jt}^M \frac{\partial P_{jt}^M}{\partial y_{jt}}}_{\text{marginal return from Medicare}} + \underbrace{\sum_{g=1}^{G_t} (P_{jgt}^C - MC_{jgt}^C) \frac{\partial Q_{jgt}^C}{\partial y_{jt}} + \sum_{g=1}^{G_t} Q_{jgt}^C \frac{dP_{jgt}^{C*}}{dy_{jt}}}_{\text{marginal return from commercial}} \\
 & = \underbrace{\alpha_Y \cdot f \left(Q_{jt}^M + \sum_{g=1}^{G_t} Q_{jgt}^C \right)}_{\text{marginal cost}}.
 \end{aligned} \tag{40}$$

That is, the hospital chooses y_{jt} such that the marginal return it obtains from both Medicare and commercially insured patients equals its increase in the cost of treating patients. I evaluate each term of the marginal return to quality at the estimated parameters, averaging across all hospitals in 2019.

The demand estimation results from Section 5.3 show that Medicare and commercial demand respond similarly to an increase in quality. On average, a 0.1 standard-deviation increase in the quality index raises Medicare demand by 8.2 percent and commercial demand by 7.4 percent. In the single-hospital model of Garthwaite et al. (2022), a larger commercial share raises quality if commercial prices are higher than public prices and commercial demand is at least as responsive to quality as public demand. My demand estimates do not deliver the second condition. However, the margins implied by the model differ greatly across payers. Medicare margins are small: the mean Medicare margin is \$90 per DRG-weight, with a median of \$870 per DRG-weight. Moreover, about 34 percent of all hospital-years have negative Medicare margins. Commercial margins implied by the model are, in contrast, much larger. The average margin for commercially insured patients is \$4,370 per DRG-weight, with a median of \$4,800. Appendix Figure 21 (right panel) plots the distributions of implied margins by payer type. The ordering of these margins is consistent with the margins hospitals report in their cost reports (Section 2), where the distribution of Medicare margins is

concentrated around zero, with commercial margins being much higher. Therefore, although higher quality attracts Medicare and commercial patients at similar rates, each additional commercial patient contributes much more to the return to quality than an additional Medicare patient. On average, the margin earned on additional commercial demand, $\sum_{g=1}^{G_t} (P_{jgt}^C - MC_{jgt}^C) \partial Q_{jgt}^C / \partial y_{jt}$, accounts for 63 percent of the total marginal return to quality, while the margin earned on additional Medicare demand, $(P_{jt}^M - MC_{jt}^M) \partial Q_{jt}^M / \partial y_{jt}$, accounts for 17 percent.

Moreover, the estimates reveal that increasing quality substantially raises bargained commercial prices. Figure 7 plots the distribution of the derivative of commercial prices with respect to quality, dP_{jgt}^{C*} / dy_{jt} in equation (40), averaged across each hospital's insurers. The price response is positive for every hospital, ranging from \$410 to \$1,710 per DRG-weight. On average, a one standard-deviation increase in the quality index raises commercial prices by \$860 per DRG-weight. Medicare's quality adjustment, in contrast, is much smaller: the same increase in quality raises the Medicare rate by \$50 per DRG-weight on average. Therefore, quality raises commercial prices roughly 17 times more than it raises the Medicare rate. The increase in bargained commercial prices received on existing commercial demand, $\sum_{g=1}^{G_t} Q_{jgt}^C dP_{jgt}^{C*} / dy_{jt}$, accounts for 18 percent of the total marginal return to quality, while the increase in the Medicare rate from Medicare's quality adjustment, $Q_{jt}^M \partial P_{jt}^M / \partial y_{jt}$, accounts for 2 percent.

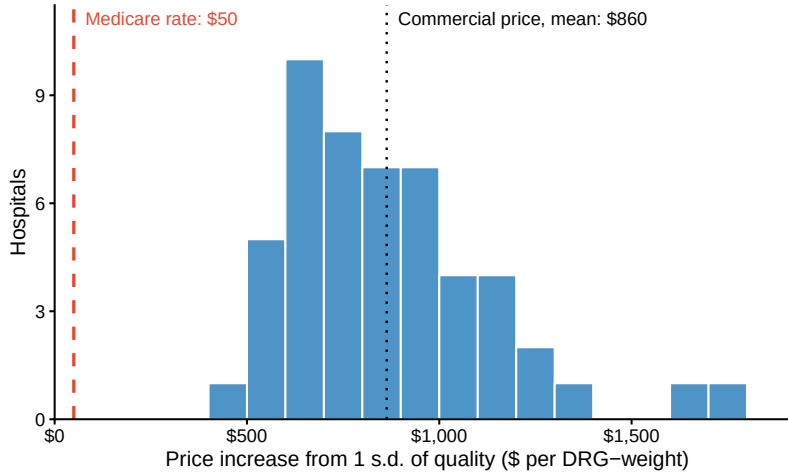


Figure 7: Distribution of dP_{jgt}^{C*} / dy_{jt} across hospitals in 2019. The dashed red line shows the average of $\partial P_{jt}^M / \partial y_{jt}$ across all hospitals.

In summary, even though demand from Medicare and commercial patients responds similarly to quality, implied commercial margins are much higher, and commercial prices increase substantially more with quality than the Medicare rate does. Therefore, on average, commercially insured patients account for 81 percent of the marginal return to quality, despite making up only 34 percent of patient demand.

7 Counterfactual Analysis

I use the estimated structural model from Sections 5–6 to conduct four counterfactual exercises. First, I estimate how hospital quality and patient welfare change under projected increases in the Medicare share of hospital demand due to population aging through 2040. Second, I estimate the increase in Medicare reimbursement rates required to offset these projected welfare losses. Third, I evaluate a policy experiment that progressively lowers the Medicare eligibility age from 65 to 50. Finally, I evaluate a budget-neutral alternative modeled on Medicare’s Hospital Value-Based Purchasing program, in which a share of Medicare payments is withheld and returned in proportion to hospital quality, and I solve for the share of Medicare payments that would have to be redistributed to prevent average patient welfare from falling.

I use the Massachusetts hospital market in 2019 as the baseline in all counterfactuals. I hold fixed the estimated demand, cost, and bargaining parameters, the structural residuals $(\xi_{jt}^p, \nu_{jgt}^C, \nu_{jt}^M)$, the set of hospitals in the market, insurers’ enrollee pools, hospital–insurer networks, and Medicare base rates. In all counterfactuals, I measure consumer welfare as average consumer surplus per patient, defined as τ times the McFadden surplus, the expected maximum utility a patient obtains from her choice set of hospitals (McFadden, 1981; Small and Rosen, 1981).²⁷ This represents the dollar value that patients derive from these hospitals. I report it separately for Medicare patients, commercial patients, and all patients.²⁸

7.1 Population Aging

In this counterfactual, I project how population aging will change the payer mix in Massachusetts through 2040, and how this will affect hospital quality and patient welfare. I project the payer mix in two steps. First, I predict the population of Massachusetts in 2040 for each age and gender group by aging the population forward using national survival rates. Second, for each age and gender group, I predict its expected DRG-weighted hospital stays using DRG-weighted hospitalization rates computed from APCD micro data (for patients under 65) or from MedPAR and MBSF (for patients 65 and older). Under this projection, the Medicare share rises from 65.3 percent in 2019 to 74.3 percent in 2040. Appendix H.1 describes the details of how I implement this projection.

To see how the payer mix affects the hospital’s optimal quality choice, note that we can rewrite the quality first-order condition (29) by dividing both sides by the hospital’s total DRG-weighted

²⁷For a patient i of payer type p , the McFadden surplus is $W_i = \log(\sum_{k \in C_i^p \cup \{0\}} \exp(V_{ik}^p))$, where V_{ik}^p is the deterministic component of utility in equation (5), and C_i^p is i ’s choice set.

²⁸The all-patient average always weights Medicare and commercial patients by their 2019 shares. Changes in the all-patient average therefore reflect changes in welfare per patient instead of changes in payer mix towards Medicare.

demand and rearranging:

$$\begin{aligned}
& \underbrace{\frac{Q_{jt}^M}{Q_{jt}^M + Q_{jt}^C}}_{\text{Medicare share}} \underbrace{\frac{1}{Q_{jt}^M} \left[(P_{jt}^M - MC_{jt}^M) \frac{\partial Q_{jt}^M}{\partial y_{jt}} + Q_{jt}^M \frac{\partial P_{jt}^M}{\partial y_{jt}} \right]}_{AMR_{jt}^M} \\
& + \underbrace{\frac{Q_{jt}^C}{Q_{jt}^M + Q_{jt}^C}}_{\text{commercial share}} \underbrace{\frac{1}{Q_{jt}^C} \sum_{g=1}^{G_t} \left[(P_{jgt}^C - MC_{jgt}^C) \frac{\partial Q_{jgt}^C}{\partial y_{jt}} + Q_{jgt}^C \frac{dP_{jgt}^{C*}}{dy_{jt}} \right]}_{AMR_{jt}^C} = \underbrace{\alpha_Y f(y_{jt})}_{\text{average marginal cost}}, \tag{41}
\end{aligned}$$

where $Q_{jt}^C = \sum_{g=1}^{G_t} Q_{jgt}^C$ is the hospital's total commercial demand, and AMR_{jt}^M (AMR_{jt}^C) is the total marginal revenue of quality from Medicare (commercial) patients, divided by total Medicare (commercial) demand at hospital j . Figure 8 plots AMR_{jt}^M , AMR_{jt}^C , their weighted average (the left-hand side of equation (41)), and the average marginal cost, as functions of the hospital's own quality y_{jt} , averaged across all hospitals in Massachusetts in 2019. The hospital's optimal quality y_{jt}^* is such that the average marginal return equals the average marginal cost. Because of the differences between commercial and Medicare margins and price responses documented in Section 6.5, at the 2019 equilibrium, the average marginal return from commercial patients for the average hospital was \$3,990 per DRG-weight, compared to \$130 for Medicare patients, and \$1,740 on average across all patients (Medicare and commercial).

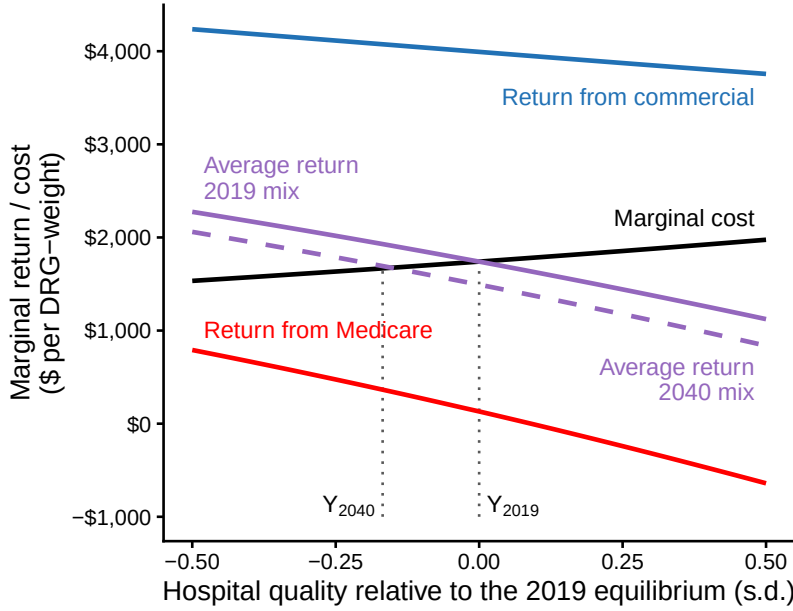


Figure 8: Marginal return and marginal cost of quality (per patient, in \$ per DRG-weight), averaged across all Massachusetts hospitals. The red and blue curves show the marginal returns from Medicare and commercial patients at the 2019 payer mix. The solid and dashed purple curves show the average marginal return at the 2019 and projected 2040 payer mixes, respectively. The black curve plots the average marginal cost per patient ($\alpha_Y f(y_{jt})$).

Note that, in the model, the marginal cost per DRG-weight does not depend on the number of patients treated. Therefore, holding the distribution of patient characteristics within each payer group fixed, a hospital’s optimal quality choice depends on the numbers of Medicare and commercial patients only through their ratio. To see this, consider doubling the numbers of Medicare and commercial patients. The hospital chooses the same quality because both sides of equation (29) double.

Because of this, to study how hospitals would readjust their quality in response to changes in payer mix through 2040, I reweight the Medicare terms on both sides of equation (29) by a factor $\lambda^M = 1.537$, since the projection implies 53.7 percent more Medicare patients per commercial patient in 2040 than in 2019. Dividing both sides by the rescaled total DRG-weighted demand gives

$$\frac{\lambda^M Q_{jt}^M}{\lambda^M Q_{jt}^M + Q_{jt}^C} AMR_{jt}^M + \frac{Q_{jt}^C}{\lambda^M Q_{jt}^M + Q_{jt}^C} AMR_{jt}^C = \alpha_Y f(y_{jt}). \quad (42)$$

I then re-solve equilibrium quality for all hospitals jointly. Appendix H.1 explains how I compute λ^M and presents the details of the implementation.

The dashed line in Figure 8 plots the average marginal return at the 2040 payer mix (the left-hand side of equation (42)). Because the share of Medicare patients relative to commercial patients increases, and $AMR_{jt}^M < AMR_{jt}^C$, the average marginal return is now lower at every quality level. Therefore, in the 2040 equilibrium, hospitals lower their quality by 0.17 standard deviations on average.

As a result, average consumer surplus per patient falls by \$118 for Medicare patients, \$200 for commercial patients, and \$152 across all patients. Commercial patients lose more welfare because quality declines are concentrated at hospitals with the highest commercial shares. Among the third of hospitals with the highest commercial shares in 2019, quality falls by 0.32 standard deviations on average, compared with 0.09 standard deviations among the third with the lowest commercial share. Welfare losses from quality declines are mitigated by patients switching to hospitals where quality falls less. If every patient kept her 2019 choice probabilities, consumer surplus would fall by \$187 per patient on average (\$145 for Medicare patients and \$245 for commercial patients), so patient reallocation offsets about 19 percent of the loss for both payers.²⁹

7.2 Medicare Rate Increase

The welfare losses documented in Section 7.1 arise because the marginal return to quality declines as the payer mix shifts toward Medicare (Section 6.5). A natural policy lever is Medicare reimbursement rates: raising rates directly increases the Medicare margin in the quality first-order condition (29), increasing the return to attracting an additional Medicare patient. In the nursing home industry, Hackmann (2019) evaluates a similar policy, finding that a 10 percent increase in the Medicaid reimbursement rate raises skilled-nurse staffing per resident by 8.7 percent. I ask: by

²⁹Holding patient i ’s 2019 choice probabilities σ_{ik}^p fixed, her change in McFadden surplus is $\sum_{k \in C_i^p} \sigma_{ik}^p \theta_i^p \Delta y_k$, where Δy_k is the change in hospital k ’s quality.

how much would Medicare rates need to increase to fully offset the projected welfare losses?

I parameterize a uniform rate increase by adding x dollars per DRG-weight to every hospital’s Medicare price:

$$P_j^M(x) = P_j^M + x. \quad (43)$$

The rate increase raises the Medicare volume term $(P_{jt}^M + x - MC_{jt}^M) \frac{\partial Q_{jt}^M}{\partial y_{jt}}$ in the hospital’s quality first-order condition (29), directly increasing the marginal return to quality. For the projected 2040 payer mix, I re-solve equilibrium quality for all hospitals jointly at each value of x and find the rate increase that restores average consumer surplus per patient across all patients to its 2019 level.

By 2040, restoring average consumer surplus per patient to its 2019 level requires a rate increase of approximately \$168 per DRG-weight, corresponding to a 2.4 percent increase over the average baseline Medicare rate. As a back-of-the-envelope calculation, applying this percentage increase to total Medicare inpatient spending in 2022, roughly \$210 billion per year including Medicare Advantage, implies an additional cost of approximately \$5.1 billion per year.³⁰

7.3 Medicare Eligibility Expansion

I next consider the effects of expanding Medicare by lowering the eligibility age. Proposals to expand Medicare coverage to near-elderly adults have received considerable attention in recent years and have broad support. In a 2026 poll of more than 19,000 voters, at least two-thirds of respondents from both parties favored lowering the eligibility age to 55 (DeGroot, 2026). House Democrats have also listed lowering the eligibility age among their priorities for the next Congress (Cost-of-Living Healthcare Working Group, 2026). I use the estimated model to study how this expansion would affect hospital quality and patient welfare.

I progressively lower the eligibility age from 65 to 50. At each eligibility age, all commercially insured patients between that age and 64 become Medicare patients. These “switchers” keep their preferences (α_i^C, θ_i^C) , their insurer’s network $\mathcal{N}_{gt}$, and their structural residuals $(\xi_{jt}^C, \nu_{jgt}^C)$, but hospitals are now paid the Medicare rate P_{jt}^M for them.

Lowering the eligibility age affects quality through the same channel as population aging in Section 7.1. Hospitals receive lower payments for the switchers, and Medicare rates respond less to quality than negotiated commercial prices, reducing the marginal return to quality and therefore lowering equilibrium quality.

Average hospital quality falls by 0.08 standard deviations when the eligibility age is lowered to 60 and by 0.19 when it is lowered to 50. Figure 9 plots the corresponding changes in average

³⁰Fee-for-service Medicare spent approximately \$111 billion on inpatient hospital services under the IPPS in fiscal year 2022 (Medicare Payment Advisory Commission, 2024a), and Medicare Advantage plans covered 48 percent of eligible beneficiaries (Ochieng, Biniek, Freed, Damico, and Neuman, 2022). Scaling the fee-for-service amount by the enrollment share, assuming comparable inpatient spending per enrollee, puts total Medicare inpatient spending at roughly \$210 billion. Medicare Advantage spending rises with fee-for-service rates as well: the plans pay hospitals rates benchmarked to fee-for-service rates (Berenson et al., 2015; Maeda and Nelson, 2018), and the payments they receive from Medicare are tied to average fee-for-service spending in their county (Medicare Payment Advisory Commission, 2024b).

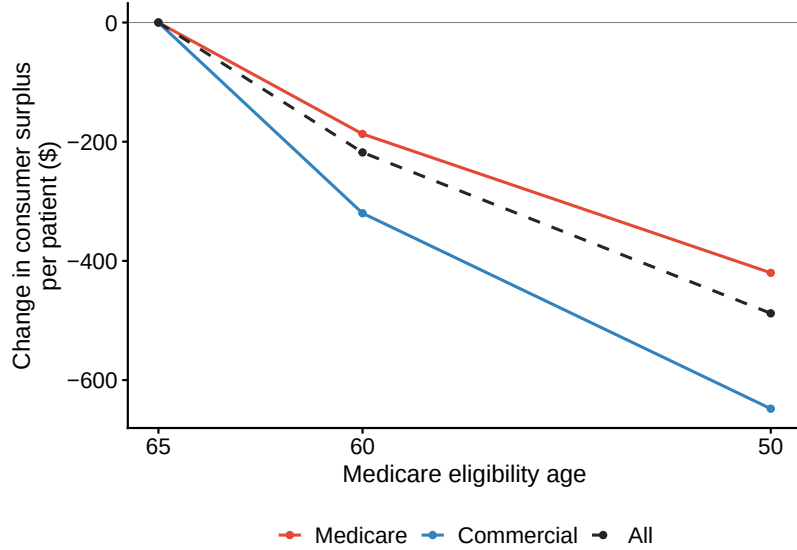


Figure 9: Change in average consumer surplus per patient as the Medicare eligibility age is lowered from 65 to 50, relative to eligibility at 65. The red and blue lines show changes for patients aged 65 and older and patients who remain commercially insured, respectively. The dashed black line shows the average across all patients, including switchers.

consumer surplus per patient. With eligibility at 60, average consumer surplus falls by \$187 for Medicare patients, \$320 for remaining commercial patients, and \$218 across all patients. With eligibility at 50, the corresponding losses are \$420, \$648, and \$488 per patient, respectively.

My model captures only one margin of Medicare expansion: how hospitals adjust their quality when they are paid the administered Medicare rate for a larger share of their patients. Therefore, the welfare changes in this counterfactual measure only the change in consumer surplus from changes in hospital quality. However, expanding Medicare could have multiple effects that are unaccounted for in this model. For instance, expanding Medicare could lower commercial premiums by lowering insurers' costs per enrollee. Medicare expansion would leave commercial insurers with a younger, and likely healthier, pool of enrollees. Moreover, because negotiated prices rise with hospital quality, lower quality would also lower the prices insurers pay hospitals. Insurers could pass part of these savings through to enrollees as lower premiums, or, for employer-sponsored plans, to employers, who could in turn pass them through to workers as higher wages ([Brot-Goldberg, Cooper, Craig, Klarnet, Lurie, and Miller, 2024](#)). These savings would generate welfare gains for patients who remain commercially insured that my measure of consumer surplus does not capture, since I do not model premiums. Changes in premiums and in the quality of hospitals in each insurer's network could also lead patients to change insurers, which my model rules out by holding enrollee pools fixed. Capturing these effects would require adding a stage in which insurers set premiums and consumers choose insurers, as in [Ho and Lee \(2017\)](#).

The counterfactual also does not take into account the effects of expanding coverage to uninsured and underinsured patients. In my model, I take the total number of hospitalizations as given:

patients choose which hospital to visit, but do not decide whether to seek hospital care. In reality, uninsured and underinsured adults may forgo or delay hospital care because of lack of adequate coverage (Card, Dobkin, and Maestas, 2008). Expanding Medicare could therefore increase hospital use, generating welfare gains from receiving care that these patients would otherwise forgo. Hospitals would also receive Medicare payments for care they currently provide to uninsured patients with little or no compensation (Garthwaite, Gross, and Notowidigdo, 2018). This would increase the marginal return to quality from these patients and could partly offset the quality decline caused by moving commercially insured patients to Medicare. In Massachusetts, however, the gains from covering the uninsured are likely small, since only 1.7 percent of residents were uninsured in 2023 (Center for Health Information and Analysis, 2024).

7.4 Making Medicare Rates More Responsive to Quality

The rate increase of Section 7.2 restores average consumer surplus, but at a cost of roughly \$5.1 billion per year in additional Medicare spending. An alternative is to make Medicare rates more responsive to quality by raising the withhold ω_t in the stylized HVBP program of Section 4.5. In the baseline, the program withholds 2 percent of each hospital’s base payments, as HVBP does today. I ask how large the withhold must be to restore average consumer surplus per patient to its 2019 level at the projected 2040 payer mix, while keeping the program budget neutral.

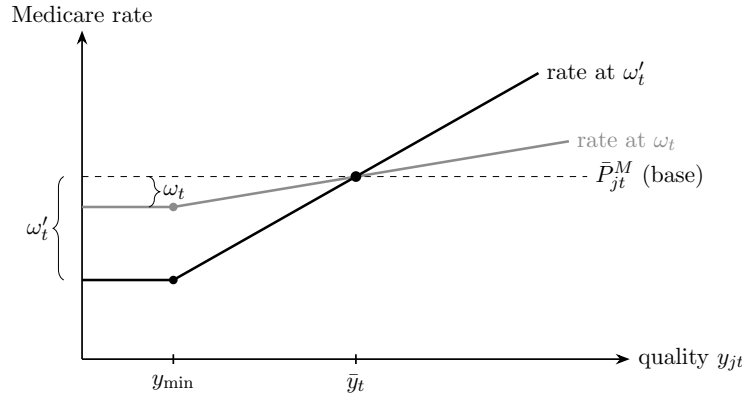


Figure 10: The realized Medicare rate under two withholds $\omega'_t > \omega_t$. The larger withhold lowers the payment to the lowest-quality hospital and finances a steeper slope, and both rates equal the base rate at the mean.

A larger withhold changes the hospital’s quality first-order condition (29) in two ways. First, it increases $\partial P_{jt}^M / \partial y_{jt} = b_t \bar{P}_{jt}^M$: as Figure 10 shows, withholding a larger share of base payments, $\omega'_t > \omega_t$, makes the Medicare rate a steeper function of quality. Second, it changes the margin the hospital earns on each additional Medicare patient it attracts, which enters the Medicare demand term $(P_{jt}^M - MC_{jt}^M) \partial Q_{jt}^M / \partial y_{jt}$. A hospital with quality above (below) the mean now receives a higher (lower) Medicare rate (26), which increases (decreases) its return to attracting an additional

Medicare patient.

I solve for the withhold ω_t^* that restores average consumer surplus per patient at the projected 2040 payer mix to its 2019 level. Figure 11 traces average consumer surplus per patient as a function of the withhold. This requires a withhold of 8.7 percent, compared with the current 2 percent withhold under HVBP. At ω_t^* , the implied payment per unit of quality averages \$206 per DRG-weight per standard deviation of the quality index, or 3.0 percent of a hospital’s base rate.

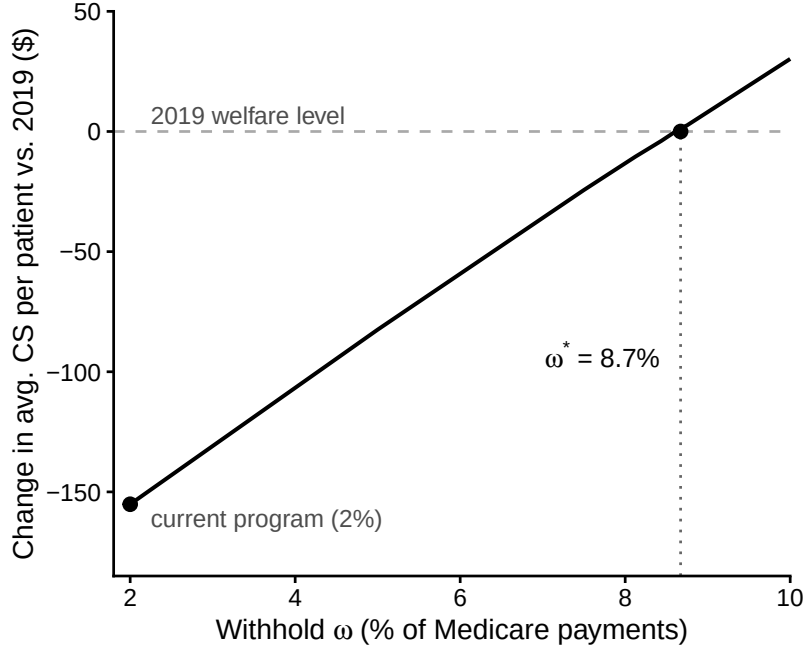


Figure 11: Change in average consumer surplus per patient at the 2040 payer mix relative to 2019, as a function of the withhold ω , starting from the current withhold of 2%. The dotted line marks the withhold ω^* that restores consumer surplus to its 2019 level. Market structure is held fixed at 2019 values.

Restoring consumer surplus by raising the withhold requires no new Medicare spending, but it redistributes the payments hospitals already receive. Because the highest-quality hospitals are concentrated in greater Boston, this redistribution may accentuate regional disparities in hospital finances. The left panel of Figure 12 maps the change in each hospital’s realized Medicare rate at the withhold $\omega^* = 8.7$ percent and the 2040 payer mix. Of the market’s 51 hospitals, only 12 see their Medicare rate rise relative to the current program. Ten of these are in the Boston metropolitan area. Massachusetts General and Brigham and Women’s, the two flagship teaching hospitals of Harvard Medical School, together receive 38 percent of the redistributed pool. The other 39 hospitals see their rate fall, by up to 6.8 percent, including 17 of the 19 hospitals outside the Boston metropolitan area.

An increase in the withhold rate moves Medicare payments toward the hospitals that already earn the highest margins (Figure 13, left panel).³¹

³¹For each hospital, the average margin is the volume-weighted average of its 2019 Medicare margin, $P_j^M - MC_j^M$,

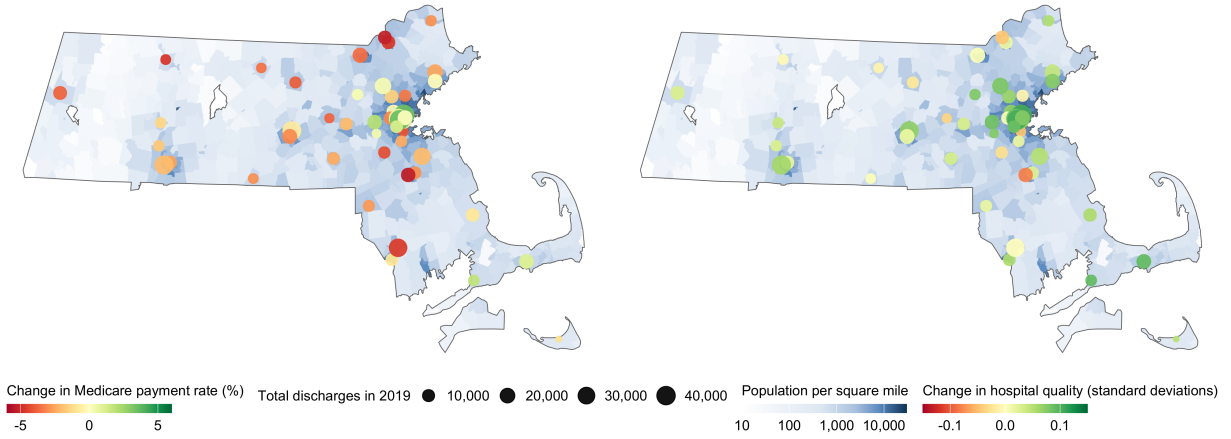


Figure 12: Geography of raising the withhold to $\omega^* = 8.7$ percent, at the 2040 payer mix. The left panel maps the change in each hospital's realized Medicare payment rate, in percent; the right panel maps the change in equilibrium hospital quality, in standard deviations of the quality index. Both changes are relative to the equilibrium under the current HVBP program. Each dot is a hospital, placed at the centroid of its zip code, with size proportional to its total discharges in 2019 from HCRIS. One hospital whose quality is held at its observed level in the solve is not shown. Background shading is 2020 Census population density by zip code area.

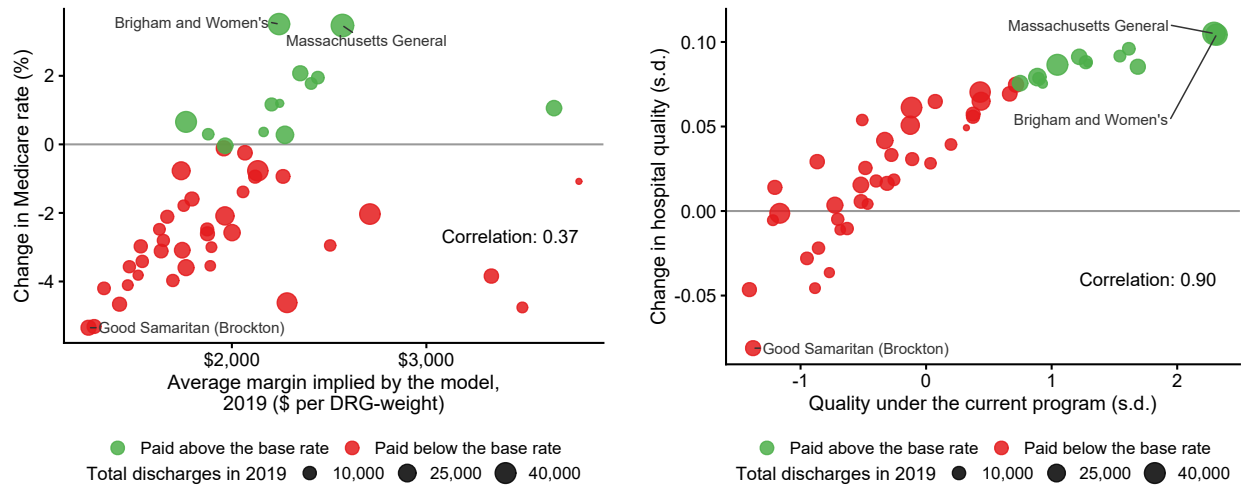


Figure 13: Effects across hospitals of raising the withhold to $\omega^* = 8.7$ percent, at the 2040 payer mix. Left panel: percent change in each hospital's Medicare rate, against the average margin implied by the model in 2019, in dollars per DRG-weight. Right panel: change in equilibrium hospital quality against quality before the change, both in standard deviations of the quality index. Both changes are relative to the equilibrium under the current HVBP program. Green (red) dots are hospitals paid above (below) their base rate under the program. Dot size is total discharges in 2019 from HCRIS. One hospital whose quality is held at its observed level in the solve is not shown.

Average consumer surplus returns to its 2019 level. However, the quality changes that produce it are uneven across hospitals. The right panel of Figure 13 plots each hospital’s quality response against its quality under the current program. Hospitals above the revenue-weighted mean quality, which the program pays above their base rate, raise quality by 0.09 standard deviations on average. Hospitals below the mean barely change quality on average, and the largest declines occur where quality was already lowest. The program therefore widens the distribution of quality across hospitals, whose standard deviation rises from 0.95 to 0.99. The geography of the quality changes mirrors the geography of the payments (Figure 12, right panel). The three largest quality increases all occur at hospitals in the Boston metropolitan area. The largest decline is at Good Samaritan in Brockton, which in 2020 became the first majority-Black city in Massachusetts.

Raising the withhold thus restores average consumer surplus at no fiscal cost, but these results show that it involves a trade-off, with distributional consequences for hospitals and patients alike. Medicare payments flow from low-margin hospitals toward the hospitals with the highest margins. Quality rises for patients treated where it was already highest and falls for patients treated where it was already lowest.

8 Conclusion

In the US, hospitals typically do not post prices and are instead reimbursed by insurers, the largest of which are commercial insurers and Medicare. Hospitals’ quality choices therefore depend on the degree to which commercial insurers and Medicare reward quality. In this paper, I study how Medicare payment policy shapes hospital quality and patient welfare. I build a structural model of hospital quality choice and hospital–insurer bargaining. In the model, quality increases demand from Medicare and commercial patients, affects negotiated commercial prices and Medicare rates, and raises hospital costs. I estimate the model using Massachusetts claims for both Medicare and commercially insured patients.

I find that—because commercial margins are higher, and commercial prices respond much more to quality than the Medicare rate does—at the average hospital, commercial patients account for 81 percent of the marginal return to quality. As a result, increases in the share of Medicare patients lead to lower hospital quality and patient welfare. By 2040, population aging lowers average consumer surplus by \$152 per patient, and lowering the eligibility age to 60 would lower it by \$218.

My analysis also highlights that Medicare could offset this decline by changing how it pays hospitals. One option is to raise Medicare rates, which increases hospitals’ incentives to provide quality, but comes at a significant fiscal cost. Another option is to make Medicare rates more responsive to quality, which requires no additional spending, but concentrates payments and quality gains at the highest-quality, highest-margin hospitals.

More broadly, these results imply that Medicare policies, such as eligibility expansions and payment reforms, should be weighed not only by their fiscal cost and the coverage they provide, and its commercial margins, $P_{jg}^C - MC_{jg}^C$, with weights equal to the shares of the hospital’s DRG-weighted patient volume accounted for by each payer. The margin is expressed in 2019 dollars per DRG-weight.

but also by their effects on the quality of care all patients receive. This logic is not specific to Medicare. Nursing homes serve Medicaid-funded and private-pay residents (Hackmann, 2019), and public universities teach in-state students at regulated tuition alongside out-of-state students who pay market rates. Wherever a firm serves a regulated and an unregulated segment and quality is largely shared across them, policies aimed at the regulated segment shape the quality that all of its consumers receive.

The analysis in this paper does not incorporate how Medicare expansion would affect insurers' premiums and enrollees' choice of insurer. Medicare expansion could lower premiums, for instance, by leaving commercial insurers with a younger, and likely healthier, pool of enrollees. Future research could capture these effects by adding a stage in which insurers set premiums and consumers choose insurers, as in Ho and Lee (2017).

References

- Linda H. Aiken, Sean P. Clarke, Douglas M. Sloane, Julie Sochalski, and Jeffrey H. Silber. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*, 288(16): 1987–1993, 2002. doi: 10.1001/jama.288.16.1987.
- David J. Balan and Jeremy A. Sandford. Mergers when price and quality are set by bargaining. Working paper, March 2026.
- Robert A. Berenson, Jonathan H. Sunshine, David Helms, and Emily Lawton. Why Medicare Advantage plans pay hospitals traditional Medicare prices. *Health Affairs*, 34(8):1289–1295, 2015. doi: 10.1377/hlthaff.2014.1427.
- Zarek Brot-Goldberg, Zack Cooper, Stuart V. Craig, Lev R. Klarnet, Ithai Lurie, and Corbin L. Miller. Who pays for rising health care prices? Evidence from hospital mergers. Working Paper 32613, National Bureau of Economic Research, 2024.
- Cory Capps, David Dranove, and Mark Satterthwaite. Competition and market power in option demand markets. *RAND Journal of Economics*, 34(4):737–763, 2003. doi: 10.2307/1593786.
- David Card, Carlos Dobkin, and Nicole Maestas. The impact of nearly universal insurance coverage on health care utilization: Evidence from Medicare. *American Economic Review*, 98(5):2242–2258, 2008. doi: 10.1257/aer.98.5.2242.
- Center for Health Information and Analysis. Findings from the 2023 Massachusetts health insurance survey: Health insurance coverage and uninsurance among residents and their families. Technical report, Center for Health Information and Analysis, June 2024. URL <https://chiamass.gov/assets/docs/r/survey/mhis-2023/MHIS-2023-02-Health-Insurance-Coverage-and-Uninsurance.pdf>.

- Center for Health Information and Analysis. Massachusetts All-Payer Claims Database. Boston, MA: Center for Health Information and Analysis, 2025. Medical Claims, Member Eligibility, and Product files, 2018–2022.
- Centers for Medicare & Medicaid Services. CMS Hospital Value-Based Purchasing Program results for fiscal year 2019. Fact sheet, December 2018. URL <https://www.cms.gov/newsroom/fact-sheets/cms-hospital-value-based-purchasing-program-results-fiscal-year-2019>.
- Centers for Medicare & Medicaid Services. CMS fast facts. Technical report, Centers for Medicare & Medicaid Services, April 2025. URL https://data.cms.gov/sites/default/files/2025-04/CMSFastFacts2025_508.pdf.
- Allan Collard-Wexler, Gautam Gowrisankaran, and Robin S. Lee. “Nash-in-Nash” bargaining: A microfoundation for applied work. *Journal of Political Economy*, 127(1):163–195, 2019. doi: 10.1086/700729.
- Zack Cooper, Stuart V. Craig, Martin Gaynor, and John Van Reenen. The price ain’t right? Hospital prices and health spending on the privately insured. *Quarterly Journal of Economics*, 134(1):51–107, 2019. doi: 10.1093/qje/qjy020.
- Zack Cooper, Joseph J. Doyle, John A. Graves, and Jonathan Gruber. Do higher-priced hospitals deliver higher-quality care? Working Paper 29809, National Bureau of Economic Research, 2022.
- Cost-of-Living Healthcare Working Group. Memo on Democratic healthcare action for the first 100 hours, first 100 days, and first year of the 120th Congress, September 2026. URL <https://www.politico.com/f/?id=000001a0-da3a-d276-aff6-fb7f2c380000>. Memorandum to House Democratic Leader Hakeem Jeffries, September 1, 2026.
- David M. Cutler. The incidence of adverse medical outcomes under prospective payment. *Econometrica*, 63(1):29–50, 1995. doi: 10.2307/2951696.
- Leemore Dafny, Kate Ho, and Robin S. Lee. The price effects of cross-market mergers: Theory and evidence from the hospital industry. *RAND Journal of Economics*, 50(2):286–325, 2019. doi: 10.1111/1756-2171.12270.
- Leemore S. Dafny. Estimation and identification of merger effects: An application to hospital mergers. *Journal of Law and Economics*, 52(3):523–550, 2009. doi: 10.1086/605702.
- Lia DeGroot. Proposals to lower health costs get broad support in battlegrounds, July 2026. URL <https://rollcall.com/2026/07/30/proposals-to-lower-health-costs-get-broad-support-in-battlegrounds/>. CQ Roll Call, July 30, 2026.

- Joseph J. Doyle, John A. Graves, Jonathan Gruber, and Samuel A. Kleiner. Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, 123(1):170–214, 2015. doi: 10.1086/677756.
- Joseph J. Doyle, John A. Graves, and Jonathan Gruber. Evaluating measures of hospital quality: Evidence from ambulance referral patterns. *Review of Economics and Statistics*, 101(5):841–852, 2019. doi: 10.1162/rest_a.00804.
- Erin Duffy, Bich Ly, and Erin Trish. Trends in patient cost sharing for hospital care and implications for urban and rural hospital revenue. *Health Affairs*, 45(4):372–377, 2026. doi: 10.1377/hlthaff.2025.01340.
- Marc N. Elliott, Alan M. Zaslavsky, Elizabeth Goldstein, William Lehrman, Katrin Hambarsoomians, Megan K. Beckett, and Laura Giordano. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Services Research*, 44(2p1):501–518, 2009. doi: 10.1111/j.1475-6773.2008.00914.x.
- Ashvin Gandhi, David Grabowski, Brian McGarry, and Ljubica Ristovska. Staff composition and patient outcomes at nursing homes. Working paper, 2025.
- Craig Garthwaite, Tal Gross, and Matthew J. Notowidigdo. Hospitals as insurers of last resort. *American Economic Journal: Applied Economics*, 10(1):1–39, 2018. doi: 10.1257/app.20150581.
- Craig Garthwaite, Christopher Ody, and Amanda Starc. Endogenous quality investments in the U.S. hospital market. *Journal of Health Economics*, 84:102636, 2022. doi: 10.1016/j.jhealeco.2022.102636.
- Nina Glasgow and David L. Brown. Rural ageing in the United States: Trends and contexts. *Journal of Rural Studies*, 28(4):422–431, 2012. doi: 10.1016/j.jrurstud.2012.01.002.
- Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, 2020. doi: 10.1257/aer.20181047.
- Gautam Gowrisankaran, Aviv Nevo, and Robert Town. Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, 105(1):172–203, 2015. doi: 10.1257/aer.20130223.
- Paul L. E. Grieco and Ryan C. McDevitt. Productivity and quality in health care: Evidence from the dialysis industry. *Review of Economic Studies*, 84(3):1071–1105, 2017. doi: 10.1093/restud/rdw042.
- Tal Gross, Adam Sacarny, Maggie Shi, and David Silver. Regulated revenues and hospital behavior: Evidence from a Medicare overhaul. *Review of Economics and Statistics*, 106(6):1709–1718, 2024. doi: 10.1162/rest_a.01191.

- Atul Gupta. Impacts of performance pay for hospitals: The readmissions reduction program. *American Economic Review*, 111(4):1241–1283, 2021. doi: 10.1257/aer.20171825.
- Martin B. Hackmann. Incentivizing better quality of care: The role of Medicaid and competition in the nursing home industry. *American Economic Review*, 109(5):1684–1716, 2019. doi: 10.1257/aer.20151057.
- Kate Ho and Robin S. Lee. Insurer competition in health care markets. *Econometrica*, 85(2): 379–417, 2017. doi: 10.3982/ECTA13570.
- Kate Ho and Robin S. Lee. Equilibrium provider networks: Bargaining and exclusion in health care markets. *American Economic Review*, 109(2):473–522, 2019. doi: 10.1257/aer.20171288.
- Katherine Ho. The welfare effects of restricted hospital choice in the US medical care market. *Journal of Applied Econometrics*, 21(7):1039–1079, 2006. doi: 10.1002/jae.896.
- Harlan M. Krumholz, Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman, and Sharon-Lise T. Normand. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*, 113(13):1683–1692, 2006. doi: 10.1161/CIRCULATIONAHA.105.611186.
- Lan Liang, Brian Moore, and Anita Soni. National inpatient hospital costs: The most expensive conditions by payer, 2017. HCUP Statistical Brief 261, Agency for Healthcare Research and Quality, July 2020. URL <https://hcup-us.ahrq.gov/reports/statbriefs/sb261-Most-Expensive-Hospital-Conditions-2017.pdf>.
- Jared Lane K. Maeda and Lyle Nelson. How do the hospital prices paid by Medicare Advantage plans and commercial plans compare with Medicare fee-for-service prices? *Inquiry*, 55: 0046958018779654, 2018. doi: 10.1177/0046958018779654.
- Nicole Maestas, Kathleen J. Mullen, and David Powell. The effect of population aging on economic growth, the labor force, and productivity. *American Economic Journal: Macroeconomics*, 15(2): 306–332, 2023. doi: 10.1257/mac.20190196.
- Daniel McFadden. Econometric models of probabilistic choice. In Charles F. Manski and Daniel McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. MIT Press, Cambridge, MA, 1981.
- Medicare Payment Advisory Commission. Hospital acute inpatient services payment system. Payment basics, MedPAC, October 2024a. URL https://www.medpac.gov/wp-content/uploads/2024/10/MedPAC_Payment_Basics_24_hospital_FINAL_SEC.pdf.
- Medicare Payment Advisory Commission. Medicare advantage program payment system. Payment basics, MedPAC, October 2024b. URL https://www.medpac.gov/wp-content/uploads/2024/10/MedPAC_Payment_Basics_24_MA_FINAL_SEC.pdf.

- National Academy for State Health Policy. NASHP hospital cost tool. <https://tool.nashp.org/>, 2024. Accessed: 2024-07-02.
- Jack Needleman, Peter Buerhaus, Soeren Mattke, Maureen Stewart, and Katya Zelevinsky. Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine*, 346(22):1715–1722, 2002. doi: 10.1056/NEJMsa012247.
- Nancy Ochieng, Jeannie Fuglesten Biniek, Meredith Freed, Anthony Damico, and Tricia Neuman. Medicare advantage in 2022: Enrollment update and key trends. Issue brief, Kaiser Family Foundation, August 2022. URL <https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2022-enrollment-update-and-key-trends/>.
- Elena Prager. Health care demand under simple prices: Evidence from tiered hospital networks. *American Economic Journal: Applied Economics*, 12(4):196–223, 2020. doi: 10.1257/app.20180422.
- RAND Corporation. RAND hospital data. Santa Monica, CA: RAND Corporation, 2025. Tool TL303.
- Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109, 1982. doi: 10.2307/1912531.
- Bradley Setzler. Labor and product market power, endogenous quality, and the consolidation of the U.S. hospital industry. Working Paper 34180, National Bureau of Economic Research, 2025.
- Kenneth A. Small and Harvey S. Rosen. Applied welfare economics with discrete choice models. *Econometrica*, 49(1):105–130, 1981.
- U.S. Census Bureau. 2023 national population projections tables: Main series. Technical report, U.S. Census Bureau, November 2023. URL <https://www.census.gov/data/tables/2023/demo/popproj/2023-summary-tables.html>.
- Chapin White and Christopher M. Whaley. Prices paid to hospitals by private health plans: Findings from round 5 of an employer-led transparency initiative. Technical Report RR-A1144-2-v2, RAND Corporation, 2024.

A Reduced-Form Appendix

A.1 Commercial-to-Medicare Price Ratio

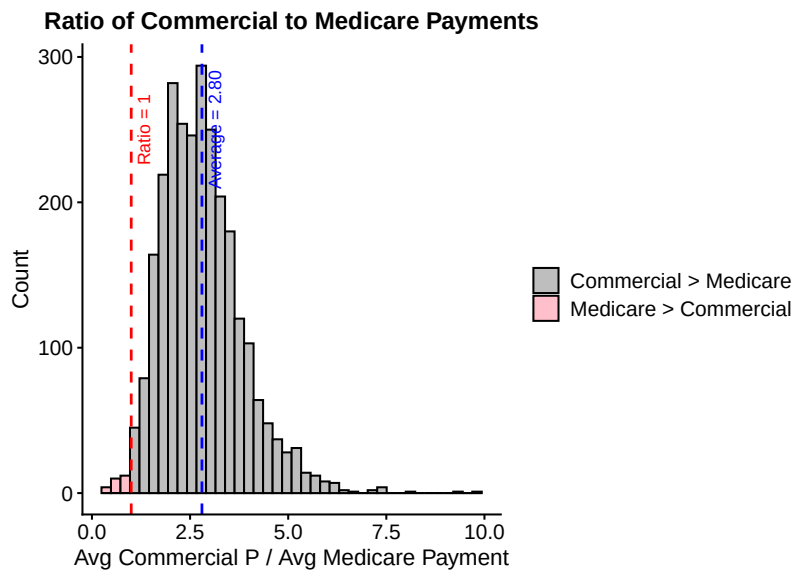


Figure 14: Distribution of the ratio of average commercially negotiated prices to average Medicare reimbursement rates across hospitals. Data are from the Healthcare Cost Report Information System (HCRIS) for general acute-care hospitals. The red dashed line marks a ratio of one (equal payment); the blue dashed line marks the sample average of 2.80. Nearly all hospitals receive higher commercial than Medicare reimbursement.

A.2 First Stage

Table 3 reports the first-stage regression of hospital Medicare share on the predetermined aging instrument.

Table 3: First-Stage Regression	
Dep. var: Medicare share of inpatient days	
Predicted % > 65	0.751*** (0.107)
Controls	Log pop., unemp. rate, log income p.c.
Fixed effects	Hospital, year
Clustering	Hospital, county
<i>F</i> -statistic (instrument)	49.4
Observations	45,225

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Regression Tables

Tables 4–8 report the full IV regression estimates underlying Figures 2–4 in the main text.

Table 4: Effect of Medicare Share on Expenditures and Staffing (per Discharge)

	Medicare Exp. (1)	Non-Med. Exp. (2)	Medical FTEs (3)	All FTEs (4)
Med. percent	−0.024*** (0.006)	−0.034*** (0.008)	−0.050*** (0.010)	−0.048*** (0.010)
Hospital & Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	45,201	35,083	32,835	32,878

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: log population, unemployment rate, log income per capita. All specifications are IV using the predetermined component of population aging as an instrument for Medicare share.

Table 5: Effect of Medicare Share on Financial Outcomes

	Non-Mdcr. Revenue (1)	Medicare Margin (2)
Med. percent	−0.038*** (0.011)	0.019*** (0.007)
Hospital & Year FE	Yes	Yes
Controls	Yes	Yes
Observations	42,994	29,027

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: log population, unemployment rate, log income per capita. All specifications are IV using the predetermined component of population aging as an instrument for Medicare share.

Table 6: Effect of Medicare Share on Quality Composites

	Mort. & Readm. (1)	ED Wait Times (2)	SoC Composite (3)	HCAHPS Composite (4)
Med. percent	-0.019 (0.018)	0.040*** (0.015)	0.110** (0.050)	-0.057*** (0.014)
Hospital & Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	43,800	35,415	15,265	38,958

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: log population, unemployment rate, log income per capita. All specifications are IV using the predetermined component of population aging as an instrument for Medicare share.

Table 7: Effect of Medicare Share on Patient Satisfaction (HCAHPS)

<i>Panel A</i>			
	Overall Rating (1)	Would Recommend (2)	Help Received (3)
Med. percent	-0.0054*** (0.0019)	-0.0084*** (0.0020)	-0.0061*** (0.0018)
Hospital & Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	38,990	38,990	38,983
<i>Panel B</i>			
	Cleanliness (4)	Info. on Medicines (5)	Rooms Kept Quiet (6)
Med. percent	-0.0032** (0.0016)	-0.0059*** (0.0017)	-0.0058** (0.0023)
Hospital & Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	38,990	38,959	38,988

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: log population, unemployment rate, log income per capita. All specifications are IV using the predetermined component of population aging as an instrument for Medicare share.

Table 8: Effect of Medicare Share on Aggregate Quality Indexes

	Quality Index (Z-Score) (1)	Quality Index (PCA) (2)
Med. percent	-0.033** (0.016)	-0.044** (0.021)
Hospital & Year FE	Yes	Yes
Controls	Yes	Yes
Observations	32,883	32,883

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: log population, unemployment rate, log income per capita. All specifications are IV using the predetermined component of population aging as an instrument for Medicare share.

A.4 Standard of Care Indicators

Figure 15 presents IV estimates for the individual components of the standard-of-care composite: central line infections (CLABSI), catheter-associated urinary tract infections (CAUTI), hospital-acquired pressure ulcers, and in-hospital falls. Table 9 reports the corresponding regression estimates.

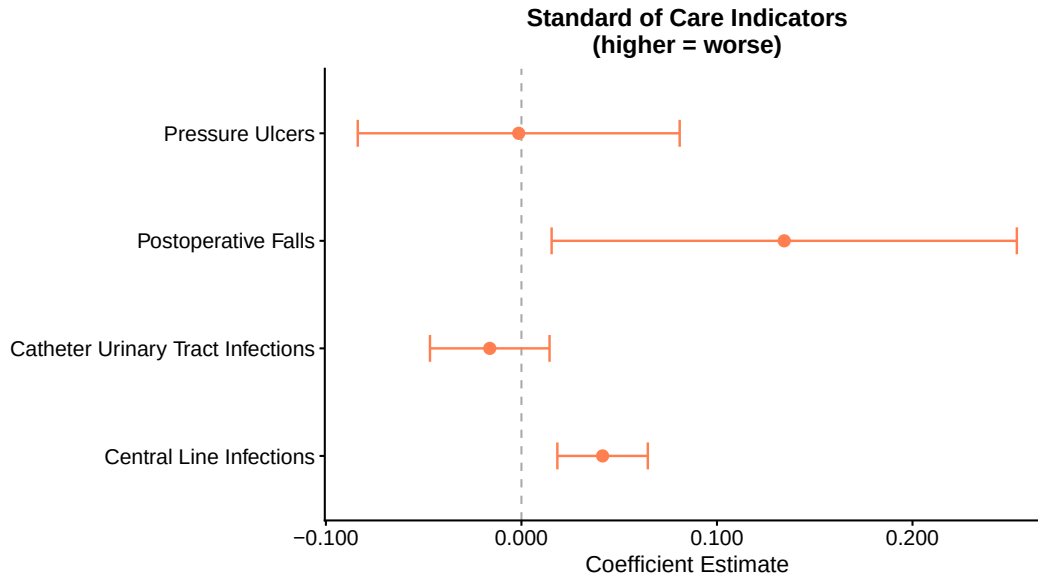


Figure 15: Effect of Medicare share on standard of care indicators. Each point is an IV estimate from a separate regression. Outcomes are standardized to zero mean and unit variance within year; coefficients are in standard-deviation units. All specifications include hospital and year fixed effects and county-level economic controls. Bars show 95% confidence intervals based on standard errors clustered at the hospital and county levels.

Table 9: Effect of Medicare Share on Standard of Care Indicators

	Pressure Ulcers (1)	Postop. Falls (2)	CAUTI (3)	CLABSI (4)
Med. percent	−0.001 (0.042)	0.134** (0.061)	−0.016 (0.016)	0.042*** (0.012)
Hospital & Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,361	15,355	39,679	39,679

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: log population, unemployment rate, log income per capita. All specifications are IV using the predetermined component of population aging as an instrument for Medicare share.

A.5 Additional Robustness Checks

A.5.1 Volume and Case-Mix Controls

Table 10 re-estimates the expenditure and staffing regressions of Section 3.3 adding, to the baseline controls, total inpatient discharges (Panel B) and the Medicare case-mix index, the average DRG weight of the hospital’s Medicare discharges (Panel C). The estimates are essentially unchanged relative to the baseline. Figure 16 plots them.

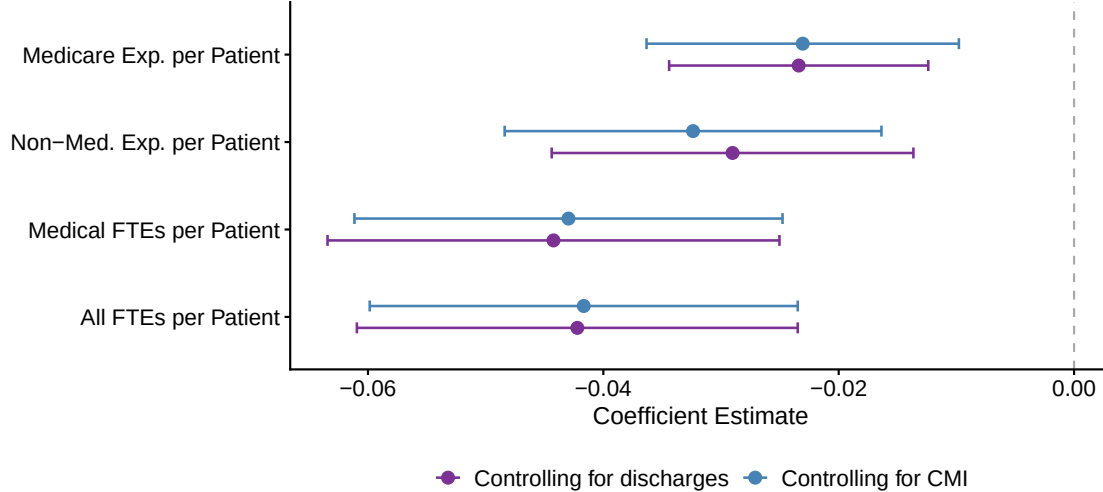


Figure 16: Effect of Medicare share on expenditures and staffing under volume and case-mix controls. Each point is an IV estimate from a separate regression; bars show 95% confidence intervals. All specifications include hospital and year fixed effects and the county-level economic controls. Table 10 reports the corresponding estimates.

Table 10: Effect of Medicare Share on Expenditures and Staffing: Volume and Case-Mix Controls

	Medicare Exp. (1)	Non-Med. Exp. (2)	Medical FTEs (3)	All FTEs (4)
<i>Panel A: Baseline</i>				
Med. percent	−0.024*** (0.006)	−0.034*** (0.008)	−0.050*** (0.010)	−0.048*** (0.010)
Observations	45,201	35,083	32,835	32,878
<i>Panel B: Controlling for total discharges</i>				
Med. percent	−0.023*** (0.006)	−0.029*** (0.008)	−0.044*** (0.010)	−0.042*** (0.010)
Observations	45,201	35,083	32,835	32,878
<i>Panel C: Controlling for Medicare case-mix index</i>				
Med. percent	−0.023*** (0.007)	−0.032*** (0.008)	−0.043*** (0.009)	−0.042*** (0.009)
Observations	32,684	32,338	32,267	32,310

Notes: Standard errors in parentheses, clustered at the hospital and county levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All outcomes in logs, per discharge. All specifications include hospital and year fixed effects and the county controls (log population, unemployment rate, log income per capita), and instrument Medicare share with the predetermined component of population aging. Panels B and C add total inpatient discharges and the Medicare case-mix index, respectively, to the baseline controls. The case-mix index, from the IPPS Final Rule Impact Files, is unavailable for hospitals not paid under the IPPS, which accounts for the smaller sample in Panel C.

A.5.2 Pre-Period Trends

I implement the test of Section 3.4 as two nine-year long differences on the balanced sample of hospitals observed in 2000, 2009, 2010, and 2019. Let $\Delta Z_c \equiv Z_{c,2019} - Z_{c,2010}$ denote the predicted change in county c 's elderly share over the sample period, constructed from equation (1). I estimate, in reduced form,

$$\Delta_{00-09} \log Y_j = \delta_0 + \lambda \Delta Z_{c(j)} + \Delta_{00-09} X_j' \beta + u_j, \quad (44)$$

$$\Delta_{10-19} \log Y_j = \delta_0 + \pi \Delta Z_{c(j)} + \Delta_{10-19} X_j' \beta + u_j, \quad (45)$$

where Δ_{00-09} and Δ_{10-19} difference each variable between the endpoints of the pre-period and of the sample period, X_j contains the county controls of Section 3.2, and standard errors are clustered by county. Because ΔZ_c is a function of the 1990 age distribution and national survival rates only, it should not predict outcome changes in the decade before the crossing cohorts reach the Medicare eligibility age. The placebo test is therefore $\lambda = 0$. The trajectory-control specification adds the hospital's own pre-period change to equation (45):

$$\Delta_{10-19} \log Y_j = \delta_0 + \pi \Delta Z_{c(j)} + \rho \Delta_{00-09} \log Y_j + \Delta_{10-19} X_j' \beta + u_j. \quad (46)$$

Table 11 reports the estimates of λ , of π , and of π under the trajectory control. Figure 17 plots them.

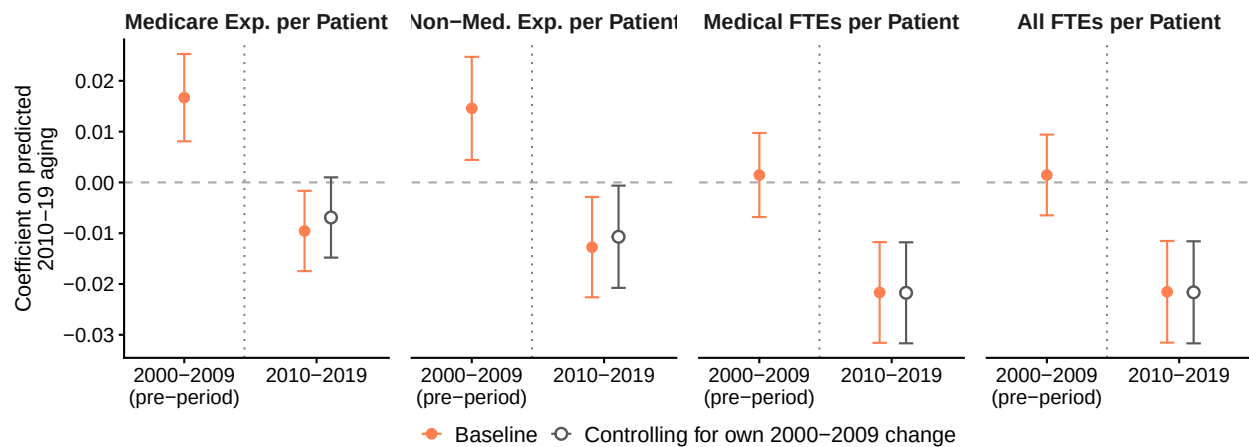


Figure 17: Pre-period placebo. Each facet plots, for one outcome, the coefficient on the predicted 2010-2019 change in the elderly share: the 2000-2009 placebo λ , the 2010-2019 effect π , and the effect under the trajectory control of equation (46). Bars show 95% confidence intervals, clustered by county. Balanced sample of hospitals observed in 2000, 2009, 2010, and 2019.

Table 11: Pre-Period Placebo: Long-Difference Estimates

	Medicare Exp. (1)	Non-Med. Exp. (2)	Medical FTEs (3)	All FTEs (4)
<i>Panel A: Placebo, 2000–2009 changes (λ)</i>				
Predicted Δ elderly share	0.017*** (0.004)	0.015*** (0.005)	0.001 (0.004)	0.001 (0.004)
<i>Panel B: Effect, 2010–2019 changes (π)</i>				
Predicted Δ elderly share	−0.010** (0.004)	−0.013** (0.005)	−0.022*** (0.005)	−0.022*** (0.005)
<i>Panel C: Effect, controlling for the 2000–2009 change</i>				
Predicted Δ elderly share	−0.007* (0.004)	−0.011** (0.005)	−0.022*** (0.005)	−0.022*** (0.005)
Observations	2,839	2,548	2,578	2,582

Notes: Standard errors in parentheses, clustered by county. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Each column is an outcome in logs, differenced between the endpoints of the indicated period. The regressor is the predicted change in the county elderly share over 2010–2019, ΔZ_c , and all specifications include the differenced county controls. Panels A and B report equations (44) and (45); Panel C reports equation (46), which adds the hospital’s own 2000–2009 outcome change as a control. Balanced sample of hospitals observed in 2000, 2009, 2010, and 2019.

B Demand Definitions

This appendix collects formal definitions of the demand-side objects referenced in Section 4.2.

Under the logit model, the expected utility that enrollee i derives from having hospital j in the network is $-\log(1 - \sigma_{ij}^C)$.³² Aggregating over enrollees:

$$GW_{jg} = - \sum_{i \in g} \log(1 - \sigma_{ij}^C). \quad (47)$$

This is the total welfare gain (in utils) that insurer g ’s enrollees derive from having hospital j in the network.

When hospital j is excluded, its patients are redistributed to remaining alternatives proportionally to their choice probabilities (IIA). The additional DRG-weighted demand at alternative k when j is excluded is:

$$\Delta_g[j, k] = \sum_{i \in g} w_i \cdot \frac{\sigma_{ij}^C \cdot \sigma_{ik}^C}{1 - \sigma_{ij}^C}, \quad k \in (\mathcal{N}_{gt} \setminus \{j\}) \cup \{0\}, \quad (48)$$

where $\sigma_{i0}^C = 1 - \sum_l \sigma_{il}^C$ is the probability of choosing the outside option. This formula applies uniformly to all alternatives, including the outside option $k = 0$.

Since removing hospital j redistributes its entire DRG-weighted demand to the remaining al-

³²Let $S = \exp(\eta_t^C) + \sum_{k \in \mathcal{N}_{gt}} \exp(V_{ik}^C)$, so $W_i(\mathcal{N}_{gt}) = \log S$. Removing j gives $S' = S - \exp(V_{ij}^C) = S(1 - \sigma_{ij}^C)$, so $W_i(\mathcal{N}_{gt}) - W_i(\mathcal{N}_{gt} \setminus j) = -\log(1 - \sigma_{ij}^C)$.

ternatives, the diversion terms sum to j 's demand:

$$\sum_{k \in (\mathcal{N}_{gt} \setminus \{j\}) \cup \{0\}} \Delta_g[j, k] = Q_{jgt}^C, \quad (49)$$

with all summands non-negative.

B.1 Quality derivatives of demand objects

Since the quality coefficient θ_i^C is heterogeneous across patients, the quality derivatives of the demand objects involve patient-level summations weighted by θ_i^C .

The own-quality and cross-quality derivatives of commercial demand are:

$$\frac{\partial Q_{jgt}^C}{\partial y_{jt}} = \sum_{i \in g} w_i \theta_i^C \sigma_{ij}^C (1 - \sigma_{ij}^C), \quad \frac{\partial Q_{kgt}^C}{\partial y_{jt}} = - \sum_{i \in g} w_i \theta_i^C \sigma_{ik}^C \sigma_{ij}^C \quad (k \neq j). \quad (50)$$

The corresponding derivatives of enrollee welfare are:

$$\frac{\partial GW_{jg}}{\partial y_{jt}} = \sum_{i \in g} \theta_i^C \sigma_{ij}^C, \quad \frac{\partial GW_{kg}}{\partial y_{jt}} = - \sum_{i \in g} \theta_i^C \frac{\sigma_{ij}^C \sigma_{ik}^C}{1 - \sigma_{ik}^C} \quad (k \neq j). \quad (51)$$

The own-quality derivative of demand diversion has a uniform formula for all $k \neq j$ (including $k = 0$):

$$\frac{\partial \Delta_g[j, k]}{\partial y_{jt}} = \sum_{i \in g} w_i \theta_i^C \sigma_{ij}^C \sigma_{ik}^C. \quad (52)$$

The cross-quality derivatives of demand diversion are:

$$\frac{\partial \Delta_g[k, l]}{\partial y_{jt}} = - \sum_{i \in g} w_i \theta_i^C \sigma_{ij}^C \sigma_{ik}^C \sigma_{il}^C \frac{2 - \sigma_{ik}^C}{(1 - \sigma_{ik}^C)^2} \quad (k \neq j, l \neq j), \quad (53)$$

and

$$\frac{\partial \Delta_g[k, j]}{\partial y_{jt}} = \sum_{i \in g} w_i \theta_i^C \sigma_{ij}^C \sigma_{ik}^C \frac{(1 - \sigma_{ij}^C)(1 - \sigma_{ik}^C) - \sigma_{ij}^C}{(1 - \sigma_{ik}^C)^2} \quad (k \neq j). \quad (54)$$

The quality derivative of Medicare demand takes the same form:

$$\frac{\partial Q_{jt}^M}{\partial y_{jt}} = \sum_{i \in \text{Med}_t} w_i \theta_i^M \sigma_{ij}^M (1 - \sigma_{ij}^M). \quad (55)$$

C Cost Function Derivation

This appendix derives the marginal cost equations (13) and (14) from Section 4.3. I assume that producing one DRG-weighted discharge requires a Leontief combination of one unit of medical service, whose cost rises with quality, and one unit of administrative service, whose cost does not

vary with quality. For the medical-service technology, I follow [Grieco and McDevitt \(2017\)](#) in assuming that hospitals combine multiple factor inputs to produce a single output that they can use either for more discharges at the same quality or for higher quality at the same quantity, with a transformation function $T(Q, y)$ specifying the medical-service output required to produce Q DRG-weighted discharges at quality y .

Let H_{jt} and S_{jt} denote the total medical-service and administrative-service inputs used by hospital j in market t . The Leontief production function to generate one unit of output is given by

$$Q_{jt} = \min\{H_{jt}, S_{jt}\}, \quad (56)$$

so at an efficient interior optimum, $H_{jt} = S_{jt} = Q_{jt}$. Each unit of medical-service input has a dollar cost $f(y_{jt})$, which depends on the hospital's quality choice and is derived below from the medical-service production technology. Each unit of administrative-service input has a dollar cost absorbed entirely into a payer-specific structural residual:

$$\text{Medicare: } \nu_{jt}^M, \quad \text{commercial insurer } g : \nu_{jgt}^C, \quad (57)$$

which captures how each hospital's administrative costs, such as pre-authorization requirements, paperwork burden, or reimbursement speed, deviate from the average. The Medicare residual ν_{jt}^M is common to all of hospital j 's Medicare discharges, while the commercial residual ν_{jgt}^C varies across hospital-insurer pairs. Total dollar cost at hospital j is then

$$C_{jt} = f(y_{jt}) Q_{jt} + \nu_{jt}^M Q_{jt} + \sum_{g=1}^{G_t} \nu_{jgt}^C Q_{jgt}^C, \quad (58)$$

and marginal cost per DRG-weighted discharge is $f(y_{jt}) + \nu_{jt}^M$ for a Medicare patient and $f(y_{jt}) + \nu_{jgt}^C$ for a commercial patient enrolled in insurer g (equations (13) and (14) of the main text). The remainder of this appendix derives $f(y_{jt})$ from the medical-service production technology.

The hospital produces medical-service output by combining K generic variable factors $X_{1jt}, \dots, X_{Kjt}$ through a constant-returns-to-scale Cobb-Douglas technology F :

$$F(X_{1jt}, \dots, X_{Kjt}) = \tilde{A}_t \prod_{k=1}^K X_{kjt}^{\beta_k}, \quad \sum_{k=1}^K \beta_k = 1, \quad (59)$$

where $\tilde{A}_t$ is total factor productivity, common to all hospitals in market t , and the output elasticities β_k sum to one (constant returns to scale). I assume that factor markets are perfectly competitive, such that factor prices $r_{1t}, \dots, r_{Kt}$ are common across hospitals in market t .

Following [Grieco and McDevitt \(2017\)](#), I capture a quality-quantity tradeoff by defining a transformation function T specifying the medical-service output required to produce Q DRG-weighted discharges at quality y . Production efficiency requires that F 's output equal T 's requirement,

yielding the constraint

$$F(X_{1jt}, \dots, X_{Kjt}) = T(Q_{jt}, y_{jt}). \quad (60)$$

I assume that T takes the following functional form:

$$T(Q_{jt}, y_{jt}) = Q_{jt} \cdot e^{\alpha_Y y_{jt}}, \quad \alpha_Y > 0. \quad (61)$$

Holding Q_{jt} fixed, a one-unit increase in y_{jt} scales the medical-service output required by a factor of e^{α_Y} . Substituting (59) and (61) into (60) gives the following production constraint

$$\tilde{A}_t \prod_{k=1}^K X_{kjt}^{\beta_k} = Q_{jt} \cdot e^{\alpha_Y y_{jt}}, \quad (62)$$

where $Q_{jt} = Q_{jt}^M + \sum_g Q_{jgt}^C$ is total DRG-weighted discharges across payers.

For given discharges Q_{jt} and quality y_{jt} , the hospital chooses factor inputs to minimize the dollar cost of medical-service production subject to the production constraint (62):

$$\min_{\{X_{kjt}\}} \sum_{k=1}^K r_{kt} X_{kjt} \quad \text{s.t.} \quad \tilde{A}_t \prod_{k=1}^K X_{kjt}^{\beta_k} = Q_{jt} \cdot e^{\alpha_Y y_{jt}}.$$

The standard CRS Cobb-Douglas solution gives optimal factor demands $X_{kjt}^* = \beta_k c_t Q_{jt} e^{\alpha_Y y_{jt}} / (r_{kt} \tilde{A}_t)$, where $c_t \equiv \prod_k (r_{kt} / \beta_k)^{\beta_k}$ is the factor-price aggregate that emerges from cost minimization. The dollar cost per unit of F 's output is therefore $c_t / \tilde{A}_t$. Substituting back yields the total dollar cost of medical-service production:

$$C_{jt}^{\text{med}} = \sum_k r_{kt} X_{kjt}^* = \frac{c_t}{\tilde{A}_t} \cdot Q_{jt} \cdot e^{\alpha_Y y_{jt}}.$$

Note that total medical cost is linear in Q_{jt} . This follows from constant returns to scale in F , competitive factor markets (so that factor prices do not respond to hospital volume), and the absence of capacity constraints. The marginal cost of medical-service production is therefore equal to the average cost. Define the effective market-time productivity $A_t \equiv \tilde{A}_t / c_t$. We can then write the per-DRG-weight marginal cost of the medical-service input (in dollars) as

$$f(y_{jt}) \equiv \frac{e^{\alpha_Y y_{jt}}}{A_t},$$

which is equation (11) of the main text. Quality enters multiplicatively through $e^{\alpha_Y y_{jt}}$, so raising quality by one unit scales the per-unit cost by e^{α_Y} . Substituting $f(y_{jt})$ into the total cost equation (58) and differentiating with respect to quantities completes the derivation of the marginal cost equations (13) and (14).

D Bargaining Derivations

D.1 Pricing formula derivation

Taking the first-order condition of the Nash product (15) with respect to P_{jgt}^C :

$$\frac{1}{2} \cdot \frac{Q_{jgt}^C}{Q_{jgt}^C(P_{jgt}^C - MC_{jgt}^C)} + \frac{1}{2} \cdot \frac{\partial \Delta V_{jg} / \partial P_{jgt}^C}{\Delta V_{jg}} = 0.$$

Under zero coinsurance, neither Q_{jgt}^C nor GW_{jg} depends on P_{jgt}^C , and the demand diversion objects are also price-independent. The only term in ΔV_{jg} that depends on the own price P_{jgt}^C is $-Q_{jgt}^C P_{jgt}^C$, so $\partial \Delta V_{jg} / \partial P_{jgt}^C = -Q_{jgt}^C$. Substituting:

$$\frac{1}{2(P_{jgt}^C - MC_{jgt}^C)} - \frac{Q_{jgt}^C}{2 \Delta V_{jg}} = 0 \implies P_{jgt}^C - MC_{jgt}^C = \frac{\Delta V_{jg}}{Q_{jgt}^C},$$

which is equation (16). The markup over marginal cost equals the insurer's surplus per DRG-weighted discharge.

D.2 Separability across insurers

Under zero coinsurance and CRS costs, three conditions ensure separability:

1. **Demand independence.** Insurer g 's demand objects (Q_{jgt}^C , GW_{jg} , $\Delta_g[j, k]$) depend only on insurer g 's enrollee pool and network composition, not on any other insurer's prices or quantities. This follows directly from zero coinsurance: σ_{ij}^C depends on quality and distance, not on prices.
2. **Cost independence.** The medical cost component $f(y_{jt})$ is independent of volume under CRS, and the commercial residual ν_{jgt}^C is insurer-specific, so marginal cost does not depend on demand from other insurers.
3. **Surplus independence.** The insurer surplus ΔV_{jg} in (10) involves only insurer g 's own demand objects and prices.

Together, these imply that the pricing system (16) for insurer g involves only insurer g 's objects. Each insurer's system can therefore be solved independently, reducing the multi-insurer problem to G_t separate bargaining problems per market.

E Quality FOC Derivations

This appendix derives the price-response system (21), its solution (22), and the total derivative $d\Delta V_{jg}/dy_{jt}$ that enters the estimating equation (37). By separability (Appendix D.2), I derive them independently for each insurer.

Differentiating the pricing equation (18) of each hospital $k \in \mathcal{N}_{gt}$ with respect to y_{jt} , using

$\partial MC_{kgt}^C / \partial y_{jt} = \mathbb{1}_{k=j} \cdot \alpha_Y f(y_{jt})$ from (12):

$$\begin{aligned} \frac{dP_{kgt}^{C*}}{dy_{jt}} = \frac{1}{2} & \left[\mathbb{1}_{k=j} \alpha_Y f(y_{jt}) + \tau \frac{d}{dy_{jt}} \left(\frac{GW_{kg}}{Q_{kgt}^C} \right) + \sum_{l \in (\mathcal{N}_{gt} \setminus \{k\}) \cup \{0\}} P_{lgt}^C \frac{d}{dy_{jt}} \left(\frac{\Delta_g[k, l]}{Q_{kgt}^C} \right) \right. \\ & \left. + \sum_{l \in \mathcal{N}_{gt} \setminus \{k\}} \frac{\Delta_g[k, l]}{Q_{kgt}^C} \frac{dP_{lgt}^{C*}}{dy_{jt}} \right]. \end{aligned} \quad (63)$$

For $k = j$ this is equation (20) of the main text, and for $k \neq j$ the indicator removes the cost-of-quality term. Stacking the $|\mathcal{N}_{gt}|$ equations gives

$$\frac{d\mathbf{P}_g^C}{dy_{jt}} = \frac{1}{2} \left(\mathbf{v}_g + \mathbf{D}^g \frac{d\mathbf{P}_g^C}{dy_{jt}} \right),$$

which is equation (21) of the main text. Multiplying both sides by two and collecting the price responses on the left-hand side yields $(2\mathbf{I} - \mathbf{D}^g) \cdot d\mathbf{P}_g^C / dy_{jt} = \mathbf{v}_g$, where the elements of $\mathbf{v}_g$ are

$$v_{kg} = \mathbb{1}_{k=j} \alpha_Y f(y_{jt}) + \tau \frac{d}{dy_{jt}} \left(\frac{GW_{kg}}{Q_{kgt}^C} \right) + \sum_{l \in (\mathcal{N}_{gt} \setminus \{k\}) \cup \{0\}} P_{lgt}^C \frac{d}{dy_{jt}} \left(\frac{\Delta_g[k, l]}{Q_{kgt}^C} \right). \quad (64)$$

The matrix $2\mathbf{I} - \mathbf{D}^g$ is strictly diagonally dominant and hence invertible, so the system has a unique solution,

$$\frac{d\mathbf{P}_g^C}{dy_{jt}} = (2\mathbf{I} - \mathbf{D}^g)^{-1} \mathbf{v}_g,$$

which is equation (22) of the main text. To see this, note that the diagonal elements equal two, while the off-diagonal row sum $\sum_{l \neq k} \Delta_g[k, l] / Q_{kgt}^C \leq 1 < 2$, since removing hospital k from the network redistributes at most its entire demand to the remaining in-network alternatives (49).

With a single in-network hospital ($|\mathcal{N}_{gt}| = 1$), $\mathbf{D}^g = 0$, and the diversion term of v_{jg} vanishes, since $\Delta_g[j, 0] / Q_{jgt}^C = 1$ (49). The system reduces to $dP_{jgt}^{C*} / dy_{jt} = v_{jg} / 2$, confirming equation (19) of the main text.

With identical enrollees, the welfare-per-patient term of equation (19) is positive. Let every enrollee have the same choice probability σ , taste θ , and severity weight w . Then the quality semi-elasticity of the welfare gain is $\theta\sigma / (-\log(1 - \sigma))$ and that of DRG-weighted demand is $\theta(1 - \sigma)$, so the difference is positive if and only if $h(\sigma) \equiv \sigma + (1 - \sigma)\log(1 - \sigma) > 0$. Since $h(0) = 0$ and $h'(\sigma) = -\log(1 - \sigma) > 0$ on $(0, 1)$, the condition holds at every interior choice probability. The welfare gain is convex in the choice probability while demand is linear in it, so quality raises the welfare j adds per patient, and negotiated commercial prices increase with quality.

The estimating equation (37) uses the *total* derivative $d\Delta V_{jg} / dy_{jt}$, which accounts for how equilibrium commercial prices respond to hospital j 's quality change. Differentiating (10) with

respect to y_{jt} :

$$\begin{aligned} \frac{d\Delta V_{jg}}{dy_{jt}} = & \tau \cdot \frac{\partial GW_{jg}}{\partial y_{jt}} - \frac{\partial Q_{jgt}^C}{\partial y_{jt}} \cdot P_{jgt}^C - Q_{jgt}^C \cdot \frac{dP_{jgt}^{C*}}{dy_{jt}} \\ & + \sum_{k \in (\mathcal{N}_{gt} \setminus \{j\}) \cup \{0\}} \left[\frac{\partial \Delta_g[j, k]}{\partial y_{jt}} \cdot P_{kgt}^C + \Delta_g[j, k] \cdot \frac{dP_{kgt}^{C*}}{dy_{jt}} \right], \end{aligned} \quad (65)$$

where $dP_{0gt}^{C*}/dy_{jt} = 0$ (the outside option price is exogenous). The partial derivatives are given in Appendix B.1. Substituting the price responses from (22) yields $d\Delta V_{jg}/dy_{jt}$.

F Demand Estimation Appendix

F.1 Medicare Type Distribution

The supply-side estimation in Section 6 and the counterfactual exercises in Section 7 require Medicare demand objects—DRG-weighted quantities Q_{jt}^M and their quality derivatives $\partial Q_{jt}^M/\partial y_{jt}$ —at the hospital-year level. Because individual Medicare records reside on a separate server from the commercial claims data and cannot be transferred, I summarize the Medicare population using a discretized type distribution that is sufficient for reconstructing all demand objects at any quality vector.

Note that each Medicare patient’s utility (30) depends on demographics only through two composite scalars: the heterogeneous distance coefficient $\alpha_i^M = \alpha_0^M + \boldsymbol{\alpha}_{\text{dem}}^M \cdot \mathbf{X}_i + \alpha_{\text{sev}}^M w_i$ and the patient’s severity w_i (DRG weight). All demand objects at any quality vector $\mathbf{y}$ are therefore fully determined by the joint distribution of (α_i^M, w_i) within each ZIP code, together with the estimated demand parameters and the distance matrix.

I discretize this joint distribution using an adaptive nested quantile binning algorithm. For each ZIP-year pair (z, t) , I sort Medicare beneficiaries by α_i^M and partition them into K_z groups of approximately equal size, then within each group sort by severity and partition again into K_z subgroups. The resolution K_z is chosen adaptively as $K_z = \min(K_{\max}, \lfloor \sqrt{N_{zt}/11} \rfloor)$, where N_{zt} is the number of unique beneficiaries in ZIP z in year t and $K_{\max} = 30$. This formula guarantees that every cell contains at least 11 unique beneficiaries, satisfying the CMS disclosure requirement by construction.

Each cell k in ZIP z is summarized by three statistics: the cell count n_{zk} , the mean distance coefficient $\bar{\alpha}_{zk}$, and the mean centered severity $\bar{w}_{zk}$ (DRG weight minus one). Given these sufficient statistics, I reconstruct Medicare demand deterministically. For each type k in ZIP z , the utility of hospital j is

$$V_{jzk} = \delta_{jt}^M + \bar{\alpha}_{zk} d_{zj} + \theta_{\text{sev}}^M \bar{w}_{zk} y_{jt},$$

where δ_{jt}^M is the hospital-year fixed effect (which absorbs the mean quality effect $\theta_0^M y_{jt}$) and d_{zj} is the distance from ZIP z to hospital j . DRG-weighted demand and its own-quality derivative are

then

$$Q_{jt}^M = \sum_z \sum_k n_{zk} (\bar{w}_{zk} + 1) \sigma_{jzk}, \quad (66)$$

$$\frac{\partial Q_{jt}^M}{\partial y_{jt}} = \sum_z \sum_k n_{zk} (\bar{w}_{zk} + 1) (\theta_0^M + \theta_{\text{sev}}^M \bar{w}_{zk}) \sigma_{jzk} (1 - \sigma_{jzk}), \quad (67)$$

where $\sigma_{jzk} = \exp(V_{jzk}) / [1 + \sum_l \exp(V_{lzk})]$ is the logit choice probability for type k in ZIP z . Because the type distribution is a deterministic summary of the patient population, the reconstruction involves no simulation draws and introduces no sampling noise.

I verify the accuracy of the discretization by comparing the grid-based demand objects to those computed directly from the individual-level Medicare data. Figure 18 plots grid-based against micro-based values of Q_{jt}^M and $\partial Q_{jt}^M / \partial y_{jt}$ for all 265 hospital-year observations in the estimation sample. The cross-hospital correlations are 0.998 for DRG-weighted demand and 0.997 for the quality derivative, confirming that the discretized type distribution closely reproduces the demand objects that would be obtained from the full patient-level data.

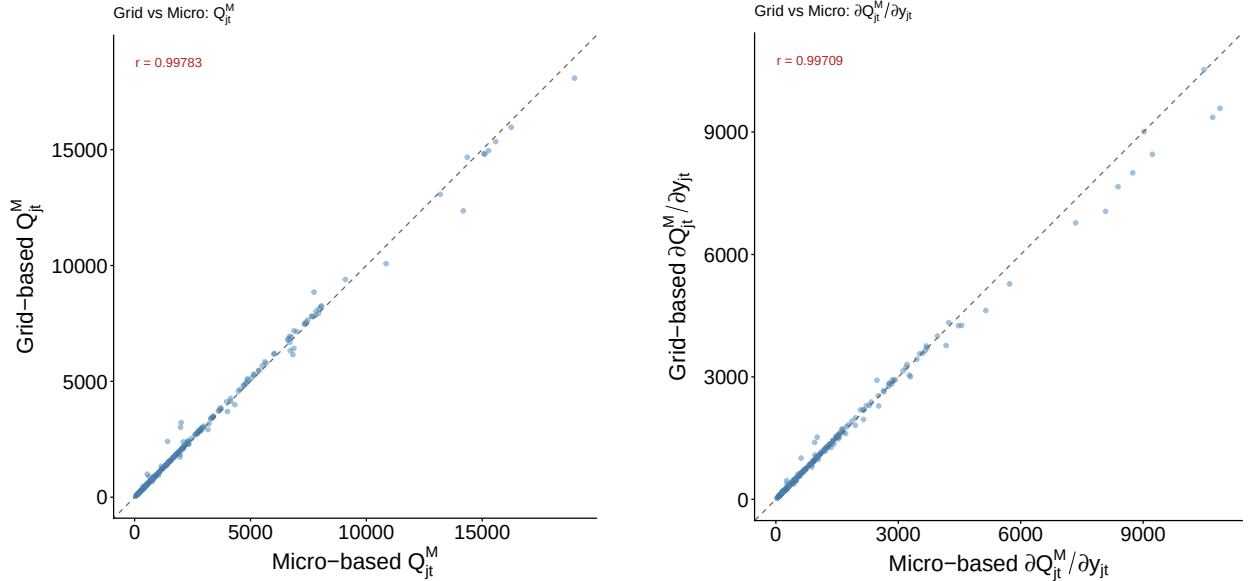


Figure 18: Validation of grid-based Medicare demand reconstruction. Each point is a hospital-year. Left panel: DRG-weighted demand Q_{jt}^M ; right panel: own-quality derivative $\partial Q_{jt}^M / \partial y_{jt}$. Dashed lines are 45-degree lines. Grid-based objects are computed from the discretized type distribution (66)–(67); micro-based objects are computed from individual-level Medicare claims.

F.2 Network Classification

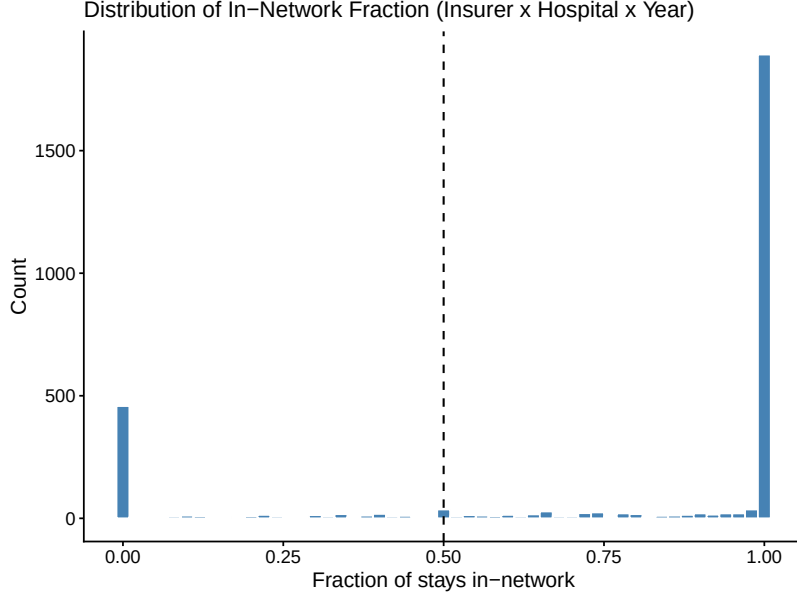


Figure 19: Distribution of in-network fraction across insurer–hospital–year cells. Each cell’s in-network fraction is the share of non-emergency stays coded as in-network. The dashed line marks the 50 percent threshold used to classify cells as in-network. The distribution is concentrated at zero and one, indicating that the classification is not sensitive to the choice of threshold.

G Cost and Bargaining Estimation Appendix

G.1 Payer-Mix Calibration

The MA-APCD captures commercially insured hospitalizations filed by insurers operating in Massachusetts but does not observe the full universe of commercial stays, because self-insured employer plans are largely absent following *Gobeille v. Liberty Mutual* (2016). The Medicare sample, by contrast, covers the universe of fee-for-service beneficiaries. If left uncorrected, the APCD-based payer mix would overstate the Medicare share of hospital demand, biasing the relative weight of the Medicare and commercial terms in the quality first-order condition (29).

I correct for this by rescaling commercial demand in each market-year to match the aggregate payer mix observed in the Healthcare Cost Report Information System (HCRIS). HCRIS reports each hospital’s total Medicare and non-Medicare inpatient days, from which I compute the Medicare share of inpatient days at Massachusetts general acute-care hospitals in each year. For each market-year t , I compute a scaling factor λ_C^t such that the ratio of Medicare to commercial DRG-weighted demand in the estimation sample matches the HCRIS Medicare share:

$$\lambda_C^t = \frac{s_t^C / (1 - s_t^C)}{s_t^{C, \text{APCD}} / (1 - s_t^{C, \text{APCD}})}, \quad (68)$$

where s_t^C is the commercial share in HCRIS and $s_t^{C, \text{APCD}}$ is the commercial share implied by the raw APCD and Medicare counts. All commercial demand objects—DRG-weighted quantities Q_{jgt}^C ,

enrollee welfare GW_{jg} , demand diversion $\Delta_g[j, k]$, and their quality derivatives—are multiplied by λ_C^t before entering the estimation. Medicare demand objects are held at their observed levels ($\lambda_M = 1$). Under the model’s constant-returns-to-scale cost structure, this rescaling is equivalent to adjusting the size of the commercial patient pool while holding the Medicare pool fixed.

G.2 Estimation Diagnostics

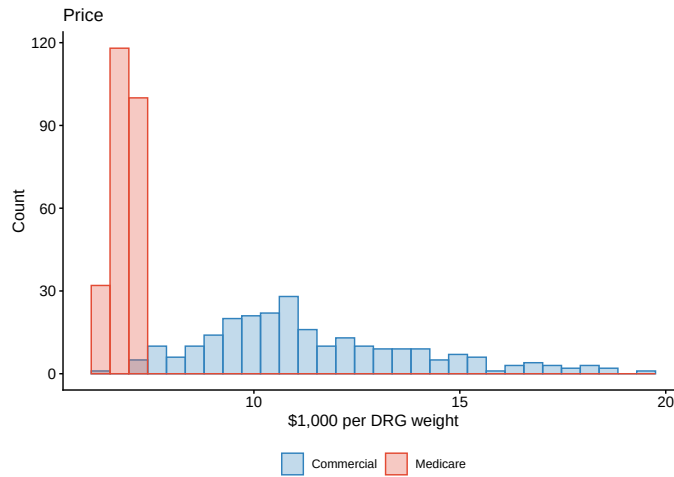


Figure 20: Distribution of prices by payer type in the estimation sample. Each observation is a hospital-year. Commercial prices are insurer-weighted averages across all in-network insurers at each hospital-year. Medicare prices are computed from the IPPS Final Rule Impact Files. Units are thousands of dollars per DRG-weight.

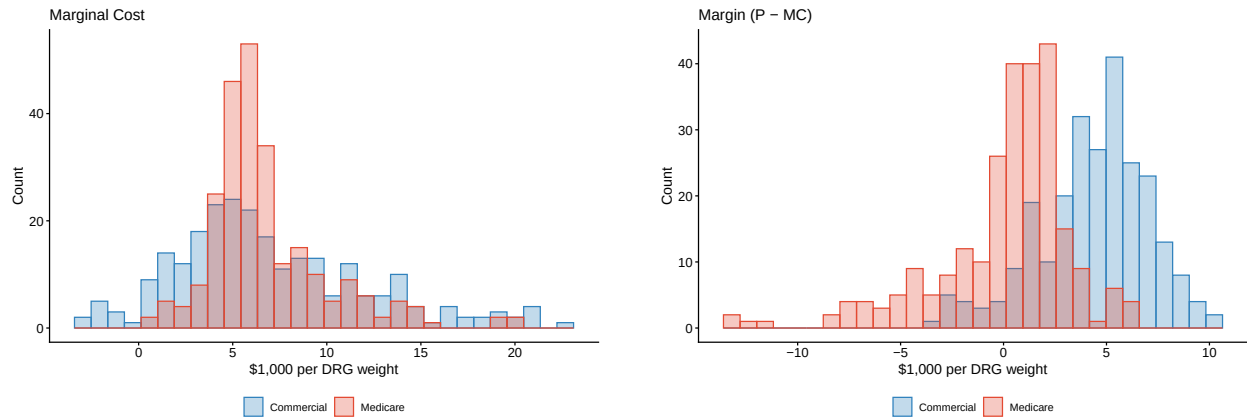


Figure 21: Distribution of implied marginal costs (left) and implied margins (right) by payer type. Each observation is a hospital-year. Commercial values are insurer-weighted averages across all in-network insurers at each hospital-year. Margins are defined as price minus marginal cost ($P - MC$) per DRG-weight. Units are thousands of dollars per DRG-weight.

H Counterfactual Details

H.1 Projecting the Payer Mix

This appendix describes how I project the Medicare share used in the population aging counterfactual of Section 7.1. I project the payer mix in Massachusetts through 2040 in three steps. First, for each age and gender group, I age the population forward using national survival rates. Second, I convert the projected population in each age and gender group into DRG-weighted hospital stays using fixed hospitalization rates by age and gender, computed from MedPAR and MBSF for patients aged 65 and older and from the APCD for patients younger than 65. Finally, I apply the projected change in the ratio of Medicare to commercial DRG-weighted stays to the last payer mix observed in HCRIS, in 2023.

Let $N_{a,x,t}$ denote the Massachusetts population of age a and gender x in year t , from the SEER population data described in Section 3.1, by single year of age from 0 to 84, with everyone aged 85 and over grouped in a single category. Starting from the observed 2022 age distribution, each cohort is aged forward one year at a time,

$$\hat{N}_{a+1,x,t+1} = S_{a,x} \hat{N}_{a,x,t}, \quad (69)$$

where $S_{a,x}$ is the national survival rate from age a to age $a + 1$, measured as the ratio of the national population of age $a + 1$ and gender x in a given year to the national population of age a and gender x in the preceding year. I estimate it by averaging this ratio over 2015–2016 through 2018–2019 and 2021–2022, excluding the COVID years. Fertility determines the cohorts born after 2022, which enter the projection at age zero. I hold fertility (births per woman aged 15–44) fixed at its 2022 rate and apply it to the projected female population aged 15–44 in each year.

The projected population is converted into inpatient demand with fixed hospitalization rates by age and gender. The hospitalization rate $h_{a,x}$ is the DRG-weighted inpatient volume per Massachusetts resident of age a and gender x per year,

$$h_{a,x} = \frac{\sum_{t \in \{2018, 2019, 2022\}} Q_{a,x,t}}{\sum_{t \in \{2018, 2019, 2022\}} N_{a,x,t}}, \quad (70)$$

where $Q_{a,x,t}$ is the total DRG-weighted inpatient stays of patients of age a and gender x in year t , measured in the Medicare claims for ages 65 and over and in the APCD for ages 0–64 (the estimation years, excluding the COVID years). The projected ratio of Medicare to commercial DRG-weighted inpatient volume in year t is

$$\hat{R}_t = \frac{\sum_{a \geq 65} \sum_x h_{a,x} \hat{N}_{a,x,t}}{\sum_{a < 65} \sum_x h_{a,x} \hat{N}_{a,x,t}}. \quad (71)$$

The projected Medicare share $\hat{s}_t^M$ grows the last observed payer mix at this demographically

predicted rate. Writing the payer mix as the ratio of the Medicare share to the commercial share,

$$\frac{\hat{s}_t^M}{\hat{s}_t^C} = \frac{s_{2023}^M}{s_{2023}^C} \times \frac{\hat{R}_t}{\hat{R}_{2023}}, \quad (72)$$

where $\hat{s}_t^C = 1 - \hat{s}_t^M$ and 2023 is the last year observed in HCRIS. The projected share therefore coincides with the observed share in 2023. Because only the ratio $\hat{R}_t/\hat{R}_{2023}$ enters, an undercount of either payer's stays cancels from equation (72) as long as the claims capture the same fraction of the payer's stays at every age: the self-insured enrollees the APCD misses must have the same distribution of DRG-weighted stays across ages as the fully insured enrollees it covers, and the Medicare Advantage beneficiaries the Medicare claims miss must have the same distribution as the fee-for-service beneficiaries they cover. Figure 22 plots the observed and projected series.

I solve for the counterfactual equilibrium quality under the 2040 payer mix by multiplying the Medicare terms in each hospital's quality first-order condition (29) by

$$\lambda^M = \frac{s_{2040}^M/(1 - s_{2040}^M)}{s_{2019}^M/(1 - s_{2019}^M)} = 1.537, \quad (73)$$

where s_{2019}^M is the Medicare share of inpatient days observed in HCRIS in 2019, and s_{2040}^M is its projected value from equation (72). Multiplying the Medicare terms by λ^M replaces Q_{jt}^M with $\lambda^M Q_{jt}^M$. Because the Medicare rate, marginal cost, and the rate's response to quality do not change, AMR_{jt}^M is unchanged:

$$\frac{1}{\lambda^M Q_{jt}^M} \left[(P_{jt}^M - MC_{jt}^M) \lambda^M \frac{\partial Q_{jt}^M}{\partial y_{jt}} + \lambda^M Q_{jt}^M \frac{\partial P_{jt}^M}{\partial y_{jt}} \right] = AMR_{jt}^M.$$

The rescaling therefore changes only the Medicare share, which becomes $\lambda^M Q_{jt}^M / (\lambda^M Q_{jt}^M + Q_{jt}^C)$, as in equation (42).

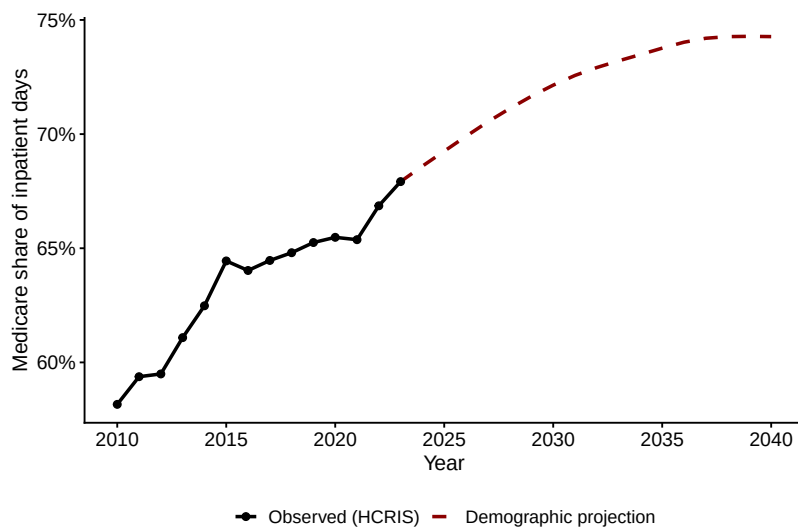


Figure 22: Medicare share of inpatient days at Massachusetts general acute-care hospitals. The solid line is the observed HCRIS series; the dashed line is the demographic projection, which ages the population forward from 2022 and converts it into inpatient demand with fixed hospitalization rates by age and gender.