THE UNIVERSITY OF CHICAGO

ESSAYS ON THE ORGANIZATION OF PRODUCTION IN HEALTHCARE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
PAULINE MOUROT

CHICAGO, ILLINOIS
JUNE 2024

To my sister, Alice,

To my fiancé, Santiago Franco Tabares,

For their unfailing support, always.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am highly indebted to my advisors Joshua Gottlieb, Neale Mahoney, Matthew Notowidigdo, and Stéphane Bonhomme for their guidance and support. I have learned tremendously from you. I will remain extremely grateful for the opportunities you have given me. Thank you for having believed in me.

I am immensely grateful for the entire community at the University of Chicago. Thank you to Professors across the Booth School of Business, the Economics Department, and the Harris School of Public Policy who have provided extremely useful feedback and comments. Particular thanks to Zarek Brot-Goldberg, Eric Budish, Jonathan Dingel, Alex Frankel, Anders Humlum, and Jacob Leshno for their time, comments, and encouragements over the years. Thank you to students across all cohorts for their comments and support. Special thanks to my amazing cohort. We have really been through this together.

Thank you to my parents, Pascale and Dominique, for having given me so much, and for still standing with me in every new challenge I set ahead. Thank you to my brother, Martin, and my sister, Alice, who have made my life infinitely richer from the moment they were born. Special thanks to you, Alice, who have stood with me at every crossroad. I do not think I would have been able to pursue this without you.

Thank you to my friends, who have shaped the person that I am today. Special thanks to Camille Mammar and Maud Hallé, who have been with me *depuis les bancs d'école*, Dalal Moosa and Sulin Sardoschau, who have inspired me deeply at so many levels, and Olivia Bordeu-Gazmuri, who has stood with me since the first day of the Ph.D.

Thank you to my fiancé, Santiago Franco Tabares, who I admire greatly, both professionally and personally. You have inspired me, encouraged me, and supported me at every step of this endeavors. Thank you for making my life more beautiful every day.

ABSTRACT

I examine the organization of production in healthcare and its consequences for healthcare quality, patient outcomes, and access to care for patients. Chapter 1 examines the production technology in healthcare, how surgeons sort across hospitals, and the consequences of this sorting on patient outcomes. Chapter 2 shows the importance of trade and economies of scale in determining the distribution of consumption and production of medical services across the U.S. Chapter 3 investigates the internal organization of healthcare firms and its consequences for physician specialization.

CHAPTER 1

SHOULD TOP SURGEONS PRACTICE AT TOP HOSPITALS? SORTING AND COMPLEMENTARITIES IN HEALTHCARE

Abstract: How does the existence of complementarities between surgeon and hospital quality impact aggregate patient outcomes? Using Medicare data, I examine the joint production function of patient survival between surgeons and hospitals in the context of coronary artery bypass graft (CABG) surgery. Cardiac surgeons tend to be independent from hospitals; they perform surgeries at multiple hospitals within the same year. I leverage this variation in a two-way fixed-effects strategy with interactions between hospital and surgeon quality. I address high-dimensionality issues in a model with two-sided heterogeneity and potential selection of patients into providers using a two-step grouped fixed-effects approach with partial endogenization of network formation. I find that cardiac surgeons engage in positive assortative matching, such that higher-survival surgeons practice at higher-survival hospitals. However, this matching does not maximize aggregate survival: low-survival surgeons have much higher returns from practicing at a high-survival hospital than high-survival surgeons do. Partial equilibrium exercises suggest that 30-day mortality from CABG could be reduced by 20% if low-survival surgeons were reallocated to high-survival hospitals. Half the gains from these national reallocations can be achieved by reallocating surgeons within hospital referral regions.

1.1 Introduction

The sorting of workers to firms and how they combine to produce output are longstanding research questions.[1] These questions are particularly relevant in healthcare, where the literature has separately documented substantial variation in doctor and hospital value-added

1. See for example Chade, Eeckhout, and Smith (2017) and Eeckhout (2018) for reviews of this literature.

in the production of health (Chandra et al., 2016*a*,*b*; Birkmeyer et al., 2013; Currie and MacLeod, 2017). Whether hospitals and doctors are complements or substitutes in the health production function and whether the observed sorting of doctors to hospitals maximizes aggregate health output have received limited attention.

This paper has two objectives. First, I estimate the value-added of surgeons and hospitals and their interactions in the production function for a surgery treating a common cardiovascular disease. Second, using the estimated production function, I evaluate the impact of current and counterfactual allocations of surgeons to hospitals on aggregate patient outcomes. To do so, I focus on coronary artery bypass graft (CABG) surgery, a common surgery performed on about 200,000 Americans at an aggregate cost of over $7 billion every year. This surgery has an unambiguously relevant and well-defined measure of output, patient operative survival, which has also been the focus of the literature investigating provider quality.[2]

I estimate the joint production of patient survival between surgeons and hospitals using a two-way fixed-effects strategy with interactions in their value-added. Although variation in this type of empirical strategy traditionally comes from "job movers", the cardiac surgery setting allows me to leverage an additional source of variation: surgeons tend to practice at multiple hospitals like freelancers, which I call "multi-homing". Using a well-defined measure of output at the surgeon-hospital level, I can address identification issues for sorting which have been outlined in two-way fixed-effect models when using worker earnings as a measure of worker productivity in the presence of frictions (Eeckhout and Kircher, 2011).[3]

Estimating individual provider value-added and their interactions in a model with two-sided heterogeneity raises two main challenges. First, individual value-added estimates are

2. In the case of CABG surgery specifically, see for example Huckman and Pisano (2006); Kolstad (2013).

3. Eeckhout and Kircher (2011) show that, in any sorting equilibrium with search frictions, worker wages are non-monotonic in firm types around the optimal allocation, which prevents from identifying firm types and sorting from wage data alone in a two-way fixed-effects model.

noisy when the outcome is a relatively rare event like mortality. To address this issue, I reduce the dimensionality of the fixed effects by first classifying both surgeons and hospitals into groups using k-means clustering in the spirit of Bonhomme, Lamadon, and Manresa (2022). I classify them using a proxy for their individual quality, provider-level risk-adjusted survival, which is used in practice and in the literature to describe providers' individual quality for CABG surgery. Second, patient selection into providers may violate the exogenous network assumption required for identification of the fixed effects in a two-way fixed-effects model. I use a control function approach leveraging distance to hospitals as an excluded instrument to identify patient selection on unobservables as recently used in Einav, Finkelstein, and Mahoney (2022). In the context of two-sided heterogeneity models, this consists in a partial endogenization of network formation by modeling the choice of hospitals by patients.

I find that surgeon and hospital quality are imperfect substitutes in the production function of survival for CABG surgery. The returns from allocating surgeons to high-survival hospitals are statistically larger for lower-survival surgeons. This result grounds a mechanism outlined in the medical literature–"failure-to-rescue"–into the economics of the technology of production (Silber et al., 1992; Ghaferi, Birkmeyer, and Dimick, 2009). High-survival hospitals achieve higher survival rates by rescuing their patients from complications, and high-skill surgeons tend to exhibit lower complication rates. My estimates suggest that high-survival surgeons tend to achieve high survival rates for their patients irrespective of the hospital at which they perform surgeries. However, low-survival surgeons exhibit much higher survival rates at higher-survival hospitals. Emphasizing the importance of interactions in the production function, I show that a simple variance decomposition without interactions would miss the crucial role of hospital value-added for low-survival surgeons.

Examining sorting, I find positive assortative matching: high-survival surgeons sort into high-survival hospitals. This sorting does not maximize aggregate patient survival in the presence of imperfect substitutability. Furthermore, there exist positive assortative matching within hospital referral regions, indicating that positive assortative matching is not entirely driven by provider location decisions across regions.

I examine the robustness of my results to accounting for selection into providers on unobservables and alternative classification strategies of the unobserved heterogeneity across surgeons and hospitals. Using a control function approach, I show that my results are robust to allowing for patient selection into providers on unobservables. Selection into treatment may be different when CABG surgery is performed in an emergency setting, so I evaluate results when excluding such emergency CABG surgeries and obtain similar findings. I obtain similar results when using a variety of alternative classification strategies for surgeons and hospitals that use different numbers of groups, additional conditional moments of the survival distribution, additional provider characteristics, or simply use quintiles of survival rather than clustering.

Using partial equilibrium reallocation exercises, I show that the current sorting of surgeons across hospitals substantially raises patient mortality. A random reallocation of surgeons to hospitals leads to a 6% decrease in average mortality from CABG surgery. Implementing negative assortative matching would lead to a large decrease in average mortality of 20%. To put these numbers in perspective, this would amount to approximately 800 lives saved within Medicare every year for CABG surgery alone. Furthermore, I find that more than 50% of the gains from national reallocations can be achieved by reallocating surgeons to hospitals within hospital referral regions. While these results do not account for general equilibrium effects that would be necessary to quantify the impact of a specific policy, they indicate that reallocating surgeons within regions may be a fruitful avenue for policy.

This paper contributes to the healthcare literature examining variation in provider quality and its determinants. Previous work has documented substantial variation in quality across providers (Chandra et al., 2016*a*,*b*; Birkmeyer et al., 2013; Currie and MacLeod, 2017; Abaluck et al., 2021; Einav, Finkelstein, and Mahoney, 2022). Hospitals with better management practices, communication technologies, and a larger amount of labor improve patient outcomes (Bloom et al., 2020; Munoz and Otero, 2023; Johnston et al., 2015; Ward et al., 2019). More experienced doctors in the specific procedure tend to produce better patient outcomes (Birkmeyer et al., 2013). This paper is most closely related to Huckman and Pisano (2006), who find evidence that surgeon performance is hospital-specific.[4] They show that surgeon performance is correlated with their volume at the specific hospital, but not to their volume at other hospitals. This paper contributes to this literature in showing that the production technology and surgeon sorting are crucial determinants of individual provider quality.

This paper also contributes to a longstanding literature on worker sorting across firms. Recent work by Kline, Saggio, and Sølvsten (2020) and Bonhomme et al. (2023) has shown substantial positive assortative matching between workers and firms in Europe and the U.S. using worker earnings with fixed effects and random effects approaches. Assuming that markets clear, the direction of sorting maps into the existence of complementarities.[5] Bonhomme, Lamadon, and Manresa (2019) estimate the existence of complementarities directly and find evidence for (weak) imperfect substitutability between workers and firms in the presence of positive assortative matching when using worker earnings.[6] This paper has two

4. In the education literature, Jackson (2013) also finds evidence that teacher effects are not fully portable across schools. He estimates that teacher-school match quality represents between 10% and 40% of what is typically attributed to teacher effects. In my paper, I explore match effects that correspond to the existence of complementarity or substitutability between surgeon and hospital value-added in the production function of health.

5. In the traditional two-way fixed-effects regression with wage data, workers are paid at their marginal productivity, so that higher-wage workers are more productive. If high-wage workers sort into high-wage firms, this indicates that their marginal productivity is higher at these firms rather than at lower-wage firms, hence identifying complementarities between firms and workers in the production function. Note that Eeckhout and Kircher (2011) show that equilibrium wages are not monotonic in firms' productivity in the presence of search frictions, hence preventing from identifying sorting from wage data alone.

6. I focus on the literature using two-way fixed-effects strategies here, closest to my paper. Another literature in labor investigates questions related to worker sorting by estimating sorting models of the labor market. Notably, Lise, Meghir, and Robin (2016) estimate a sorting model with search frictions and find evidence that worker and job characteristics exhibit no complementarities for unskilled workers while they find evidence for complementarities among higher-skill workers.

main contributions to this literature. First, it documents positive assortative matching in a specific yet important labor market using a direct measure of output. Second, it shows the importance of including interactions between workers and firms to both better understand their relative contributions to output, but also to quantify the impact of worker sorting on aggregate output.

The rest of the paper is organized as follows. I describe the institutional setting and provide an overview of the data in Section 1.2. I delineate the empirical strategy in Section 1.3. Section 1.4 evaluates the sensibility of estimated parameters. I detail the imperfect substitutability and sorting results and assess their robustness in Section 1.5. Finally, I quantify the impact of surgeon sorting across hospitals on aggregate patient survival using partial equilibrium reallocation exercises in Section 2.7. Section 1.7 concludes.

1.2 Setting & data description

To study the joint production function of physicians and hospitals, I focus on a complex yet common surgery: coronary artery bypass graft (CABG) surgery. In this section, I first describe the institutional setting of Medicare and CABG surgery. I next detail how I obtain the analysis sample and illustrate basic facts in the data. I describe the existence of "multi-homing" in the data, i.e., the fact that surgeons have operating privileges at multiple hospitals, and detail basic facts about the variation in patient outcomes across surgeons and hospitals separately.

1.2.1 Institutional Setting

Medicare

I use data from Medicare, which is the health-insurer for Americans aged 65 and older and the disabled. In addition to being one of the largest health-insurer in the U.S. with about 60 million insurees every year, Medicare is a federal health-insurance program that provides a relatively even geographic coverage of patients and healthcare providers compared to individual private insurers. About a third of patients 65 years old and older opt for Medicare Advantage plans, which are administered by private health-insurers and usually offer additional coverage such as prescription drugs (Part D). I focus on Traditional Medicare (TM) which includes about 40 million enrollees every year.

Traditional Medicare has two key advantages to study physicians sorting across hospitals. First, it has no network restrictions for enrollees, who can go to any doctor or hospital in the country. Second, patients can access these healthcare providers at the same price. This allows me to abstract away from concerns about the endogeneity of hospital or physician choice to the type of health-insurance.

Coronary artery bypass graft (CABG) surgeries

CABG surgery is one of the treatments of coronary artery disease, the most prevalent heart disease in the U.S., responsible for more than 375,000 deaths in the U.S. in 2021 (Centers of Disease Control and Prevention, 2023). Coronary artery disease is the narrowing of the blood vessels bringing oxygen to the heart muscle. A severe coronary artery blockage can result in an acute myocardial infraction (AMI), an emergent condition that requires immediate treatment to minimize tissue damage and ensure survival.

It is a common and expensive surgery. It is performed on about 200,000 Americans every year, of whom about half are 65 years old and older, and was in the top 20 most common operating room procedures in 2018 (McDermott and Liang, 2021). CABG surgery is also an expensive surgery: it costs about $47,000 on average per hospital stay, for an aggregate cost of more than $7 billion every year, bringing it to the top 6 in aggregate cost in 2018 (McDermott and Liang, 2021). CABG surgery represents a large fraction of cardiac surgeons' activity: this surgery is their most common surgery on average on Medicare patients, followed by heart-valve replacement and aortic surgery.

Patient outcomes after CABG surgery represent a meaningful measure of provider quality for both surgeons and hospitals. Both the hospital and the operating surgeon have a substantial role to play in determining patient outcomes from this surgery. While the operating surgeon's skill is crucial to successfully restore blood flow, the hospital determines the rest of the team needed to successfully treat CABG patients, both during and after the surgery. I give more details on the processes involved during CABG surgery in Appendix 1.9.1. Probably for this reason, measures of provider-level risk-adjusted operative mortality for CABG surgery started being reported in the 1990s as part of report-card programs as a signal for provider quality. Hospital 30-day risk-adjusted mortality rates after CABG surgery are publicly reported yearly by CMS and integrated in their hospital five-star rating measure. For surgeons, 30-day risk-adjusted mortality calculated at the surgeon level for CABG surgery started being publicly reported in the state of New York in 1991, followed by other states including Pennsylvania and Massachussetts.[7]

Cardiac surgeons operate on their patients at multiple hospitals within the same year without an actual change of employment, which I leverage as additional variation in my empirical strategy. A large fraction of cardiothoracic surgeons are not employed by hospitals, but are rather independent in private practices, like freelancers (Huckman and Pisano, 2006; Kolstad, 2013). To get access to an operating room to perform surgery, they need to obtain operating privileges at hospitals. Obtaining such privileges is relatively low cost, under the form of a one time administrative cost, and there is no limit in the number of hospitals at which they can obtain operating privileges. While potentially costly, operating at multiple hospitals allows for flexibility for surgeons, as detailed in Appendix 1.9.1.

7. An extensive literature has investigated the impact of provider-level quality reporting on consumer choice and provider behavior. See for example Kolstad and Chernew (2009) for a review of this literature.

1.2.2 *Data*

Sample construction

I infer the identity of the surgeon operating on a specific patient, and in which hospital, by merging Medicare professional fee files–the Carrier files–to inpatient hospital data–the MedPAR files–following a similar approach as Chen (2021). I use procedure codes in the Carrier files to select CABG surgeries and identify the operating surgeon using the performing physician. The Carrier file contains professional fees for a random sample of 20% of Medicare beneficiaries each year. To identify the hospital in which the surgery took place, I use the claim date and the patient identifier to merge the professional fee into the file containing the universe of hospital stays for Medicare, the MedPAR files, which identify the hospital in which a surgery took place. Further details on this matching process are included in Appendix 1.9.2. I am able to match 90% of the CABG surgeries from the 20% sample Carrier file as reported in Table 1.1.

I restrict the sample in four main dimensions as reported in Table 1.1. First, I restrict my attention to surgeons whose specialty is consistent with CABG surgery to make sure that I capture the operating surgeon and I exclude residents. I do so using external data from the National Plan and Provider Enumeration System (NPPES) to identify the specialty of the physician. Second, I exclude patients who have been admitted at the end of 2010, but discharged in 2011 when my data starts, since I do not observe all claims from 2010. Third, I restrict my attention to surgeon-hospital pairs with more than five observations in the time period. This imposes a minimum of five surgeries per hospital and per surgeon, so that the activity of a surgeon at a specific hospital can be more precisely estimated. Very low Medicare volume surgeons and hospitals are therefore excluded. Fourth, I exclude patients residing or treated by providers outside of mainland U.S. to ensure that patients can be matched to a hospital referral region (HRR). To align the samples when using a control function or not, I also exclude hospital-surgeon pairs in hospital referral regions where patients only received CABG surgery from hospitals outside the HRR, since I cannot estimate demand for these patients using HRR as market definition.

The final sample includes a total of 111,059 patients treated by 2,911 surgeons across 1,167 hospitals between 2011 and 2017.[8] Patient covariates are reported in Table 1.2. The Charlson score is a measure of health: it aggregates seventeen comorbidities based on severity from diagnoses listed in all claims in the past twelve months prior to surgery into a score from 0 to 24, with a larger score indicating poorer health. Patients undergoing CABG surgery exhibit an average Charlson score of 3.41, indicating moderate to high health risk. 40% of patients in the sample have had an acute myocardial infarction (AMI) and 42% of them received a diagnosis of congestive heart failure (CHF) in the year prior to surgery. Consequently, short-term mortality after CABG surgery is non-null: the average mortality after CABG surgery is 5% within 30 days and 6% within 60 days in the final sample.

Surgeons practicing at multiple hospitals within a year: "multi-homing"

The fact that surgeons operate on patients at multiple hospitals within the same year, which I call "multi-homing," is a sizeable additional variation to the usual variation provided by job movers in employer-employee matched data. I define multi-homers as surgeons observed in more than one hospital in at least four years of the final sample, so more than half of the sample time frame. I consider all other surgeons observed in more than one hospital as "traditional movers," i.e surgeons who shifted their entire practice from one hospital to the next. This data-driven definition is imperfect, but it allows me to get a sense of the additional variation provided by multi-homers. Note that this definition assumes that

8. The total number of observations is larger than the total number of patients for two reasons. First, 604 patients received CABG surgery more than once in the 2011-2017 final sample. Second, 16.5% of surgeries exhibit more than one performing surgeon in the final sample. When more than one surgeon operates on a patient, I assign the patients' outcome from the surgery to both surgeons assuming these observations are independent.

surgeons would not change hospital employment four or more time in seven years, but also that a surgeon observed at multiple hospitals within a year three times or less necessarily changed employment. It can therefore both under- and overestimate the share of multi-homers in the data, but it has no impact on the aggregate variation used for identification. Based on this definition, Figure 1.1 shows that multi-homers represent close to 13% of surgeons and 19% of surgeries in the final sample, as compared to 25% for traditional movers. Overall, the number of surgeons observed at multiple hospitals represents almost 40% of surgeons in the final sample.

Differences in patients treated for each surgeon category are reported in Appendix Table 1.9. Overall, multi-homers are more likely to treat younger, lower-income patients in large-population ZIP codes, who tend to be sicker at baseline. These differences likely reflect the location of surgeons: multi-homers are more likely to practice in larger cities, which is consistent with the capacity constraint explanation for the source of multi-homing for surgeons. Another notable difference is that traditional movers tend to have graduated from medical school more recently than single-homers and multi-homers, consistent with early carrier job moves.

Multi-homing happens for a susbtantial share of a surgeon's activity. Appendix Table 1.10 describes how multi-homers split their activity when they multi-home. When a surgeon operates at two hospitals in a year, on average 30% of their surgeries are performed outside of their top-choice hospital. When a surgeon operates at three or more hospitals, more than 40% of their surgeries are performed outside of their top-choice hospital. In conclusion, multi-homing is not a marginal practice at the surgeon level.

Motivating facts

The variation in patient survival across surgeons and hospitals separately is quantitatively important for CABG surgery. In Figure 1.2, I depict the average 30-day risk-adjusted survival for hospitals and surgeons separately, after adjusting for measurement error using empirical Bayes shrinkage techniques detailed in Appendix 1.9.4. I compute the risk-adjusted survival at the patient level using a logit model including patient observables as delineated in Appendix 1.9.3. The standard deviation across hospitals represents about 2 percentage points of 30-day survival after adjusting for measurement error. This represents about 40% of the average 30-day mortality for CABG surgery. It is comparable for surgeons.[9]

Note that the averages across hospitals include the impact of surgeons and vice versa in Figure 1.2: we cannot immediately isolate the relative contributions of surgeons versus hospitals in patient outcomes. Are surgeons substantial contributors to the variance in patient outcomes compared to hospitals? Does a surgeon's ability to save patients vary with the quality of the hospital? Are high-survival hospitals high quality only because they attract top surgeons? Would more patients survive if surgeons were allocated to different hospitals? The goal of this paper is to answer these questions to gain novel insights into the determinants of provider quality.

1.3 Empirical strategy

I leverage the existence of multi-homers and traditional movers in a two-way fixed-effects approach including interactions between surgeon and hospital unobserved heterogeneity. I first classify hospitals and surgeons into groups using k-means clustering in the spirit of Bonhomme, Lamadon, and Manresa (2019, 2022). Thanks to this classification step, I can both address estimation error on parameters of interests and estimate the existence and strength of complementarities between surgeons and hospitals, using a non-parametric production function of survival in the second step. Assuming selection on observables, the patient-

9. Without measurement error adjustment, the variation across hospitals and surgeons is larger and larger for surgeons, as delineated in Appendix Table 1.11. The standard deviation across hospitals amounts to 2.8 percentage points of 30-day survival against 3.8 percentage points of 30-day survival across surgeons. Results are similar after risk adjustment.

surgeon-hospital match is assumed to be exogenous conditional on patient observables. I show how to relax this assumption by partially endogenizing network formation through modeling the choice of hospitals and using distances to hospitals as excluded instruments.

1.3.1 Production function of survival for CABG

Assume a production function of survival for patient i treated by surgeon j in hospital h such that

$$Y^*_{ijht} = g(\alpha_j, \psi_h, X_{it}) + \epsilon_{ijht}$$

where α_j and ψ_h are, respectively, the unobserved heterogeneity of the surgeon and hospital, X_{it} are patient observables, and ϵ_{ijht} are unobserved health shocks. For simplicity, I will assume that the unobserved shocks ϵ_{ijht} are mutually independent. This assumption rules out spillover effects where a surgeon may become better as they perform more at a hospital, for example.

Y^*_{ijht} is the potential outcome, here 30-day survival, of patient i treated by surgeon j in hospital h: it takes values 0 if the patients dies and 1 if the patient survives. Note that this is not a latent variable model. The function g describes how surgeon and hospital heterogeneity and patient observables combine to produce patient survival. Assuming $\mathbb{E}[\epsilon_{ijht}|\alpha_j, \psi_h, X_{it}] = 0$, the conditional expectation of patient survival is equal to the production function g.

$$\mathbb{E}[Y^*_{ijht}|\alpha_j, \psi_h, X_{it}] = g(\alpha_j, \psi_h, X_{it})$$

I seek to estimate, rather than assume, the existence and magnitude of complementarities between surgeons and hospitals. Both the direction and magnitude of complementarities are needed to quantify the potential gains and losses from alternative allocations of surgeons to hospitals. I first assume that the production function g is monotonic in α_j and ψ_h, a reasonable assumption when examining an output measure directly. This assumption does not restrict the pattern of complementarities between surgeon and hospital quality. Fix patient observables such that $X_{it} = \bar{X}$. Complementarities between surgeon and hospital quality are represented by the sign and magnitude of the cross-partial derivatives in the production function.

$$\frac{\partial^2 g(\alpha_j, \psi_h, \bar{X})}{\partial \alpha_j \partial \psi_h} \tag{1.1}$$

When equation (1.1) is positive, surgeons and hospitals are complements: the return to allocating high-α_j surgeons to high-ψ_h hospitals is larger than for low-α_j surgeons. The production function is supermodular. When equation (1.1) is negative, surgeons and hospitals are imperfect substitutes: the return to allocating low-α_j surgeons to high-ψ_h hospitals is larger than for high-α_j surgeons. The production function is submodular. Finally, when equation (1.1) is equal to zero, the contributions of surgeons and hospitals to the production function are independent.

Figure 1.3 illustrates these differences graphically: the cross-partial derivative of the production function can be evaluated as the differences in slopes across surgeons in these graphs. In Figure 1.3a, hospital and surgeon quality are assumed to be separable. This notably corresponds to production functions where α_j and ψ_h enter additively. In this case, the slopes are identical across surgeons: the return to allocating surgeons to high-ψ_h hospitals is independent of the surgeon. Figure 1.3b shows the case where surgeon and hospital quality are complements: the slope is larger for high-α_j surgeons. Consequently, the return to allocating surgeons to high-ψ_h hospitals is larger for high-α_j surgeons. Conversely, hospital and surgeon quality are assumed to be imperfect substitutes in Figure 1.3a: the slope is larger for lower-α_j surgeons, such that the return to allocating surgeons to high-ψ_h hospitals is larger for low-α_j surgeons.

Denote the observed survival of patient i treated by surgeon j in hospital h as Y_{ijht} such that

$$Y_{ijht} = D_{ijht} Y^*_{ijht}$$

where D_{ijht} is an indicator for the existence of a match with patient i treated by j in hospital h. The matrix D describes the network of patients, physicians, and hospitals. Note that observed and potential survival coincide when $D_{ijht} = 1$:

$$\begin{aligned}\mathbb{E}[Y_{ijht}|D_{ijht} = 1, \alpha_j, \psi_h, X_{it}] &= \mathbb{E}[Y^*_{ijht}|D_{ijht} = 1, \alpha_j, \psi_h, X_{it}] \\ &= g(\alpha_j, \psi_h, X_{it}) + \mathbb{E}[\epsilon_{ijht}|D_{ijht} = 1, \alpha_j, \psi_h, X_{it}] \qquad (1.2)\end{aligned}$$

There are two main challenges to recover complementarities between surgeon and hospital quality as well as the sorting of surgeons across hospitals. First, estimates for α_j and ψ_h are noisy measures for quality since operative mortality from CABG surgery is a rare event. In particular, the average 30-day mortality rate is 5% in the sample, while the mean and median number of surgeries per surgeon is 44 and 37, respectively, against 95 and 69 for hospitals. The noise in these individual estimates of provider quality is a challenge to estimate sorting, but also to recover the magnitude of complementarities between surgeon and hospital quality.[10] I address this issue by grouping surgeons and hospitals in a first step in the spirit of Bonhomme, Lamadon, and Manresa (2022). I cluster providers using their average risk-adjusted survival as a proxy for their individual quality using a k-means algorithm, as detailed in the next subsection. Using this classification, I can next recover grouped fixed effects estimates for surgeons' and hospitals' types as well as their interactions in a second step.

Second, parameters α_j and ψ_h are identified if and only if the network is exogenous, i.e.,

$$\epsilon_{ijht} \perp D_{ijht}|\alpha_j, \psi_h, X_{it}, \ \ \forall i, j, h, t$$

This assumption implies that $\mathbb{E}[\epsilon_{ijht}|D_{ijht}, \alpha_j, \psi_h, X_{it}] = \mathbb{E}[\epsilon_{ijht}|\alpha_j, \psi_h, X_{it}] = 0$ in equa-

10. Noise in the individual fixed effects also magnifies the bias on the sorting parameter, called limited mobility bias, which has been shown to be quantitatively important in the literature (Kline, Saggio, and Sølvsten, 2020; Bonhomme, Lamadon, and Manresa, 2022).

tion (1.2). This is the exogenous network assumption, common in two-way fixed-effects model as in Abowd, Kramarz, and Margolis (1999). The probability for a patient to be treated at a hospital h by a surgeon j can depend on the individual hospital and surgeon heterogeneity and patient observables, but it cannot depend on unobservables at the patient level that have an impact on their survival Y^*_{ijht}. In the case of patient outcomes, this assumption may be violated if patients select into hospitals or surgeons on unobservables. I address this concern using distances to providers as instruments to identify selection on patient unobservables. I delineate this approach in subsection 1.3.3.

1.3.2 Classifying hospitals and surgeons

I group individual surgeons and hospitals to reduce the dimensionality of the fixed effects in the two-sided heterogeneity model.[11] This dimensionality reduction addresses the noise in individual fixed effects by estimating the quality at the provider *group* level, which in turn allows me to estimate interactions in the production function and sorting. This consists in a two-step grouped fixed-effects approach similar to Bonhomme, Lamadon, and Manresa (2019, 2022), with the difference that I cluster both sides, i.e., hospitals *and* surgeons.

To group surgeons and hospitals, I first need to obtain individual provider moments that identify individual provider types. Using provider-level average risk-adjusted survival to identify the individual provider types requires a provider's average risk-adjusted survival to be increasing in the individual provider type. To see what this assumption implies, assume away patient observables for simplicity here and consider the average survival at hospital h

11. Theoretical properties of the grouped fixed-effects estimator has been established in Bonhomme and Manresa (2015); Bonhomme, Lamadon, and Manresa (2019) when assuming the unobserved heterogeneity is discrete in the underlying population and the number of types is known.

in the population as

$$\mathbb{E}[Y_{ijht}|\psi_h] = \int \underbrace{g(\alpha_j, \psi_h)}_{\text{Production function}} \underbrace{f(\alpha_j|\psi_h)}_{\text{Sorting}} d\alpha_j \tag{1.3}$$

where $g(\alpha_j, \psi_h)$ is the production function, assumed to be monotonic in individual surgeon and hospital quality, and $f(\alpha_j|\psi_h)$ describes the probability to observe surgeon j conditional on hospital h, which describes sorting of surgeons across hospitals. In the absence of sorting, individual quality is identified from the average survival: the average survival at hospital h is increasing in its individual quality ψ_h since g is monotone in ψ_h.

With sorting, $f(\alpha_j|\psi_h)$ depends on ψ_h and identification may fail with negative assortative matching. To see this, take the example of a linear and additive production function abstracting away from patient observables for simplicity such that $g(\alpha_j, \psi_h) = \alpha_j + \psi_h$. In this example, the average survival at hospital h in the population can be written as

$$\mathbb{E}[Y_{ijht}|\psi_h] = \psi_h + \mathbb{E}[\alpha_j|\psi_h]$$

With positive assortative matching, $\mathbb{E}[\alpha_j|\psi_h]$ is increasing in ψ_h, so average survival at hospital h is increasing in its quality and identifies the individual hospital type. With negative assortative matching, $\mathbb{E}[\alpha_j|\psi_h]$ is decreasing in ψ_h, which adds an "offsetting" effect to the individual quality of the hospital. In that case, average survival is not necessarily increasing in ψ_h and identification of the individual type from average survival may fail. Two different hospital types may exhibit the same average survival, which prevents from separating their individual types from their average survival. Consequently, identification of the individual types from individual moments requires that these individual moments are increasing in the provider's type. This may not always be true under negative assortative matching. I will use several alternative moments to evaluate the robustness of the results to this assumption.

The goal of this classification is to cluster hospitals and surgeons into groups capturing their individual quality. Average risk-adjusted survival from CABG surgery is used as a measure of quality in practice in report-cards for surgeons and hospitals. It is publicly reported and used in CMS quality ratings for hospitals, for example. It is also used in the literature evaluating hospital and surgeon quality (Huckman and Pisano, 2006; Ghaferi, Birkmeyer, and Dimick, 2009; Kolstad, 2013). I use provider-level risk-adjusted survival as a proxy for individual quality to group surgeons and hospitals into quality groups.

I follow the methodology used in the literature to compute risk-adjusted survival. In particular, the predicted probability of survival for each patient is estimated using a logit model:

$$\ln\left(\frac{Pr[Y_{ijht}=1|X_{it}]}{1-Pr[Y_{ijht}=1|X_{it}]}\right)=\alpha+\beta X_{it}$$

The fitted values are used to form the expected survival rate (ESR) at the hospital or surgeon level. I then obtain the average risk adjusted survival rate (RASR) for a hospital as

$$RASR_h=\left(\frac{OSR_h}{ESR_h}\right)\times OSR$$

where OSR_h is the average observed survival rate of patients treated at hospital h, ESR_h is the average expected survival rate of patients treated at hospital h obtained from the logit model, and OSR is the national average survival rate.

I group surgeons and hospitals using k-means clustering on the computed average risk-adjusted survival as a proxy for individual quality. The groups should capture the underlying heterogeneity in quality across individual providers. K-means is well-suited for this purpose, as it creates the groupings by maximizing the distance in the average moments across groups, and minimizes the distance in the individual average moments within groups, using the euclidian distance. The number of groups needs to be specified by the researcher ex-ante. I will examine several alternative number of groups for both surgeons and hospitals. More details on the k-means algorithm are reported in Appendix 1.9.5. I will show that the variance in survival across k-means groups indeed represents a substantial fraction, over 80%, of the variance in survival across individual providers. I will also examine alternative grouping methods, such as simple quintiles of risk-adjusted survival.

When the production function is monotonic and positive assortative matching exists, I can accurately recover individual surgeon and hospital types from the grouped fixed effects. As shown in Monte Carlo exercises in Appendix 1.9.6, the correlation between true and estimated surgeons' and hospitals' grouped fixed effects is over 0.9. With negative assortative matching, surgeons and hospitals are misclassified, which biases against finding any sorting of surgeons across hospitals. Increasing the number of groups partially alleviates misclassification and allows us to recover the direction of sorting.[12]

Grouping surgeons and hospitals also addresses biases in the variance and covariance estimates in two-way fixed-effects models (Bonhomme, Lamadon, and Manresa, 2019, 2022). This bias, called limited mobility bias (Abowd et al., 2004; Andrews et al., 2008, 2012), is related to the weak connectivity of the network (Jochmans and Weidner, 2019). In matched employer-employee data, the few number of movers per firm leads to overestimate the variance of the firm fixed effect and underestimate the covariance between the firm and the worker fixed effects. Recent work has shown that this bias is quantitatively important and correcting for it leads to different conclusions on the respective contributions of worker and firm heterogeneity in wage dispersion, as well as on the direction of sorting of workers into firms (Kline, Saggio, and Sølvsten, 2020; Bonhomme et al., 2023). Using a rare event as an outcome magnifies this concern and supports the use of groups to accurately recover the sorting parameter while also allowing for interactions in surgeon and hospital quality.

12. Increasing the number of groups can remedy some of the classification error but at the cost of a larger limited mobility bias. The limited mobility bias results in a large downward bias on the sorting estimates, and large upward bias on the firm variance estimates (Kline, Saggio, and Sølvsten, 2020; Bonhomme et al., 2023). Consequently, increasing the number of ex-ante groups K results in a trade-off between classification error from k-means clustering and the limited mobility bias through weaker firm network connections (Jochmans and Weidner, 2019; Bonhomme, Lamadon, and Manresa, 2022).

Using the groups recovered from the classification steps, I seek to estimate the production function of survival with functional form

$$g(\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}) = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} \tag{1.4}$$

where $\alpha_{l(j)}$ is the grouped type of surgeon j, $\psi_{k(h)}$ is the grouped type of hospital h, and $\kappa_{l(j)k(h)}$ are interactions between surgeon and hospital grouped types in the production function. This production function is non-parametric in the sense that the existence and magnitude of the cross-partials in the production function as in equation (1.1) are estimated directly through the interaction terms.

Note that this specification assumes away complementarities between surgeon or hospital quality and patient observables. In other words, I assume that hospital and surgeon quality have an homogenous treatment effect on patients. My estimates therefore reveal complementarities between surgeon and hospital quality on the average patient.

1.3.3 Controlling for patient selection into providers

I make two alternative assumptions on the relationship between unobserved health shocks ϵ_{ijht} and the network D_{ijht}.

Approach A: Exogenous network conditional on observables. Network formation, i.e., the formation of patient-surgeon-hospital triplets, is exogenous conditional on patients observables X_{it}, the year unobserved heterogeneity γ_t, and unobserved heterogeneity $\alpha_{l(j)}$, $\psi_{k(h)}$, $\kappa_{l(j)k(h)}$:

$$\epsilon_{ijht} \perp D_{ijht} | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, \ \ \forall i, j, h, t \tag{1.5}$$

In other words, the realization of a link D_{ijht} is independent of unobservables ϵ_{ijht} conditional on observables X_{it} and unobserved heterogeneity $\alpha_{l(j)}$, $\psi_{k(h)}$, $\kappa_{l(j)k(h)}$, γ_t. Note that this assumption implies that the probability for a patient to be treated by a specific surgeon j at a hospital h cannot depend on ϵ_{ijht}, but it allows this probability to depend on patient observables X_{it}, the surgeon and hospital unobserved heterogeneity $\alpha_{l(j)}$, $\psi_{k(h)}$, and $\kappa_{l(j)k(h)}$, and the year unobserved heterogeneity γ_t.

The network exogeneity assumption requires that patient selection into surgeons and hospitals happens on observables. I use a rich set of patient observables from the Medicare claims data that includes various demographics, including age, gender, Medicaid eligibility, income and population in the ZIP code of residence, but also health status based on diagnoses on claims in the 12 months prior the surgery. These diagnoses identify seventeen comorbidities that are aggregated in a health score, the Charlson score. Yet, if selection happens on patient unobservables, the network exogeneity assumption is violated. I relax this assumption below.

Under the network exogeneity assumption, I can recover surgeon and hospital grouped fixed effects from estimating in a second step:

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t] = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t \quad (1.6)$$

Approach B: Partially endogenous network using distance to the hospital as an excluded instrument. Despite a rich set of patient covariates, there may still be selection on patient unobservables: patients may select into providers based on private information not captured in the claims data. To identify selection on unobservables, I use the distance between the hospital and the patient ZIP codes as an excluded instrument, as used recently in Einav, Finkelstein, and Mahoney (2022) also in the context of a control function. Distance to the hospital is a strong predictor of hospital choice for CABG surgery, as reported in Subsection 1.5.3.

I partially endogenize network formation by modeling the choice of hospitals. Recall the observed survival Y_{ijht} for patient i treated by surgeon j in hospital h is

$$Y_{ijht} = D_{ijht} Y^*_{ijht}$$

but I now assume that

$$\begin{aligned} D_{ijht} &= 1\{u_{ih} \geq u_{ih'},\ \forall h'\} \\ \text{with } u_{ih} &= \delta_h - \tau \ln(d_{ih}) + \eta_{ih} \end{aligned} \tag{1.7}$$

where u_{ih} is the utility from patient i from receiving the surgery at hospital h, δ_h is the perceived quality of hospital h, on which all patients agree within a market, and d_{ih} is the distance between the patient ZIP code and the hospital ZIP code. I assume η_{ih} are type-I extreme value error terms.[13] The outside option is defined as choosing a hospital outside of the patient's hospital referral region (HRR) of residence.

Here, part of the network is endogenous, so that ϵ_{ijht} and η_{ih} are correlated. Following Dubin and McFadden (1984), I impose the following linear structure to the conditional expectation of ϵ_{ijht}:

$$\mathbb{E}[\epsilon_{ijht} | \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, X_{it}, \eta_{i1}, ..., \eta_{iH}, D_i = h] = \sum_{s \in \mathcal{H}} \phi_s (\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta) \tag{1.8}$$

where μ_η is the Euler constant (mean of logit errors), $\mathcal{H}$ the set of hospitals, and D_i indicates the chosen hospital.

The expected survival conditional on the fixed effects, patient observables X_{it}, the choice

13. Surveys reported in the medical literature indicate that the hospital is chosen by the operating surgeon and the patient jointly, with more role for surgeons for cardiovascular surgeries (Wilson, Woloshin, and Schwartz, 2007). The choice model above supports a joint decision between surgeons and patients.

of hospital D_i, and the unobserved logit shocks $\eta_{i1}, ..., \eta_{iH}$ can be written as

$$\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, X_{it}, \eta_{i1}, ..., \eta_{iH}, D_i = h] = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} + \gamma_t + \sum_{s \in \mathcal{H}} \phi_s(\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta) \tag{1.9}$$

Integrating equation (1.9) over the unobserved demand shocks $\eta_{i1}, ..., \eta_{iH}$ delivers the estimating equation

$$\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, X_{it}, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \alpha_{l(j)} + \alpha_h + \kappa_{l(j)k(h)} + \beta X_{it} + \gamma_t + \sum_{s \in \mathcal{H}} \phi_s \theta_{is}(h) + \varphi \theta_{ih}(h) \tag{1.10}$$

where $\theta_{is}(h) = \mathbb{E}[\eta_{is} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = h]$ are the control functions such that

$$\theta_{is}(h) = \begin{cases} -\ln \hat{p}_{is} & \text{if } s = h \\ \frac{\hat{p}_{is}}{1-\hat{p}_{is}} \ln \hat{p}_{is} & \text{if } s \neq h \end{cases}$$

and $\hat{p}_{is}$ is the predicted probability for patient i to choose hospital s from the demand model in equation (1.7). Derivations are included in Appendix 1.9.7. Note that the control function is positive when $s = h$ but negative otherwise since $\ln \hat{p}_{is} < 0$ with $0 < \hat{p}_{is} < 1$.

Parameter φ is choice-specific and captures Roy-type selection or selection on gains: if a patient is choosing a hospital because he is idiosyncratically more likely to improve there, then $\varphi > 0$. The intuition for identification is the following: when patients travel farther than expected for a hospital, leading to a larger η_{ih}, are more likely to survive after CABG surgery, then the probability to survive after CABG surgery and η_{ih} are positively correlated and this identifies selection on gains.

Parameters ϕ_s are hospital-specific and capture selection into specific hospitals. If sick patients select into high-quality hospitals, $\phi_s < 0$. The intuition for identification is similar as above: when patients travelling farther for a specific hospital are consistently less likely to survive after CABG surgery, then $\phi_s < 0$ and these patients must be sicker. If healthier patients select into high-quality hospitals, $\phi_s > 0$.

I estimate the demand model market by market, using hospital referral regions (HRRs) as market definitions.[14] I then construct the control functions $\hat{\theta}_{is}$ based on the estimated predicted probabilities $\hat{p}_{is}$, and recover surgeon and hospital grouped fixed effects from the following regression:

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t + \sum_{s \in \mathcal{H}} \phi_s \hat{\theta}_{is}(h) + \varphi \hat{\theta}_{ih}(h) \quad (1.11)$$

Distance to the hospital is an excluded instrument here, since it is excluded from the last step: distance to the hospital can only have an impact on patient survival through the choice of hospital. Since CABG surgery is usually performed in non-emergency settings, distance to the hospital as time to treatment should have no impact on patient survival.[15] Yet, the exclusion restriction may fail if hospital locations are endogenous to patient characteristics relevant for survival. I confirm the plausibility of this assumption using patient observables in Subsection 1.5.3.

Note that there are two potential sources of selection: selection into providers and selection into treatment. The distance between the patient and the hospital is an instrument for hospital-surgeon pairs which addresses selection into providers. Since selection into

14. Hospital referral regions (HRRs) are healthcare market definitions constructed by the Dartmouth Atlas based on where patients receive care in the U.S. Receiving CABG surgery outside of the patients' HRR of residence is defined as the outside option.

15. 77% of surgeries in the final sample are not emergent since patients have no emergency room expenses in their hospital stay. Even in the case of emergencies, CABG surgery is performed on stable patients for whom distance to the hospital was probably not critical. I investigate robustness of results by excluding patients with emergency-room expenses, and find similar results.

treatment is probably limited in this setting, I do not address it directly. As detailed in Appendix 1.9.1, treatment decisions are likely to be made prior to the referral to cardiac surgeons, and alternative treatments are performed by distinct types of physicians. CABG surgery is also usually performed in non-emergency settings. Therefore, there exists limited scope for selection into treatment by surgeons or hospitals.

1.4 Estimated parameters

I summarize the estimated parameters from the empirical model. I first show that there remains substantial variation in risk-adjusted survival across providers after classifying them into grouped types using k-means clustering. I then investigate the sensibility of recovered parameters. Risk-adjustment coefficients for patient observables are sensible and statistically significant. Surgeon and hospital estimated effects are correlated with external measures of quality. Finally, I find limited evidence for selection of patients into providers using patient observables, suggesting that patient selection into providers may not play a major role for my parameters of interest.

1.4.1 Grouped types of surgeons and hospitals

I group surgeons and hospitals using k-means clustering on average risk-adjusted survival as delineated in Section 1.3.2. I impose five distinct groups for both hospitals and surgeons, which is the greatest symmetric number of groups that allows me to observe patients for each hospital-surgeon group interaction. I show that results are robust to alternative number of groups in Subsection 1.5.3.

There exist a substantial amount of variation in survival across groups. Figure 1.4 shows the difference between average observed and average predicted survival across hospital and surgeons groups. The maximum difference is about 12 percentage points across hospital groups, and about 20 percentage points across surgeon groups. Overall, the variance across groups represents 84% of the variance in 30-day survival across individual providers, for both surgeons and hospitals. Note that the ordering of the groups displayed in Figure 1.4 does not come from k-means clustering: the classification step only clusters hospitals and surgeons into groups, but does not impose any meaningful ordering on them. Groups are also of varying sizes. Importantly, note that the average risk-adjusted survival for hospitals and surgeons still includes the combination of the hospital and surgeon effects in Figure 1.4.

Since the relevant source of variation comes from surgeons operating across multiple hospitals, clustering hospitals into groups has a cost since surgeons now have to operate across multiple hospital *groups*. While more than a third of surgeons were observed at more than one hospital in Figure 1.1, 30% of surgeons are observed at more than one hospital *group*, as depicted in Appendix Figure 1.15. Yet, the number of surgeons observed at multiple hospital groups remains large.

1.4.2 Estimated parameters are sensible

Risk-adjustment parameters. The coefficients on patient observables for risk adjustment are sensible. Table 1.3 reports coefficients on patient covariates from estimating equations (1.6) and (1.11). In both specifications, older and sicker patients are less likely to survive 30 days after surgery. Women are also less likely to survive after CABG surgery, consistent with several studies in the medical literature (Zwischenberger, Jawitz, and Lawton, 2021).

Correlations with external measures of quality. Estimated effects for surgeon and hospital groups are correlated with external measures of quality of these providers. I find evidence that better surgeons tend to have more experience in the procedure during the time period, which is in line with the importance of learning-by-doing to determine physicians' skills outlined in the literature (Birkmeyer et al., 2013; Currie and MacLeod, 2017). As reported in Figure 1.5, higher surgeon-group estimates are positively correlated with a surgeon's recent experience in CABG procedures, measured in frequency and in revenue in Medicare between 2012 and 2017. Surgeon group estimates are also positively correlated with a surgeon's recent experience in surgical procedures overall, but these relationships are not statistically significant.[16] There is no relationship with a surgeons' tenured experience measured as the number of years since the surgeon graduated from medical school, which is consistent with previous evidence from Birkmeyer et al. (2013) notably. Note that these surgeon covariates only explain a small fraction of the variation across fixed effects: the R^2 of the regression including all covariates remains below 0.01, as shown in Appendix Table 1.14.

I also find that higher hospital-group effects are positively correlated with external measures of quality, such as CMS ratings and CMS 30-day risk-adjusted survival for six conditions and procedures as reported in Figure 1.6. Larger estimated hospital-group effects are positively correlated with CMS five-star ratings, but this relationship is not statistically significant. Larger estimated hospital-group effects are negatively correlated with CMS 30-day risk-adjusted mortality measures, especially for heart-related conditions. Note that these CMS measures include both surgeons and hospital effects: this is why the relationship between the estimated hospital effects from the empirical model described in equation (1.6) and the CMS risk-adjusted mortality measure for CABG surgery is highly but not perfectly correlated. The R^2 when including all CMS quality measures amounts to about 0.05 for both specifications. I investigate correlations with other hospital-level characteritics in Appendix Table 1.15. None of these relationships are statistically different from zero, and the R^2 when including all covariates available for at least 1,000 hospitals is below 0.01 for both specifications.

16. Correlations using yearly experience are similar; the correlation in these surgeons' activity within Medicare year to year is above 0.8.

1.4.3 Limited evidence of patient selection using observables

To gain insights into the existence of patient selection into providers for CABG surgery, I examine the relationship between patient observables and the ranking of their provider groups. I find no evidence of systematic adverse selection of patients into higher-survival providers using patient observables. I use the predicted survival for patients net of provider effects, i.e., the predicted survival only driven by patient covariates, to examine systematic relationships with the ranking of a surgeon or hospital group. As shown in Appendix Table 1.16, there is no systematic relationship between predicted survival based on patient covariates and their provider rankings. If anything, higher-survival surgeons tend to operate on slightly healthier patients, but this is not statistically significant. Investigating each group separately in Panel of Appendix Figure 1.20, I find potential evidence of adverse selection into the top hospital or surgeon group using patient observables. The lowest provider groups appear to treat observably sicker patients, which reflects their location as shown in Subsection 1.5.2.

Since surgeons tend to multi-home, they may be able to "triage" their patients across hospitals, taking their sickest patients into their best available hospitals. However, as shown in Figure 1.7 and Appendix Table 1.17, I find no systematic relationship between patient covariates or predicted survival based on patient covariates and hospital rankings within surgeons. The evidence is similar for hospitals: there is no evidence of adverse selection into higher-survival surgeons within hospitals as shown in Appendix Figure 1.16 and Appendix Table 1.17. If anything, there appear to be weak evidence of advantageous selection of patients into higher-survival surgeons within hospitals. This is more consistent with surgeons taking their own patients to the hospital rather than hospitals assigning patients to surgeons.[17] Overall, patient selection into higher-survival providers appears to be minimal

17. This is indeed consistent with the medical literature. Surveys suggest that the surgeon is the main driver of that decision, especially in the case of cardiovascular surgery (Wilson, Woloshin, and Schwartz, 2007). Since patients are referred to a cardiothoracic surgeons prior to surgery, surgeons and patients may choose at which hospital to perform the surgery jointly.

in this setting.

1.5 Imperfect substitutability and sorting results

I find that surgeon and hospital quality are imperfect substitutes in the production function of survival for CABG surgery: the return to allocating low-survival surgeons to high-survival hospitals is greater than for high-survival surgeons. This finding is consistent with a mechanism of "failure-to-rescue" emphasized in the medical literature. I find evidence of positive assortative matching, where high-survival surgeons sort into high-survival hospitals. The current positive assortative matching of surgeons across hospitals does not maximize aggregate survival: negative assortative matching would increase aggregate survival while reducing dispersion in survival across patients. I show the robustness of these results to allowing for selection on unobservables, to alternative number of groups in the classification, alternative classifications, and alternative samples.

1.5.1 Surgeon and hospital quality are imperfect substitutes

To investigate the differential returns to allocating high- and low-survival surgeons to alternative hospitals, I report the average predicted 30-day survival for each hospital-surgeon group interaction separately in Figure 1.8a, similarly to Figure 1.3 in Section 1.3.1. To rank surgeon groups, I calculate the predicted 30-day risk-adjusted survival for each group, assuming each interaction with a hospital group is equally likely. Similarly, I rank hospital groups using the predicted 30-day risk-adjusted survival for each group, assuming each interaction with a surgeon group is equally likely. The differential returns between high- and low-survival surgeon groups to being allocated to higher-survival hospital groups can be inferred directly from the differentials in slope of the average predicted survival across hospital groups for each surgeon group. Equal slopes across surgeon groups would suggest that surgeon and hospital quality do not depend on each other. A larger slope for higher-survival surgeon groups would suggest complementarities between surgeon and hospital quality. Conversely, a larger slope for lower-survival surgeon groups would suggest imperfect substitutability between surgeon and hospital quality.

Surgeon and hospital quality are imperfect substitutes in the production function of 30-day survival for CABG surgery. As indicated in Figure 1.8a, the predicted survival gains from allocating surgeons to higher-survival hospital groups are larger for low-survival surgeon groups. I estimate the slope for each surgeon group in Table 1.4. Lower-survival surgeons exhibit a larger slope than high-survival surgeons in both specifications, and these differences are statistically significant. These suggest that the magnitude of the imperfect substitutability between surgeon and hospital quality may be quantitatively large. The production function of 30-day survival for CABG is submodular: the cross-derivative in surgeon and hospital quality is negative.[18]

The dispersion in predicted risk-adjusted survival across surgeon and hospital groups is large. The standard deviation in surgeon types' effects amounts to 3.1 percentage points of 30-day survival. It is comparable to the standard deviation in the predicted survival based on patients' observables, which amounts to 2.9 percentage points of 30-day survival. The standard deviation in hospital types' effects is smaller but still large, amounting to 1.9 percentage points of 30-day survival.

Surgeons are key contributors to the variance in patient outcomes, but the value-added of hospitals plays a crucial role for low-survival surgeons. High-survival surgeons are the primary drivers of their patient outcomes: their patients tend to exhibit high survival rates irrespective of the quality of the hospital they operate at. Among low-survival surgeons, the quality of the hospital plays a crucial role in determining patient outcomes: the predicted

18. In terms of functional form specification, this result rejects production functions for 30-day survival where hospital and surgeon quality are additive, such as $g(\alpha_{l(j)}, \psi_{k(h)}) = \alpha_{l(j)} + \psi_{k(h)}$. However, it is consistent with a logit production function for 30-day survival with additive hospital and surgeon group fixed effects. The logit is submodular for probabilities above 0.5, and average 30-day survival probability is 0.95 for CABG.

survival for their patients varies widely with the quality of the hospital they operate at. These facts are illustrated in Figure 1.9. Overall, high-survival surgeons tend to achieve high survival rates at all hospitals, and high-survival hospitals tend to achieve high survival rates no matter what surgeon type is operating on patients.

A model without interactions between surgeon and hospital groups fails to uncover the crucial role of hospital quality for low-survival surgeons. To see this, I estimate a model without interactions such that

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \gamma_t] = \alpha_{l(j)} + \psi_{k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t \tag{1.12}$$

and I decompose the explained variance in 30-day survival net of covariates such that

$$Var(Y_{ijht} - \sum_p \hat{\beta}_p X_{it,p} - \hat{\gamma}_t - \hat{\epsilon}_{ijht}) = Var(\hat{\alpha}_{l(j)}) + Var(\hat{\psi}_{k(h)}) + 2 \times cov(\hat{\alpha}_{l(j)}, \hat{\psi}_{k(h)}) \tag{1.13}$$

Table 1.5 shows results from this decomposition, expressing relative contributions as percentages of the explained variance in 30-day survival net of covariates $Var(Y_{ijht} - \sum_p \hat{\beta}_p X_{it,p} - \hat{\gamma}_t - \hat{\epsilon}_{ijht})$. The term $Var(\hat{\alpha}_{l(j)})$ captures the contribution of surgeon groups, term $Var(\hat{\psi}_{k(h)})$ captures the contribution of hospital groups, and $cov(\hat{\alpha}_{l(j)}, \hat{\psi}_{k(h)})$ captures the direction and contribution of sorting.

The contribution of hospital groups amounts to about 10% of the explained variance in 30-day survival net of covariates. The contribution of surgeon groups is much larger than for hospitals in this decomposition, representing up to six times the contribution of hospitals. The contribution of surgeons to the variance in patient outcomes is also similar to the contribution of patients observables, as shown in Appendix Table 1.18. However, this leads to underestimate the contribution of hospital quality for low-survival surgeons.

Mechanism: failure-to-rescue. While imperfect substitutability between surgeons and hospitals in terms of patient survival may appear surprising at first glance, it seems consistent with a mechanism highlighted in the medical literature: failure-to-rescue. "Failure-to-rescue"–defined as the probability of death given complications–describes the inability of a hospital to save patients from complications. This term was coined by Silber et al. (1992) who showed that hospital-level complication measures tended to be less sensitive to hospital characteristics than mortality measures. They found that failure-to-rescue measures were highly correlated with both hospital-level mortality measures and hospital characteristics, suggesting that low-mortality hospitals tend to achieve low-mortality outcomes through their ability to rescue patients from complications.

Ghaferi, Birkmeyer, and Dimick (2009) extend these findings to six high-risk surgical procedures in the entire Medicare population. In the case of CABG, they find a complication rate of 24.2% for high risk-adjusted mortality hospitals versus 21.1% for low risk-adjusted mortality hospitals. However, failure-to-rescue was 18.9% at high risk-adjusted mortality hospitals versus 6.2% at low risk-adjusted mortality hospitals: a three-fold difference. Post-operative complications need to be noticed quickly, and handled both correctly and rapidly. Hospitals with greater ability to rescue patients from complications have been shown to have better nurse and physician staffing and better communication processes (Ghaferi, Birkmeyer, and Dimick, 2009; Johnston et al., 2015; Ward et al., 2019).

In the context of a joint production function between surgeons and hospitals, high-survival surgeons may be able to prevent a larger fraction of complications or make complications less severe (Birkmeyer et al., 2013). Therefore, no matter at which hospital they perform surgery, these surgeons' patients recover normally and survive, with little role for the hospital. However, low-survival surgeons may not be able to prevent as many complications: the hospital at which they perform the surgery becomes crucial for patient survival, since it is the hospital that will handle post-operative complications. In the context of CABG surgery, hospitals potentially have a large role to play, since patients stay on average 10 days in the hospital.

Whether results for CABG surgery are generalizable to other surgeries, in particular in terms of imperfect substitutability between surgeons and hospitals, remains to be investigated. Note that the failure-to-rescue mechanism has been shown to arise for several other procedures (Ghaferi, Birkmeyer, and Dimick, 2009). Additionally, most common surgical procedures are similar in processes to CABG surgery, with one surgeon in charge of performing the surgery along with a surgical team and the hospital taking charge of pre- and post-operative care. Hip and knee arthroplasties or heart-valve replacements, which also appear in the top 10 operating room procedures in aggregate cost, are such examples (McDermott and Liang, 2021). However, more novel or frontier surgeries may exhibit complementarities, especially when a team of surgeons is involved in the surgery.

1.5.2 High-survival surgeons sort into high-survival hospitals

I find positive assortative matching where high-survival surgeons sort into high-survival hospitals. Figure 1.8b describes the share of surgeries at a hospital group performed by each surgeon group, where groups are described by their relative rankings. Surgeries at high-survival hospitals are performed by high-survival surgeons, while surgeries at low-survival hospitals are performed by low-survival surgeons. The positive assortative matching appears to be relatively strong, with estimated correlations ranging between 0.3 and 0.5 across specifications. Variance decompositions in Table 1.5 suggest that sorting explains between 20 and 25% of the explained variance in survival net of covariates.

Positive assortative matching at the national level reflects provider location decisions across space. High-survival hospitals and surgeons tend to co-locate in space into larger and higher-income regions. Examining the correlation between estimated provider rankings and patient covariates, I find that high-survival providers tend to be located in more populated and higher-income locations. As reported in Figure 1.10, higher-survival hospital and surgeon groups tend to treat older patients living in highly populated high income ZIP codes. This is driven by the location of surgeons and hospitals rather than selection of patients across providers: these statistically significant relationships disappear when controlling for the hospital's or surgeon's HRR, as reported in Figure 1.11. This is consistent with Dingel et al. (2023), which shows that larger cities tend to produce higher-quality medical services, notably through division of labor.

However, positive assortative matching at the national level is not entirely driven by providers co-locating across space, since positive assortative matching is also substantial even *within* regions. I compute the correlation between the estimated surgeon and hospital group effects for the subset of patients treated in each specific HRR and report the distribution of these correlations across HRRs in Figure 1.12. I find that a substantial fraction of HRRs exhibit strong positive assortative matching, suggesting that surgeon sorting is substantial within regions too.

The current sorting of surgeons across hospitals does not maximize aggregate 30-day survival after CABG surgery. There exist larger returns from allocating low-survival surgeons to high-survival hospitals compared to lower-survival surgeons. Yet, high-survival surgeons sort into high-survival hospitals. This suggests that we could increase 30-day survival after CABG surgery by reallocating low-survival surgeons to high-survival hospitals. I quantify the impact of surgeon sorting across hospitals on aggregate patient outcomes using partial equilibrium counterfactual reallocations in the next section.

1.5.3 Robustness

Addressing selection on unobservables: distance to hospitals as an excluded instrument. Since patients may select into surgeon-hospital pairs on characteristics that I do not observe in the claims data, I control for selection on unobservables using a control function approach as delineated in Section 1.3.3 as approach B. I use the distance between the patient and the hospital ZIP codes as an instrument to identify patient demand. Patients tend to be treated at hospitals close to their ZIP code of residence, as depicted in Appendix Figure 1.17. 21% of surgeries are performed outside of the patient's HRR.[19] The relationship between the chosen hospital and distance also appears log-linear, supporting the functional form assumption in equation (1.7).

I examine whether the distance to the hospital is predictive of hospital choice in Panel 1.13a of Figure 1.13. This figure depicts the relationship between the predicted probabilities to choose a hospital from the hospital choice model depicted in equation (1.7) estimated for each HRR separately and the distance between the patient ZIP code of residence and the hospital ZIP code. Appendix Figure 1.18 reports these predicted probabilities for two specific HRRs: Boston and Chicago. The probability to choose a hospital within an HRR declines sharply with distance to the hospital: the first stage of the distance instrument is strong. For this reason, it is an instrument commonly used to model healthcare provider choice (for example Einav, Finkelstein, and Williams (2016); Card, Fenizia, and Silver (2023); Einav, Finkelstein, and Mahoney (2022)).

The key identifying assumption relies on the exclusion restriction: the distance to the hospital should only have an impact on patient survival through the choice of hospital. Usual balance tests–examining covariates balance with the instrument–cannot be straightforwardly performed here since the instrument is the distance between the patient and every hospital in her choice set. To evaluate the plausibility of the exclusion restriction assumption, I perform two different exercises. First, I evaluate the stability of the relationship between 30-day survival and distance to the chosen hospital with and without patient observables. As shown in Appendix Table 1.19, the coefficient on the logarithm of distance is relatively

19. For the average HRR, 26% of surgeries are performed outside of it. There is substantial variation across HRRs, with patients from more populous HRRs tending to remain in their HRR to receive CABG surgery. It is 6% in Boston, MA versus 58% in Altoona, PA for example (Dingel et al., 2023).

stable with and without covariates, lending support for the exclusion restriction.

Second, I examine the stability of distance parameters in the demand model described by equation (1.7) when allowing δ_h, the perceived quality from hospital h, to depend on patients observables $\delta_h(X_i)$ as in Einav, Finkelstein, and Mahoney (2022). If the estimated distance parameter $\hat{\tau}$ does not vary with the inclusion of patient covariates, allowing the perceived quality of hospital h to depend on patient covariates does not change the impact of distance on patient utility, which suggests that distance only impacts the choice of hospital. Panel 1.13b of Figure 1.13 compares the estimated demand parameters for the logarithm of distance without patient observables to including patient age, ZIP code income per capita, and Charlson score in δ_h. The parameters are almost identical between the two specifications, with a correlation above 0.99.[20]

Estimated parameters of the control function are consistent with the expected direction of patient selection. The coefficients on patient observables for risk adjustment are sensible, as shown in the second column of Table 1.3. The estimated parameter for selection on gains $\hat{\psi}$ is positive, even though it remains non-statistically significant. Parameters $\hat{\phi}_s$ capturing selection into specific hospitals are consistent with sicker patients selecting into better hospitals in some cases, but also with healthier patients selecting into better hospital in other cases, as reported in Appendix Figure 1.19. While the former is consistent with the idea of surgeons triaging their sicker patients into better hospitals, the latter is consistent with healthier, and correlatedly wealthier, patients being able to sort into better hospitals because they may have a lower distance elasticity or may be better informed (Dingel et al., 2023) .

Control function parameters suggest adverse selection into higher-survival providers. I examine the relationship between patient observables when including control function arguments and provider rankings in Appendix Figure 1.20. While there is no systematic

20. Including all patient observables only allows to estimate demand for 263 out of 305 HRRs because of collinearity issues in patient observables at the option level. The distance parameters for the 263 HRRs when including no versus all patient observables in δ_h are also very similar with a correlation over 0.85.

relationship between predicted survival net of provider fixed effects and provider rankings when including only observables, adding the control function arguments suggests a negative relationship between predicted survival net of provider fixed effects and provider rankings, which is statistically different from zero. While there is evidence of adverse selection into both surgeon and hospital groups, adverse selection into higher-survival providers appears to be stronger for surgeons.

Results are robust when using the control function approach. As shown in Figure 1.14, surgeon groups and hospitals groups are imperfect substitutes in the production function of survival. Estimated slopes in column (2) of Table 1.7 for each surgeon group are similar to the selection on observables approach. Positive assortative matching is also strong, with a similar correlation between hospital and surgeon group effects of 0.47. The variance decomposition of a model without interactions between surgeon and hospital groups also leads to similar conclusions. As reported in Table 1.5, the relative contribution of surgeon groups compared to hospital groups would be even larger. Overall, the main results are not altered by allowing for patient selection into providers on unobservables.

Alternative production function. I also examine the robustness of the variance decomposition to assuming an alternative production function. I use a logit production function, which matches the imperfect substitutability finding, where surgeon and hospital group quality enter additively such that[21]

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}] = \frac{\exp\Big(\alpha_{l(j)} + \psi_{k(h)} + \sum_s \beta_s X_{it,s}\Big)}{1 + \exp\Big(\alpha_{l(j)} + \psi_{k(h)} + \sum_p \beta_p X_{it,p}\Big)} \qquad (1.14)$$

Since the predicted log odds of survival are linear in the hospital and surgeon group fixed

21. Note that hospital and surgeon group quality are imperfect substitutes in this production function, since average 30-day survival is 0.95 which corresponds to the submodular part of the logit function.

effects and patient covariates, I decompose the variance as

$$Var\left(\ln\left(\frac{\hat{p}_{ijht}}{1-\hat{p}_{ijht}}\right)-\sum_{p}\hat{\beta}_p X_{it,p}\right) = Var(\hat{\alpha}_{l(j)})+Var(\hat{\psi}_{k(h)})+2\times cov(\hat{\psi}_{k(h)},\hat{\alpha}_{l(j)}) \quad (1.15)$$

where $\hat{p}_{ijht}$ corresponds to the predicted 30-day survival from the estimated logit model.

Results are extremely similar with the logit production function, as shown in Panel B of Table 1.5. The contribution of surgeons to the variance in predicted log-odds is more than six times larger than the contribution of hospitals. Sorting is positive and strong, with a correlation of 0.5, and represents about 25% of the variance in predicted log-odds.

Alternative number of groups. I investigate the robustness of results when varying the number of surgeon and hospital groups in Panel C of Table 1.5. Variance decomposition results remain similar when increasing the number of groups for hospitals, surgeons, or both. The contribution of surgeon groups remains largely greater than hospital groups contribution, and the sorting is positive and of similar magnitude. The imperfect substitutability result is also robust to these alternative number of groups, as delineated in Appendix Table 1.20.

Appendix Figure 1.21 shows the results from the variance decomposition of the predicted log-odds of 30-day survival, i.e., using a logit model, when varying the number of hospital groups holding the number of surgeon groups fixed, when varying the number of surgeon groups holding the number of hospital groups fixed, and when varying the number of surgeon and hospital groups jointly. Results remain relatively stable across all alternative number of groups.

Alternative classifications. I examine the robustness of results to using alternative classifications in Panel D of Table 1.5. First, I add conditional moments to the k-means clustering for hospitals. In particular, I group hospitals using eight conditional moments: average 30-day risk-adjusted survival for patients above and below the median Charlson score, above and below the median ZIP code of residence income per capita and population, and for males and females. Second, I use simple quintiles of risk-adjusted survival for hospitals and surgeons where each quintile contains the same number of surgeries. Third, I explore results when adding provider-level covariates to the k-means algorithm. For hospitals, I add the size of the hospital, captured by the number of beds, and the median income per capita and population in the hospital's ZIP code to risk-adjusted survival. For surgeons, I add surgeon's experience using the cumulated 2011-2017 Medicare activity, the Medicare surgical activity, and the CABG activity, measured in total reimbursement. I also include the median income per capita and population in the surgeon's average primary practice ZIP code.

Results remain robust to these alternative classifications. The contribution of surgeon groups remains larger than hospitals, except when adding additional covariates to k-means on the surgeon side only, where their relative contribution becomes similar. Sorting remains positive across classifications, with a correlation between 0.31 and 0.54. Column (3) of Table 1.7 reports the slopes across surgeon groups for the specification adding provider covariates to the k-means algorithm. Surgeon and hospital quality are imperfect substitutes with this alternative classification: the slope is decreasing in the surgeon's rank, even though it is not perfectly monotonic as for the baseline specification.

Alternative sample. Finally, I investigate robustness of results when excluding CABG surgery performed in an emergency setting in Panel E of Table 1.5. Emergency CABG surgery may be different from elective CABG surgery. Emergency CABG surgeries are performed by a potentially different set of surgeons who are employed at the hospital. In addition, selection may be different for emergency CABG surgery compared to elective CABG surgery; there may be differences in selection into treatment depending on hospitals' comparative advantages or the stability of the patient when reaching the ER. Consequently, I test for the stability of my results when focusing on elective CABG surgery. I do so by excluding surgeries associated with a non-zero emergency department expense in the hospital stay. This excludes about 23% of observations in the main sample.

Results are robust to excluding emergency CABG surgery. As shown in Panel E of Table 1.5, the relative contribution of surgeon groups compared to hospital groups is similar, and the sorting is positive and of similar magnitude. Column (4) of Table 1.7 also indicates imperfect substitutability between surgeon and hospital quality.

1.6 Counterfactual allocations of surgeons to hospitals

The existence of strong positive assortative matching when the production function is submodular, i.e., when surgeons and hospitals are imperfect substitutes, suggests that the current allocation of surgeons to hospitals is worse than random. The strength of the imperfect substitutability determines the impact of alternative allocations of surgeons to hospitals on aggregate patient survival. Using partial equilibrium counterfactual exercises, I show that surgeon sorting has a large impact on aggregate patient survival from CABG surgery. In particular, reallocating low-survival surgeons to high-survival hospitals decreases average mortality by 20%. I next investigate how much of the reduction in CABG 30-day mortality can be achieved by reallocating surgeons across hospital types within regions. I find that only reallocating surgeons across hospitals within HRRs achieves more than 50% of the benefits from a national reallocation.

1.6.1 Surgeon sorting has a large impact on aggregate survival

To examine the impact of surgeon sorting on aggregate patient survival, I simulate aggregate 30-day mortality from CABG surgery under two alternative allocations of surgeons: random sorting of surgeons to hospital groups and negative assortative matching. The main goal of this exercise is to evaluate the strength of the imperfect substitutability between surgeons and hospitals in the production function. It does not intend to give an exact estimate of the impact of a particular policy reallocating surgeons across hospitals, to solve for the optimal policy, nor to capture welfare. This exercise assumes away general equilibrium effects: I assume away spillover effects or learning from coworkers, for example. Surgeon and hospital spatial locations are also assumed to be fixed. I only focus on aggregate patient survival, which is only one of the many dimensions of welfare.

I first focus on reallocations where surgeons are reallocated to hospitals nationally, irrespective of the patient's or surgeon's initial location. I reallocate patients randomly to surgeons based on the national number of surgeries available per surgeon group. For the random reallocation, I next randomly assign patient-surgeon pairs to hospital groups, also taking into account the national number of surgeries available per hospital group. For the negative assortative matching reallocation, I allocate patients treated by the lowest-survival surgeon group to the highest-survival hospital group until all surgeries available at this hospital group are taken or until all surgeries performed by the lowest-surgeon group are taken, and I move to the next group. I do so until all surgeries are assigned to a surgeon and hospital group. Across all simulations, the total number of surgeries performed by each group of surgeon and each group of hospital are identical, but the number of surgeries performed by each surgeon-hospital group pair differs. I next predict 30-day mortality using estimated parameters from equation (1.6) using the new assigned surgeon and hospital groups.

As reported in Table 1.8, randomly reallocating patient-surgeon pairs to hospital types nationally decreases average 30-day mortality by about 3 deaths per thousand patients, corresponding to a 6% decrease in 30-day mortality compared to the current positive sorting. It also reduces the dispersion in 30-day mortality across patients by 7%. Consistent with the existence of imperfect substitutability between surgeons and hospitals, moving away from positive assortative matching is beneficial for patients in terms of 30-day mortality.

The magnitude of these changes is large. To put these numbers in perspective, assuming that 80,000 patients undergo CABG surgery every year within Medicare, this corresponds to about 240 lives saved per year. These gains in terms of lower 30-day mortality for CABG surgery would be even larger when taking into account non-Medicare patients. Once again, these numbers should be taken with caution, since they assume away general equilibrium effects. However, they indicate that the imperfect substitutability of surgeons and hospitals in the production function of survival for CABG surgery is quantitatively significant.

Implementing the negative assortative matching allocation leads to a reduction in average 30-day mortality that is more than three times larger than that of the random reallocation. Simulation results suggest that about 10 deaths per thousand patients would be averted every year, corresponding to a 20% decrease in 30-day mortality compared to the current positive sorting. Furthermore, the dispersion in 30-day mortality across patients would be reduced by 29% compared to the current sorting. Assuming 80,000 Medicare patients undergo CABG surgery every year, this corresponds to about 800 lives saved per year within Medicare.

There are two main takeaways from this simple reallocation exercise. First, the production function of 30-day survival for CABG surgery exhibits a strong imperfect substitutability between surgeon and hospital quality. This indicates that the allocation of surgeons to hospitals has large consequences for aggregate patient survival and its dispersion. This also suggests that the current allocation of surgeons to hospitals, exhibiting positive assortative matching, leads to a large loss in terms of patient lives. While this exercise cannot give exact estimates of the impact of alternative policies that would reallocate surgeons across hospitals, it indicates that gains from such policies may be large and beneficial to patients, and may consequently be fruitful avenues for policy.

Second, these results emphasize the importance of examining the existence of complementarity or substitutability when evaluating sorting of workers to firms. Traditional two-way fixed-effects models used to assume the shape of the production function, in particular assuming complementarities in level and substitutability in logs in the case of worker-employee match on wages (Abowd, Kramarz, and Margolis, 1999; Card, Heining, and Kline, 2013). Recent work has emphasized the importance of the production function, and in particular the existence of complementarities, to evaluate sorting patterns (Bonhomme, Lamadon, and

Manresa, 2019; Adhvaryu et al., 2020).[22] Interestingly, my results are consistent with results from Bonhomme, Lamadon, and Manresa (2019), who also find evidence for imperfect substitutability in the presence of strong positive assortative matching for worker sorting across firms on wages. In their context, the imperfect substitutability between workers and firms did not appear to be quantitatively large. In my setting, I find that it is, on the contrary, quantitatively significant.

1.6.2 Do we have to move surgeons across space?

Do high-survival surgeons and hospitals sort into larger cities? Is the national sorting of surgeons across hospitals due to the spatial sorting of providers? Examining the spatial distribution of doctors across the U.S., a large literature has now shown that more specialized doctors locate in larger cities (Newhouse et al., 1982*a*,*c*; Baumgardner, 1988*a*; Rosenthal, Zaslavsky, and Newhouse, 2005*a*; Dingel et al., 2023). However, as indicated in Section 1.5, most HRRs exhibit positive assortative matching of surgeons across hospital groups. This suggests that not all of the variation in provider types is due to sorting across space, at least within the cardiac surgery specialty. To further investigate this question, I compare results from the national reallocation exercises to reallocation exercises *within* HRRs.

In this exercise, I reallocate surgeon types operating in an HRR to alternative hospitals within the same HRR. Patients will be treated in the same HRR, and surgeon types will operate in the same HRR as in the data. Patients are allocated to surgeon types based on the number of surgeries performed by this type of surgeon at hospitals within the same HRR. Next, patient-surgeon pairs are similarly allocated to hospital types based on the number of surgeries performed by this type of hospital within the same HRR.

22. Adhvaryu et al. (2020) investigate the sorting of workers to managers within a firm using a direct measure of output. They find negative assortative matching of workers to managers while they cannot reject complementarities between the productivity of workers and managers in the production function. Also using direct measures of output, Metcalfe, Sollaci, and Syverson (2023) find evidence of negative assortative matching within the firm, where better managers tend to be allocated to worse-performing retail stores.

Results reported in Table 1.8 indicate that a random reallocation within HRRs produces a national correlation between surgeon and hospital type fixed effect of about 0.20. The difference between the correlation when allowing for a national reallocation, amounting to zero, and 0.20 suggests that there exists some non-negligible sorting of providers across space. However, only reallocating surgeons to hospital types within HRRs reduced the correlation from 0.48 in the current allocation to 0.20, confirming that not all of the variation is due to sorting of providers across space.

Randomly reallocating surgeons within HRRs leads to a decrease of about 1.7 deaths per thousand patients, representing 56% of the mortality gains from a national reallocation. In the case of negative assortative matching, the reallocation leads to about 5 deaths per thousand patients, representing 54% of the gains from a national reallocation. Reductions in the standard deviation follow similar patterns, with between 53% and 69% of the reduction in inequality from a national reallocation being achieved by reallocations within HRRs.

This exercise has interesting consequences for healthcare policy. Reallocating surgeons across regions is likely to offer very different trade-offs and costs than reallocating surgeons within regions. Dingel et al. (2023) emphasize proximity-concentration trade-offs when reallocating medical services production across space. Relocating doctors closer to patients in rural areas decreases travel costs for these patients but foregoes the benefits from region-level economies of scale. Similar trade-offs arise here in the context of cardiac surgeons. However, more than 50% of the impact on patient survival is unrelated to provider location decisions across space, in the case of CABG surgery. This suggests that incentivizing surgeons to perform surgeries at different hospitals without requiring them to move to a different region may be a fruitful avenue for policy, especially as surgeons already tend to "multi-home." It can also complement potential spatial health policies. Reallocations of surgeons within regions may increase inequality between regions, but policies facilitating patient travel to high-survival regions may alleviate inequality in patient access to high-survival providers.

1.6.3 Discussion: payments in healthcare

These results highlight that the current sorting of surgeons across hospitals does not maximize aggregate patient outcomes. This may be surprising since hospitals and surgeons have strong incentives to maximize their patients' outcomes. This is particularly true for CABG surgery, since patient outcomes for this surgery are publicly reported. Why would the sorting of surgeons across hospitals be worse than random?

The goal of this paper is not to offer a comprehensive model of surgeon sorting, so the sorting model used is restricted to patient survival. Yet, surgeons are likely to sort on many other dimensions than patient survival. Surgeons may sort into hospitals based on "prestige," to attract more patients, for example. Using external measures of quality for hospitals in Subsection 1.4.2 indicates that high-survival hospitals tend to be high CMS ratings hospitals. Other reasons for positive assortative matching in the absence of complementarities may include better amenities at higher-survival hospitals. Bloom et al. (2020) and Munoz and Otero (2023) have shown that better management practices translate into better patient outcomes, which probably also translates into better work amenities for surgeons such as ease of scheduling, better technology, or higher-quality peers.

How to incentivize high-survival surgeons to practice in lower-survival hospitals? It is likely difficult for low-survival hospitals to attract high-survival surgeons. Hospitals cannot offer financial incentives to surgeons if they are not employed by the hospital. The federal anti-kickback statute prohibits hospitals from paying doctors for referrals, with the goal of removing financial incentives from doctors' clinical decisions. This means that low-survival hospitals will not be able to offer higher compensations to attract high-survival surgeons that they do not employ. In addition, existing policies aiming at incentivizing hospital quality actually penalize low-quality hospitals that do not meet minimum quality standards, like the CMS Hospital Readmission Reduction Program. This makes it even harder for such hospitals to attract high-survival surgeons with higher salaried compensation.

In the current fee-for-service system in Medicare, surgeons and hospitals are paid separately a predetermined amount. With bundled payment, the current situation could be exacerbated since high-survival hospitals may capture larger net payments. A potential avenue to incentivize high-survival surgeons to practice at low-survival hospitals would be reimbursements based on value-added. This idea is not new: it has notably been proposed in the context of teacher payments in the education literature (Hoxby, 2014).[23] The value-added of high-survival surgeons being much larger at low-survival hospitals than at high-survival ones, they will then have incentives to take some of their patients to lower-survival hospitals with little impact on their patients' survival. A similar argument can be made for hospitals, paying them proportionally to their value-added to the operating surgeon. Obviously, value-added payments are extremely hard to put in place since value-added is hard to compute. Furthermore, more work is needed to evaluate general equilibrium effects of such reallocations of doctors across hospitals, in particular taking into account learning and spillovers across physicians.

1.7 Conclusion

Healthcare providers care jointly for their patients. Substantial variation across physicians and hospitals has been documented in the literature, yet little was known about this joint production process and its consequences for our understanding of provider quality. This paper directly examines the joint production function of patient survival between surgeons and hospitals in the context of CABG surgery, and its consequences for aggregate patient outcomes.

This paper highlights the importance of interactions between the quality of different provider types in the production function of patient outcomes. In the context of survival

23. The education literature has long recognized the importance of teacher and school value-added and their interactions, or match effects, for student outcomes. Jackson (2013) notably estimates that these match effects account for about a quarter of teacher quality.

from CABG surgery, I find that surgeon and hospital quality are imperfect substitutes, so that the return to allocating low-survival surgeons to high-survival hospitals is larger than for high-survival surgeons. The value-added of high-survival surgeons is larger at low-survival hospitals, and the value-added of high-survival hospitals is larger for low-survival surgeons. These findings relate the economics of the production technology to well-known facts in the medical literature related to “failure-to-rescue” mechanisms. High-survival hospitals tend to achieve higher survival rates by saving patients from complications. At the same time, higher-skill surgeons tend to achieve lower complication rates.

This paper also provides evidence of positive assortative matching of surgeons into hospitals. These findings complement empirical evidence from the labor literature investigating worker sorting across firms using worker earnings (Abowd, Kramarz, and Margolis, 1999; Card, Heining, and Kline, 2013; Kline, Saggio, and Sølvsten, 2020; Bonhomme, Lamadon, and Manresa, 2022). I estimate sorting in a specific yet important labor market using a direct measure of output: patient survival. I also document a situation where surgeon sorting does not maximize aggregate outcomes: surgeon and hospital quality are imperfect substitutes, but high-survival surgeons sort into high-survival hospitals. Surgeon sorting is worse than random for aggregate patient survival. I also show that positive assortative matching at the national level is partially driven by providers co-locating across space in the U.S. Yet, there remain substantial positive assortative matching within regions.

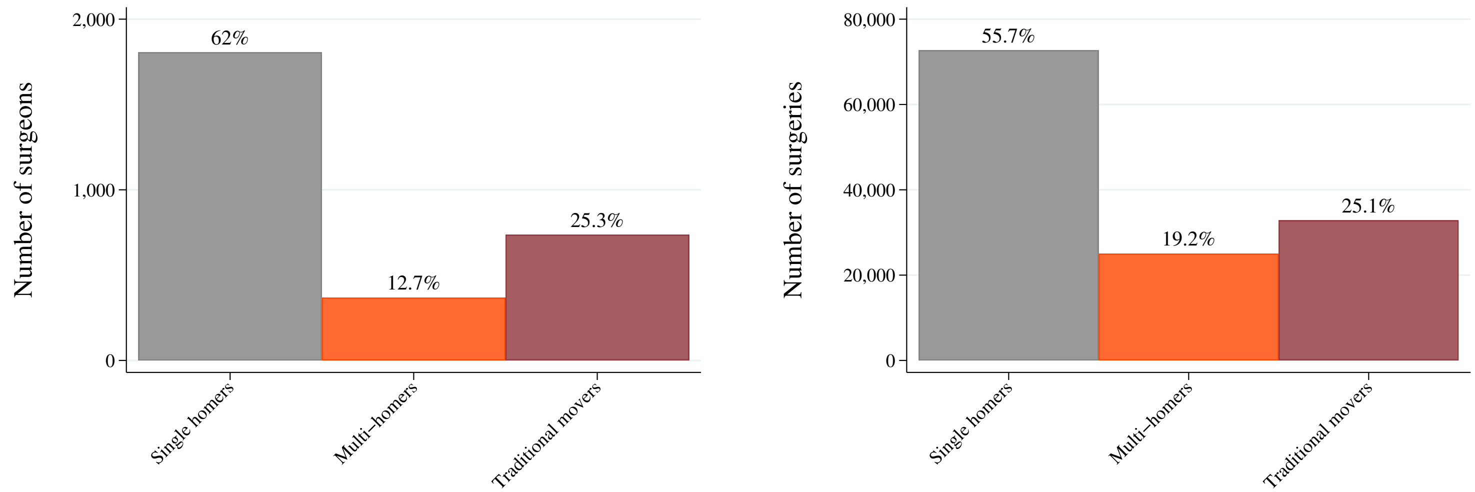
This detailed quantification of the joint production function between surgeons and hospitals at the micro-level has large consequences at the aggregate level for patient outcomes. I use partial equilibrium reallocation exercises to quantify the magnitude of the imperfect substitutability between surgeon and hospital quality. Simply randomly reallocating surgeons across hospitals would save about 200 lives a year within Medicare. Implementing the negative assortative matching allocation increases this number to 800 lives saved a year within Medicare. This implies that surgeon sorting has large consequences for aggregate patient outcomes. I also use these reallocation exercises to explore how much of the positive assortative matching result at the national level is driven by providers sorting across space. I estimate that at least 50% of the gains from national reallocations could be achieved by reallocating surgeons to hospitals *within* regions. This suggests that reallocating surgeons within regions may be a fruitful avenue for policy.

In outlining the importance of understanding the production function of hospitals for aggregate outcomes, this paper is in line with the literature showing the importance of endogenizing firms' internal organization to explain aggregate market outcomes. Adenbaum (2022) emphasizes the role of labor in explaining variation in firms' TFPs, by estimating the role for worker productivity and endogeneous specialization to represent about 75% of the variance in firm-level TFP. Freund (2022) estimates that the rise in coworker complementarities explains a quarter to a half of the rise in wage inequality in Germany. Taking into account team formation within hospitals, as well as co-workers complementarities within the hospital, which includes surgeons, hospitals, nurses, and nonmedical staff, is likely to provide fruitful insights into the determinants of hospital productivity.

1.8 Figures and Tables

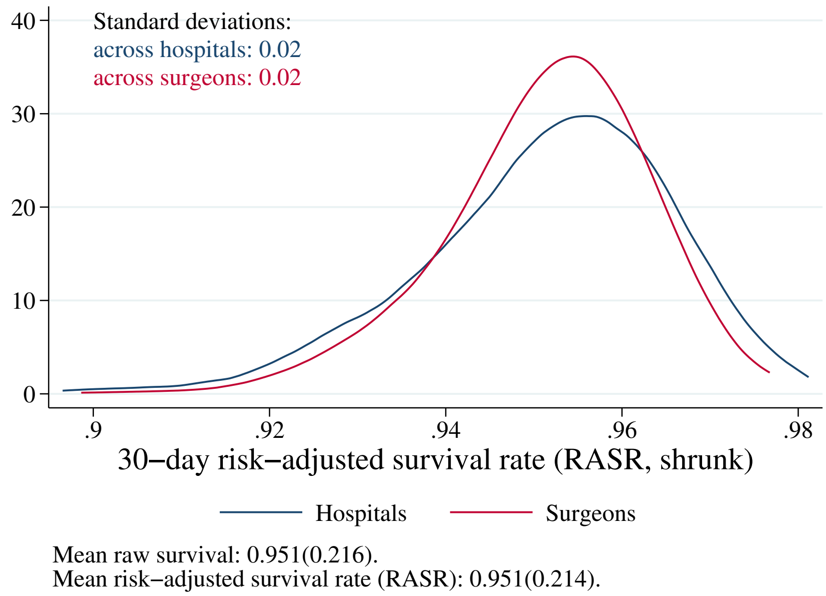
Figure 1.1: Proportion of “single-homers,” “multi-homers,” and “traditional movers”

(a) Proportion of surgeons (b) Proportion of surgeries



Notes: “Multi-homers” are defined as surgeons who performed CABG surgeries at more than one hospital within a year for four years of more in the sample. They represent 12.7% of all surgeons performing CABG in the sample, and 19.2% of CABG surgeries in the sample. “Traditional movers” are surgeons who performed CABG surgeries at more than one hospital in one, two, or three years in the sample. They represent 25.3% of all surgeons performing CABG in the sample, and 25.1% of CABG surgeries in the sample. “Single homers” include surgeons who only performed CABG surgeries at a unique hospital in the sample. They represent 62% of all surgeons performing CABG in the sample, and 55.7% of CABG surgeries in the sample. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

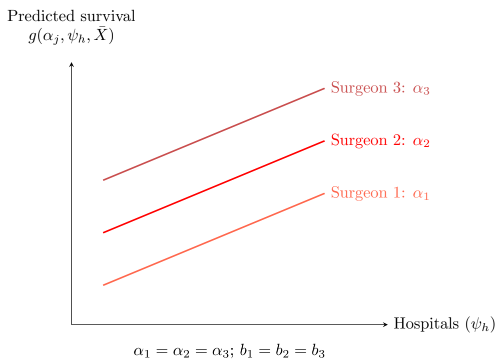
Figure 1.2: Distribution of 30-day risk-adjusted survival across surgeons and hospitals



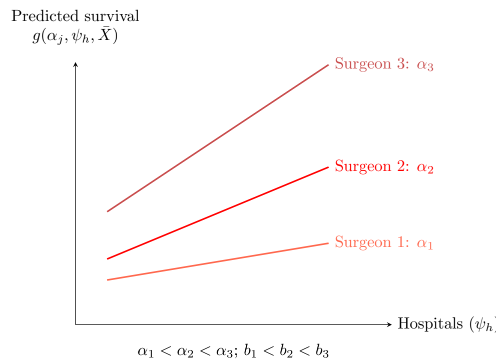
Notes: This graph depicts the distribution of average risk-adjusted 30-day survival (RASR) across surgeons and hospitals, weighted by the number of patients at each provider. These averages are adjusted for measurement error using empirical Bayes shrinkage as detailed in Appendix 1.9.4, which "shrinks" noisily estimated averages toward the mean. Risk-adjustment is performed by predicting 30-day survival using a logit model as delineated in Appendix 1.9.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.3: Impact of alternative assumptions on interactions between surgeon and hospital quality on predicted survival

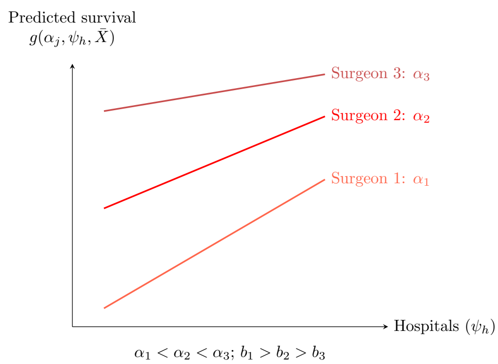
(a) Perfectly separable



(b) Complements



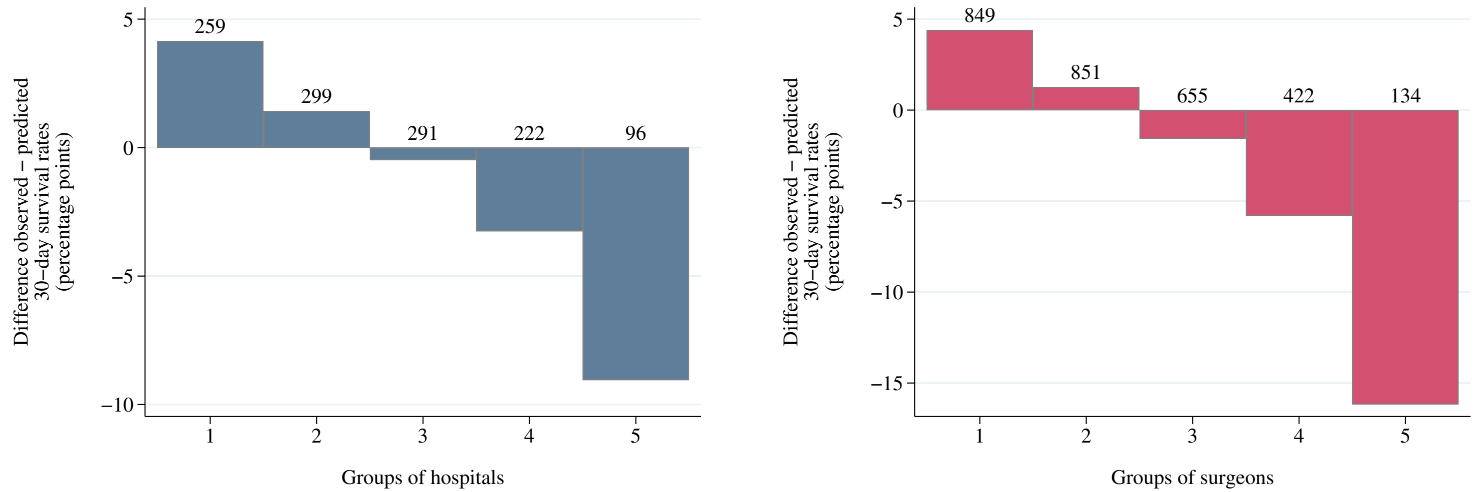
(c) Imperfect substitutes



Notes: These figures illustrate the impact of alternative assumptions on the interactions between surgeon and hospital quality on predicted patient survival across providers. Panel 1.3a describes the case where surgeons and hospitals are perfectly separable. In this case, the slope across hospitals types b_j is equal for all surgeons: the return to allocating surgeons to high-ψ_h hospitals is independent of the surgeon. Panel 1.3b illustrates the case where surgeons and hospitals are complements. The slope across hospital types is greater for high-α_j surgeons: the return to allocating surgeons to high-ψ_h hospitals is greater for high-α_j surgeons than for lower-α_j ones. Panel 1.3c details the imperfect substitutability case. Now the slope across hospital types is greater for low-α_j surgeons: the return to allocating surgeons to high-ψ_h hospitals is greater for low-α_j surgeons than for higher-α_j ones.

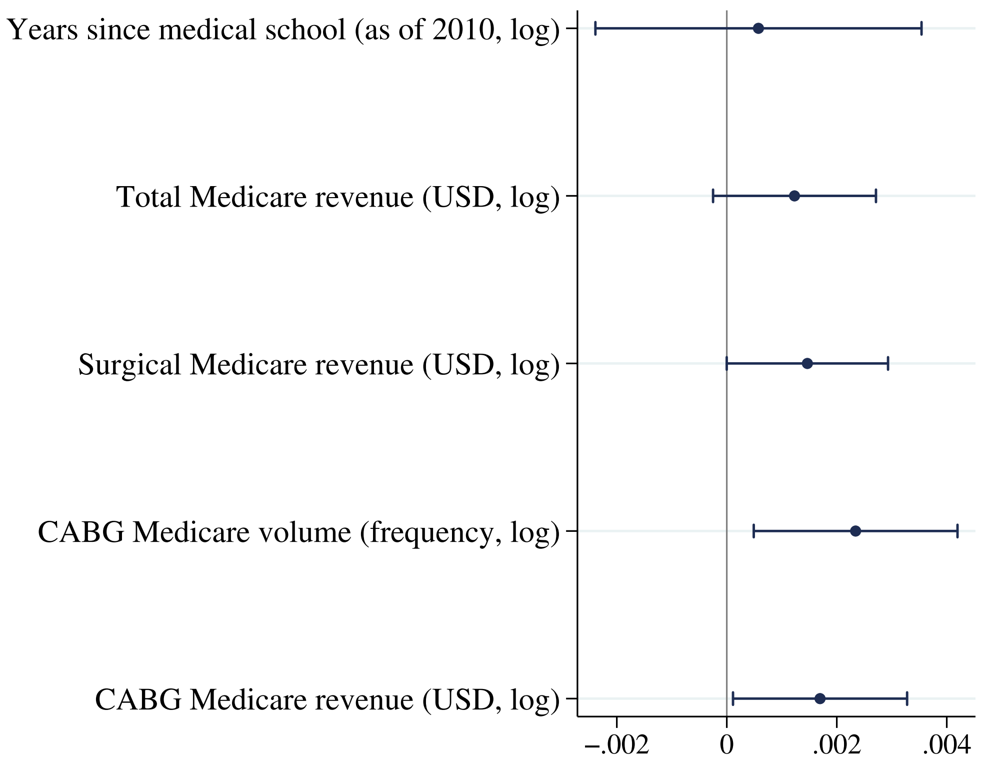
Figure 1.4: Risk-adjusted survival rate variation across k-means groups

(a) Hospitals (b) Surgeons



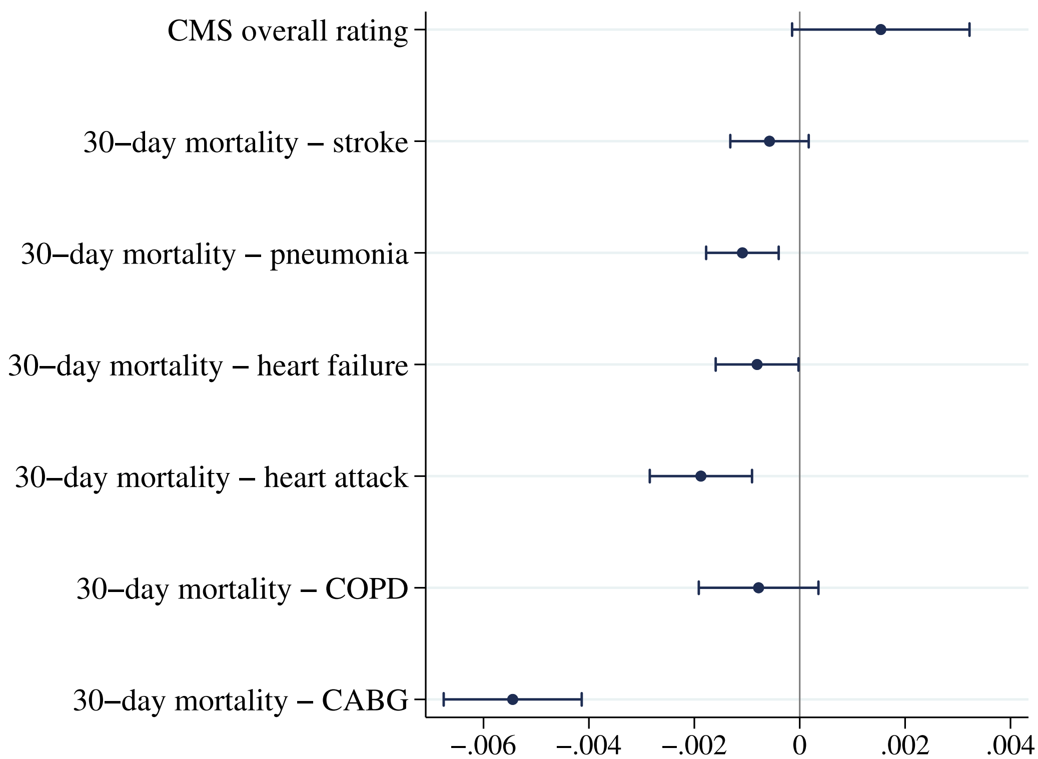
Notes: The variation in risk-adjusted 30-day survival across groups of hospitals and surgeons resulting from k-means clustering is large. Risk-adjusted survival is expressed as the difference between the average observed and average predicted 30-day survival. Numbers on top of bars indicate the number of hospitals or surgeons per group. Predicted survival is computed as described in Section 1.3 using a logit model and including all patient covariates and year fixed effects. K-means clustering is performed using average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.5: Correlation of estimated surgeon group effects with external measures of surgeons' skill



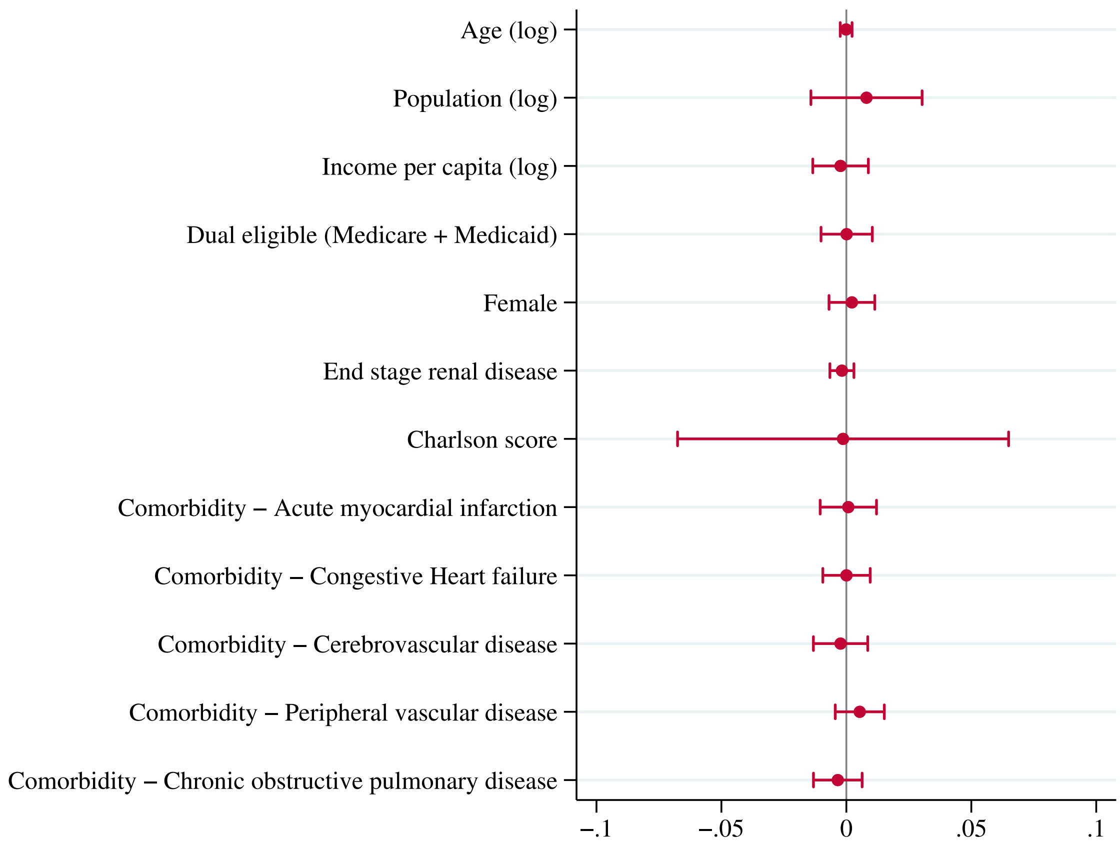
Notes: This graph reports the point estimates and 95% confidence intervals from regression of the surgeon-group estimated effects on surgeon-level covariates. The surgeon-group estimates include the fixed effects with interactions as $\widehat{\alpha_l} + \frac{1}{K}\sum_k \widehat{\kappa_{lk}}$ from equation (1.6), i.e., weighting each interaction with each hospital group equally. Surgeon-group effects are positively correlated with surgeons experience in performing CABG within Medicare, measured in log-revenue or log-frequency. It is statistically significant. Surgeon-group effects are also positively correlated with surgeons' surgical and overall experience, measured as surgical and total Medicare revenues respectively, but this is not statistically significant. The relationship with tenured experience - measured as the number of years since medical school graduation - is not statistically different from zero. The R^2 of the regression including all surgeon covariates amount to less than 0.01. Surgeons' Medicare revenues and frequency are calculated for years 2012 to 2017 from the CMS Medicare Physician & Other Practitioners file. Years since medical school graduation is calculated as of 2010 based on the medical school graduation in the CMS Doctors and Clinicians dataset. Confidence intervals displayed are at 95% constructed using robust standard errors.

Figure 1.6: Correlation of estimated hospital group effects with external measures of hospital quality



Notes: This graph reports the point estimates and 95% confidence intervals from regression of the estimated hospital-group estimated effects on hospital-level covariates. The hospital-group estimates include the fixed effects with interactions as $\widehat{\alpha_k} + \frac{1}{L}\sum_l \widehat{\kappa_{lk}}$ from equation (1.6), i.e., weighting each interaction with each surgeon group equally. Hospital-group estimates are positively correlated with hospital CMS five-star ratings but this is not statistically significant. They are negatively correlated with 30-day risk-adjusted mortality for six conditions publicly reported by CMS as part of the five-star rating. The R^2 of the regression including all CMS quality measures amount to less about 0.045. The CMS five-star ratings and mortality measures are obtained from the CMS Hospital General Information and Complications and Deaths datasets for 2017. Confidence intervals displayed are at 95% constructed using robust standard errors.

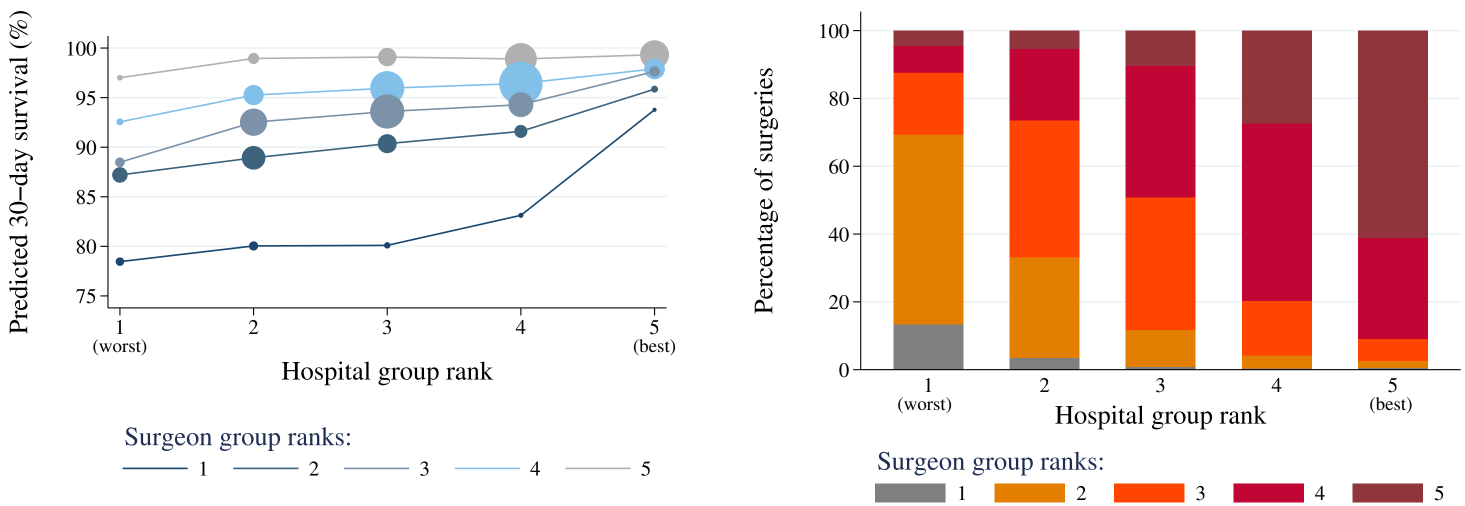
Figure 1.7: No evidence that surgeons systematically triage sicker patients into higher-survival hospitals



Notes: This graph examines the existence of "triaging" for multi-homers, i.e., whether surgeons tend to operate on sicker patients at higher-survival hospitals using patient observables. All coefficients are close to zero and statistically insignificant, suggesting a limited role for triaging into hospitals using patient observables. Coefficients reported in this graph correspond to the estimated $\hat{\beta}$ from the regression $x_{ijh} = \alpha + \beta \text{rank}_{k(h)} + \lambda_j + \epsilon_{ijh}$. x_{ijh} correspond to the covariates of patients treated by surgeon j at hospital h, and λ_j are individual surgeon fixed effects. The ranks of hospital groups are computed as the rank in predicted survival based the model from equation (1.6) assuming each surgeon group is equally likely for each hospital group. "Multi-homers" are defined as surgeons who performed CABG surgeries at more than one hospital *group* within a year for four years of more in the sample. Surgeon and hospital groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Confidence intervals displayed are 95% confidence intervals constructed using clustered standard errors at the surgeon level.

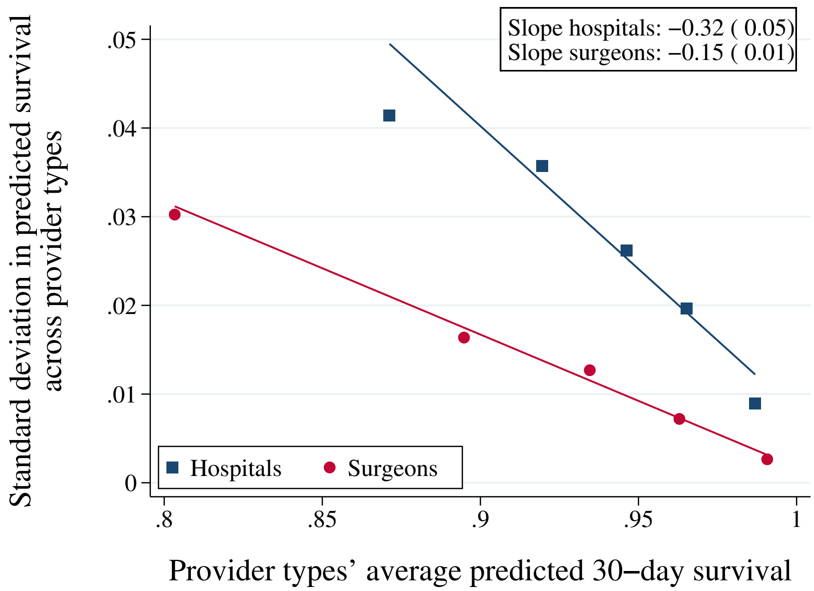
Figure 1.8: Imperfect substitutability and sorting

(a) Imperfect substitutability (b) Sorting



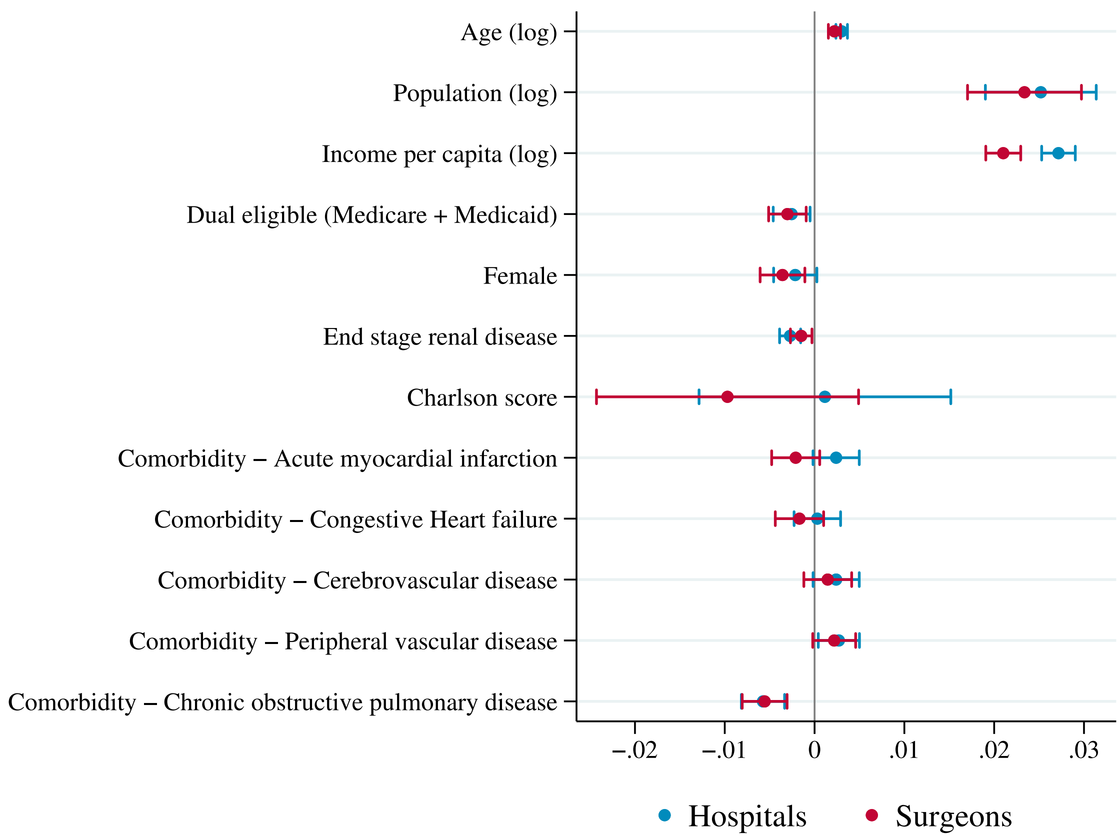
Notes: These graphs show results when assuming selection on observables as delineated in equation (1.6). Panel 1.8a displays the predicted 30-day survival for the average patient in the data across hospital and surgeon groups where groups are described by their relative rankings. The production function of survival appears to be sub-modular: the return of allocating low-survival surgeons to high-survival hospitals is greater than for high-survival surgeons. The slopes of fitted lines across hospital rankings for each surgeon group are reported in Table 1.4: the slope for lower-rank surgeons is greater than for high-survival surgeons. Marker sizes are proportional to the number of surgeries performed by each hospital-surgeon group. Panel 1.8b describes the percentage of surgeries performed by each surgeon group at each hospital group, where groups are described by their relative rankings. Surgeries at high-survival hospitals tend to be performed by high-survival surgeons: high-survival surgeons sort into high-survival hospitals. Surgeon groups are ranked based on the predicted 30-day risk-adjusted survival for each group assuming each interaction with a hospital group is equally likely. Similarly, hospital groups are ranked using the predicted 30-day risk-adjusted survival for each group assuming each interaction with a surgeon group is equally likely. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.9: Lower dispersion in survival among high-survival providers



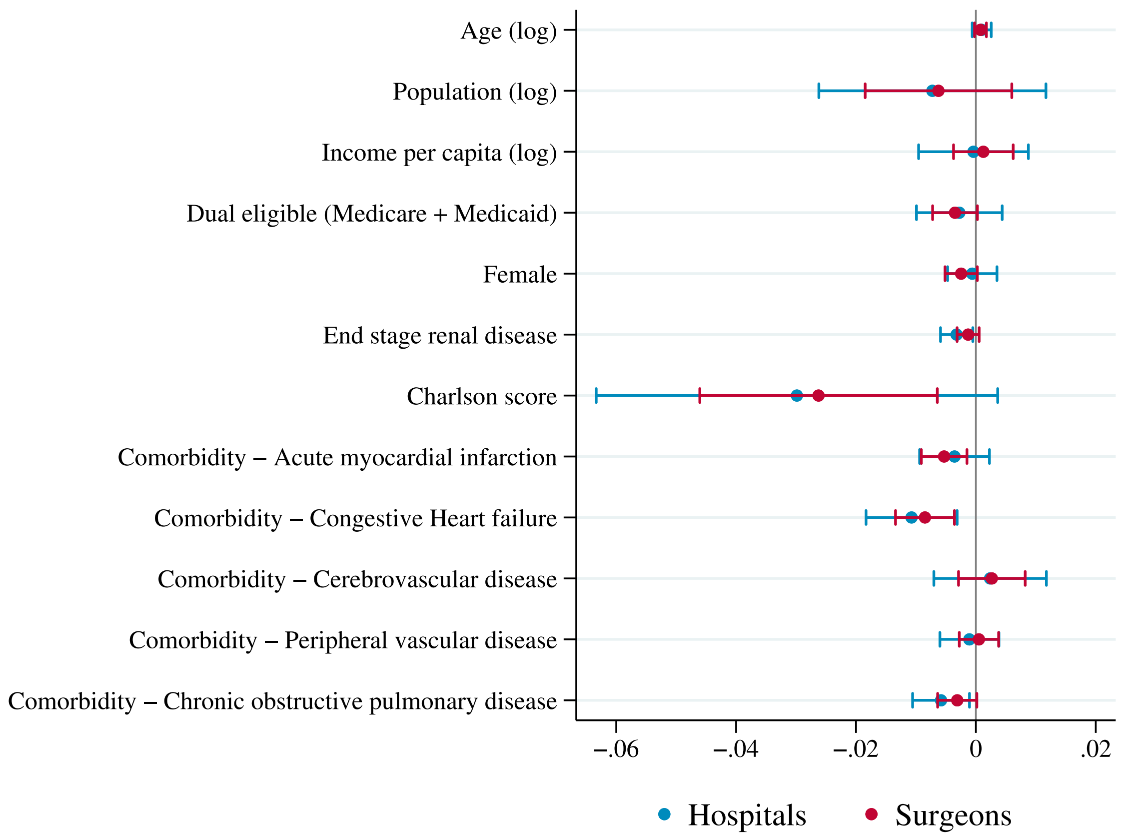
Notes: This graph shows that high-survival hospitals exhibit more similar survival across surgeons. This is depicted by the hospitals serie, which relates the standard deviation in predicted 30-day risk-adjusted survival across surgeon groups for each hospital group to the average predicted 30-day risk-adjusted survival at this hospital group. Similarly, higher-survival surgeons achieve more similar survival across hospitals. This is depicted by the surgeons serie, which relates the standard deviation in predicted 30-day risk-adjusted survival across hospital groups for each surgeon group to the predicted 30-day risk-adjusted survival achieved by this surgeon group. Average predicted 30-day risk-adjusted survival is calculated for the average patient in the average year and weighted by the number of surgeries performed by each surgeon group at each hospital group. Standard deviations in predicted 30-day risk-adjusted survival across providers are computed using the total number of surgeries performed by each surgeon group for the hospital serie, and using the total number of surgeries performed at each hospital group for the surgeon serie. Predictions come from estimated parameters from the model delineated in equation (1.6). Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.10: Relationship between patient observables and provider rankings



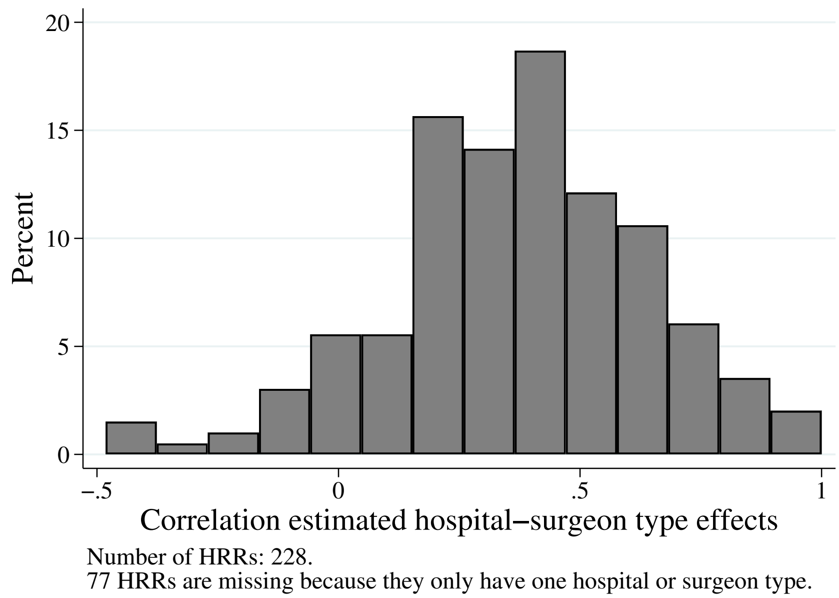
Notes: This graph shows the relationship between patient observables and the rank of the provider they receive surgery from. Highly ranked surgeons and hospitals tend to treat older patients living in highly populated high income ZIP codes, suggesting that high-type providers tend to be located in such locations. Coefficients reported in this graph correspond to the estimated $\hat{\beta}$ from the regression $x_{il(j)} = \alpha + \beta \text{rank}_{l(j)} + \epsilon_{il(j)}$. $x_{il(j)}$ correspond to the covariates of patients treated by provider group $l(j)$. The ranks of surgeon and hospital groups are computed as the rank in predicted risk-adjusted survival based the model from equation (1.6). Confidence intervals displayed are at 95% constructed using robust standard errors. Provider groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Income per capita and population are computed from the patient ZIP code of residence and come from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Professional fees come from the Medicare 20% carrier Research Identifiable Files, hospital stays from the Medicare MedPAR Research Identifiable Files, and beneficiary information from the Medicare Beneficiary Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.11: Relationship between patient observables and provider rankings, controlling for provider locations



Notes: This graph shows the relationship between patient observables and the rank of the provider group they receive surgery from after controlling for the provider's locations. The relationships with patient socioeconomic status such as ZIP code income and population and dual eligibility become statistically insignificant after controlling for the provider locations. However, there exist suggestive evidence of advantageous selection into provider types using health measures, such as the Charlson score or the fraction of patients with congestive heart failure. Coefficients reported in this graph correspond to the estimated $\hat{\beta}$ from the regression $x_{il(j)} = \alpha + \beta \text{rank}_{l(j)} + \lambda_z(j) + \epsilon_{il(j)}$. $x_{il(j)}$ correspond to the covariates of patients treated by provider group $l(j)$, and $\lambda_z(j)$ are provider HRR fixed effects using the ZIP code of their primary practice. The ranks of surgeon and hospital groups is computed as the rank in predicted risk-adjusted survival based the model from equation (1.6). Confidence intervals displayed are 95% confidence intervals constructed using clustered standard errors at the HRR level. Provider groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Income per capita and population are computed from the patient ZIP code of residence and come from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. The surgeon's ZIP code is the primary practice ZIP code recorded in the NPPES data. Professional fees come from the Medicare 20% carrier Research Identifiable Files, hospital stays from the Medicare MedPAR Research Identifiable Files, and beneficiary information from the Medicare Beneficiary Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.12: Sorting within hospital referral regions (HRRs)

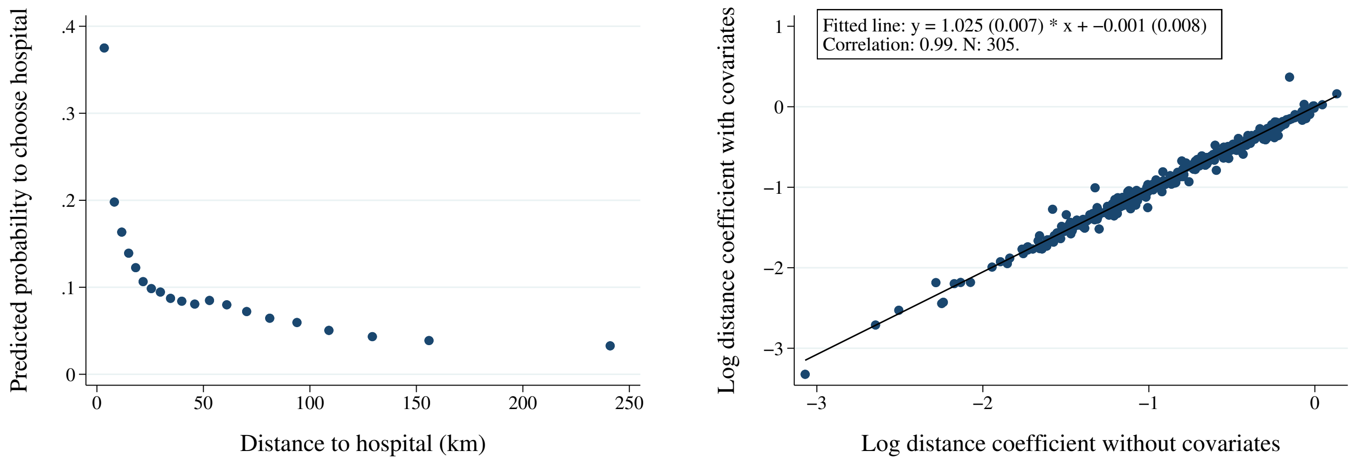


Notes: This graph shows the distribution of the estimated correlation between surgeon and hospital group effects computed for each HRR separately. There exist substantial positive assortative matching *within* HRRs for a substantial fraction of HRRs. Predictions come from estimating the model described by equation (1.6). Surgeon group effects are calculated assuming equal probability for each hospital group interaction. Similarly, hospital group effects are calculated assuming equal probability for each surgeon group interaction. The correlations between surgeon and hospital group effects are computed for the subset of patients treated in a hospital located in each specific HRR. The definition of hospital referral regions (HRRs) follows the definition of the Dartmouth Atlas Project. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.13: Distance to the hospital is a strong predictor of hospital choice within an HRR

(a) Probability to choose a hospital and distance

(b) Distance parameters per HRR

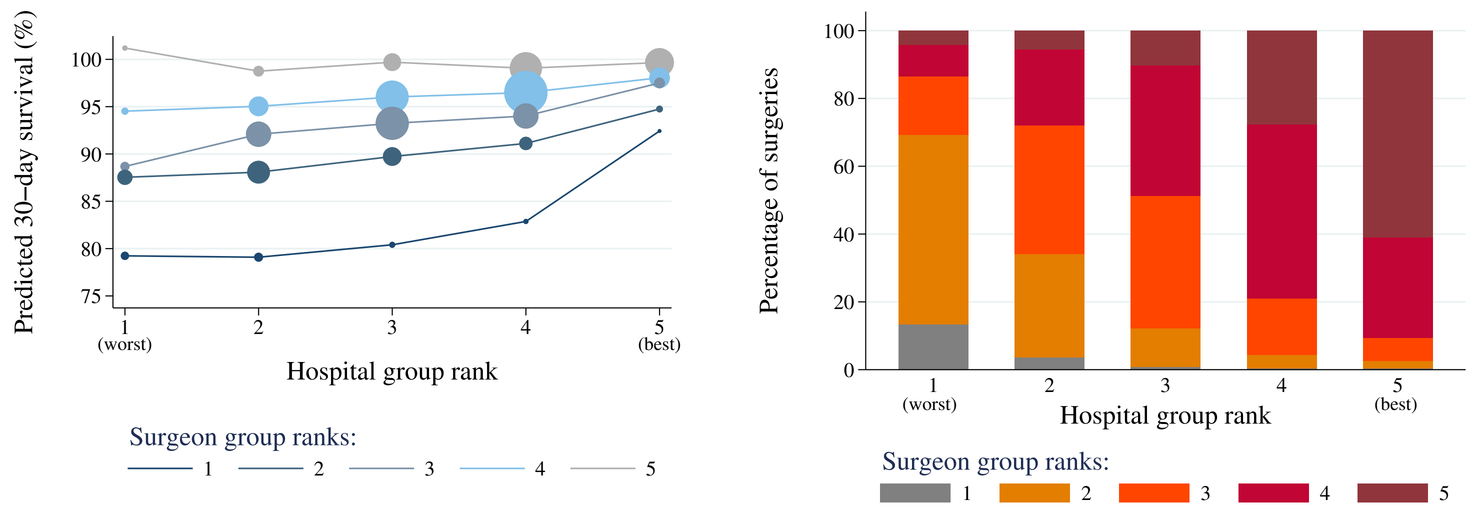


Notes: Panel 1.13a depicts the relationship between the predicted probabilities to choose a hospital using the demand model delineated in equation (1.7), estimated HRR by HRR, and the distance between the patient and the hospital ZIP codes. Only predicted probabilities for hospitals within a patient's residential HRR are included. The graphs summarize this relationship using a binned scatter plot with twenty equally sized bins. Panel 1.13b depicts the estimated demand parameter for the logarithm of distance τ in the specification without patient observables from equation (1.7) and the specification with patient observables such that $u_{j(i)h} = \delta_h(X_i) - \tau \ln(d_{ih}) + \kappa_j + \eta_{j(i)h}$. The estimated parameters for the logarithm of distance are extremely similar across the two specifications, with a correlation over 0.99, hence lending support to the exclusion restriction assumption. Included patient covariates are patient age, charlson score, and ZIP code log income per capita. Hospital ZIP codes come from the 2017 National Plan and Provider Enumeration System (NPPES) data, and beneficiary ZIP codes from the Medicare Beneficiary Research Identifiable Files. Distances are calculated using ZCTA-to-ZCTA distances for distances below 100 miles, using HSA-to-HSA distances when above 100 miles and when patient and provider HSAs differ, and capped at 100 miles when patients and providers are in the same HSA but with ZCTAs distant over 100 miles. Patients' residential ZIP codes are mapped to income per capita and total population using the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. The definition of hospital referral regions (HRRs) follows the definition of the Dartmouth Atlas Project. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.14: Imperfect substitutability and sorting with a control function approach

(a) Imperfect substitutability

(b) Sorting



Notes: These graphs show results when using the control function approach delineated in equation (1.11). Panel 1.14a displays the predicted 30-day survival for the average patient in the data across hospital and surgeon groups where groups are described by their relative rankings. The production function of survival appears to be sub-modular: the return of allocating lower-rank surgeons to high-survival hospitals is greater than for high-survival surgeons. The slopes of fitted lines across hospital rankings for each surgeon group are reported in Table 1.7: the slope for lower-rank surgeons is greater than for high-survival surgeons. Marker sizes are proportional to the number of surgeries performed by each hospital-surgeon group. Panel 1.14b describes the percentage of surgeries performed by each surgeon group at each hospital group, where groups are described by their relative rankings. Surgeries at high-survival hospitals tend to be performed by high-survival surgeons: high-survival surgeons sort into high-survival hospitals. Surgeon groups are ranked based on the predicted 30-day risk-adjusted survival for each group assuming each interaction with a hospital group is equally likely. Similarly, hospital groups are ranked using the predicted 30-day risk-adjusted survival for each group assuming each interaction with a surgeon group is equally likely. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.1: Exclusions to final sample

Sample	Number of observations	Number of patients	Number of surgeons	Number of hospitals
CABG professional fee claims	154,655	122,531	3,815	-
	(100%)	(100%)	(100%)	-
Matched professional fees to hospitals stays	139,166	115,925	3,780	1,327
	(90%)	(95%)	(99%)	(100%)
Excluding inconsistent specialties	136,718	114,752	3,578	1,323
	(88%)	(94%)	(94%)	(100%)
Exclude surgeries performed end of 2010	136,600	114,652	3,578	1,323
	(88%)	(94%)	(94%)	(100%)
Exclude less than five surgeries per surgeon at a hospital	131,214	111,370	2,923	1,174
	(85%)	(91%)	(77%)	(88%)
Exclude patients or providers located outside of mainland U.S.	130,844	111,059	2,911	1,167
or in an HRR where all patients chose a hospital outside the HRR	(85%)	(91%)	(76%)	(88%)

Notes: Percentages in parenthesis are expressed as a percentage of the number in the initial CABG sample in the first line. Professional fee claims for coronary artery bypass graft (CABG) surgery are isolated using healthcare common procedure coding system (HCPCS) codes 33510-33516, 33533-33536, and 33517-33523 in the claim line file. The operating surgeon is identified as the performing provider for the claim line relative to a CABG HCPCS code. The professional fee claims are matched to hospital stays if the professional fee claim date falls within the admission and discharge date of a unique hospital stay for the patient. The total number of observations is larger than the total number of patients because some patients undergo CABG surgery multiple times in the final sample time period and because some surgeries are linked to multiple performing physicians in the professional claim lines. 604 patients received CABG surgery more than once in the 2011-2017 final sample. 16.5% of surgeries exhibit more than one performing surgeon in the final sample. Physician specialties and hospital ZIP codes are identified by linking the provider's unique national provider identifier (NPI) to the National Plan and Provider Enumeration System (NPPES) data. Included primary specialties as defined in the NPPES are thoracic surgery, surgery, specialist, vascular surgery, cardiovascular disease, transplant surgery, vascular specialist, and surgical critical care. Patient and hospital ZIP codes are linked to hospital referral regions (HRRs) as defined by the Dartmouth Atlas Project. Professional fees come from the Medicare 20% carrier Research Identifiable Files, hospital stays from the Medicare MedPAR Research Identifiable Files, and beneficiary information from the Medicare Beneficiary Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.2: Patients summary statistics

	Mean	Standard Deviation
Socio-demographics & health		
Age		
Less than 65	0.11	0.31
[65;70[	0.24	0.43
[70;75[	0.25	0.43
[75;80[	0.21	0.41
[80;85[	0.14	0.34
[85;90[	0.05	0.21
[90;95[	0.00	0.07
[95;100[	0.00	0.01
More than 100	0.00	0.00
Female	0.30	0.46
Dual eligible (Medicaid + Medicare)	0.17	0.38
Income per capita (USD, x1,000)	33.41	13.96
ZIP code population (x1,000)	25.28	18.98
End Stage Renal Disease	0.05	0.21
Charlson score	3.41	2.66
Comorbidities		
Acute myocardial infarction	0.40	0.49
Congestive heart failure	0.42	0.49
Peripheral vascular disease	0.26	0.44
Cerebrovascular disease	0.40	0.49
Chronic obstructive pulmonary disease	0.30	0.46
Outcomes		
30-day mortality	0.05	0.22
60-day mortality	0.06	0.23
Length of stay	10.32	7.50
N	111,059	

Notes: Patient residential ZIP codes are mapped to income per capita and total population using the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. The Charlson score and comorbidities are obtained using all diagnoses appearing in inpatient, outpatient, and professional fee claims up to twelve months prior to the surgery. Professional fees come from the Medicare 20% carrier Research Identifiable Files, hospital stays from the Medicare MedPAR Research Identifiable Files, and beneficiary information from the Medicare Beneficiary Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.3: Estimated coefficients on patient observables for risk adjustment

	Selection on observables	Control function
Age - [65;70[	0.0036	0.0034
	(0.0023)	(0.0023)
Age - [70;75[	-0.0037	-0.0038
	(0.0023)	(0.0023)
Age - [75;80[	-0.0153***	-0.0154***
	(0.0023)	(0.0024)
Age - [80;85[	-0.0310***	-0.0313***
	(0.0025)	(0.0026)
Age - [85;90[	-0.0349***	-0.0348***
	(0.0034)	(0.0034)
Age - [90;95[	-0.0803***	-0.0799***
	(0.0085)	(0.0086)
Age - [95;100[	-0.0272	-0.0234
	(0.0450)	(0.0454)
Female	-0.0190***	-0.0191***
	(0.0013)	(0.0013)
Dual eligible (Medicare + Medicaid)	0.0072***	0.0075***
	(0.0017)	(0.0018)
Income per Capita (log)	0.0051***	0.0044**
	(0.0017)	(0.0020)
Population (log)	-0.0004	-0.0003
	(0.0005)	(0.0006)
End stage renal disease	-0.0315***	-0.0315***
	(0.0030)	(0.0030)
Charlson score	-0.0023***	-0.0023***
	(0.0003)	(0.0003)
Comorbidity - Acute myocardial infarction	-0.0097***	-0.0098***
	(0.0013)	(0.0013)
Comorbidity - Congestive Heart failure	-0.0293***	-0.0296***
	(0.0014)	(0.0014)
Comorbidity - Peripheral vascular disease	-0.0110***	-0.0108***
	(0.0015)	(0.0015)
Comorbidity - Cerebrovascular disease	0.0006	0.0007
	(0.0013)	(0.0014)
Comorbidity - Chronic obstructive pulmonary disease	-0.0023*	-0.0027*
	(0.0014)	(0.0014)
Roy selection		0.0005
		(0.0008)
N	130,844	130,844

Notes: This table reports the estimated coefficient for patient covariates from running regressions delineated in equations (1.6) and (1.11). Income per capita and population are computed from the patient's ZIP code of residence and come from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Statistical significance: *** 2.5% , ** 5%, and * 10%. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.4: Relationship between predicted 30-day risk-adjusted survival and the ranking of hospital groups per surgeon group

	Predicted survival
Slope surgeon rank 1 (worst)	2.31
	(0.08)
Slope surgeon rank 2	1.64
	(0.01)
Slope surgeon rank 3	1.33
	(0.01)
Slope surgeon rank 4	0.81
	(0.00)
Slope surgeon rank 5 (best)	0.20
	(0.00)
p-value: equality of slopes	< 0.01
p-value: slope rank 5 $\geq$ 1	< 0.01
p-value: slope rank 4 $\geq$ 2	< 0.01
Observations	130,844
R-squared	0.99
Physician type FEs:	X

Notes: This table reports the estimated slope coefficient per surgeon group $\hat{\beta}^L$ from the regression $\hat{y}_{ijht} = \sum_{L=1}^{5} 1\{j \in L\}\, \beta^L \text{rank}_{k(h)} + \lambda_L + \epsilon_{ijht}$ where $\hat{y}_{ijht}$ is the predicted 30-day risk-adjusted survival from the model delineated in equation (1.6), L is the rank of the surgeon group, $k(h)$ is the group of hospital h, $\text{rank}_{k(h)}$ is the rank of hospital group $k(h)$ in terms of predicted 30-day survival, and λ_L are surgeon group fixed effects. The predicted survival is expressed in percentage points of survival. These slope coefficients correspond to the slope of fitted line across hospital rankings for each surgeon group displayed in Figure 1.8a. The production function of survival appears to be sub-modular: the slope for low-survival surgeons is larger than for high-survival surgeons. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included. Standard errors in parenthesis are robust standard errors.

Table 1.5: Decomposition of the explained variance in 30-day survival, net of covariates

	Percentage of explained variance net of covariates (%)	
	Selection on observables	Control function
Hospitals	8.77	5.80
$Var(\psi_{k(h)})$	(1.35)	(1.71)
Surgeons	66.42	72.78
$Var(\alpha_{l(j)})$	(2.68)	(4.34)
Sorting	24.80	21.42
$2 \times cov(\alpha_{l(j)}, \psi_{k(h)})$	(1.36)	(2.69)
Correlation surgeon-hospital FE	0.51	0.52
$Corr(\alpha_{l(j)}, \psi_{k(h)})$	(0.01)	(0.02)
N patients	111,059	111,059
N surgeons	2,911	2,911
N hospitals	1,167	1,167

Notes: This table shows the decomposition of the explained variance in 30-day survival net of covariates for a model without interactions as delineated in equation (1.13). Fixed effects are estimated following equation (1.12). The contribution of surgeon groups is large, and larger than the contribution of hospital groups. The covariance between estimated surgeon and hospital group fixed effects is positive and large, revealing strong positive assortative matching of surgeon groups across hospital groups. The sum of hospitals', surgeons', and the sorting contributions is equal to 100%. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included. Standard errors in parenthesis are obtained by bootstrapping the main sample with 200 replications.

Table 1.6: Robustness of the variance decomposition

	Percentage of explained variance net of covariates (%)			
	Surgeons $\frac{Var(\alpha_{l(j)})}{Var(\hat{y}^E)}$	Hospitals $\frac{Var(\psi_{k(h)})}{Var(\hat{y}^E)}$	Sorting $2 \times \frac{cov(\psi_{k(h)},\alpha_{l(j)})}{Var(\hat{y}^E)}$	Correlation $corr(\psi_{k(h)}, \alpha_{l(j)})$
Baseline	66.42	8.77	24.80	0.51
		A. Control function		
Control function	72.78	5.80	21.42	0.52
		B. Logit production function		
Logit	65.80	9.45	24.75	0.50
		C. Alternative number of groups		
K=5, L=10	71.57	6.28	22.15	0.52
K=10, L=5	63.72	10.40	25.87	0.50
K=10, L=10	68.76	7.72	23.52	0.51
K=20, L=20	69.06	7.86	23.08	0.50
K=50, L=50	70.22	7.67	22.10	0.48
		D. Alternative classifications		
Cond. moments hospitals	72.29	6.70	21.01	0.48
Quintiles risk-adjusted survival	67.25	7.89	24.87	0.54
Add. covariates to k-means, hospitals only	81.48	4.39	14.13	0.37
Add. covariates to k-means, surgeons only	34.40	38.13	27.47	0.38
Add. covariates to k-means, both	57.76	20.55	21.69	0.31
		E. Alternative samples		
Excluding emergencies	69.15	7.90	22.95	0.49

Notes: This table reports the variance decomposition as delineated in equation (1.13) for alternative specifications. The variance in predicted log-odds of 30-day survival is used for the logit model. Conditional moments used for k-means clustering for hospitals include risk-adjusted 30-day survival for patients above/below the median Charlson score, age, income per capita, and male/female. Quintiles include the same number of surgeries. For hospitals, additional covariates used for k-means clustering include the total number of beds, income per capita, and population in the hospital ZIP code. For surgeons, additional covariates used for k-means clustering include the total 2011-2017 cumulated Medicare activity, surgical activity, and CABG surgery amount in USD, income per capita, and population in the surgeon's average primary practice ZIP code. The sample without emergencies excludes all hospital claims with non-zero emergency department amounts. Primary practice locations come from the NPPES data, and ZIP code level information from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Surgeons' Medicare activity is computed using the national Medicare Provider Utilization and Payment Data. The number of beds per hospital comes from the CMS provider of services data. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.7: Robustness of the imperfect substitutability result

	(1) Baseline	(2) Control function	(3) Additional covariates	(4) No emergencies
Slope surgeon rank 1 (worst)	2.31	1.90	0.75	2.13
	(0.08)	(0.08)	(0.01)	(0.19)
Slope surgeon rank 2	1.64	1.44	0.56	1.73
	(0.01)	(0.01)	(0.00)	(0.03)
Slope surgeon rank 3	1.33	1.37	0.73	0.73
	(0.01)	(0.01)	(0.01)	(0.02)
Slope surgeon rank 4	0.81	0.83	0.69	0.61
	(0.00)	(0.00)	(0.01)	(0.00)
Slope surgeon rank 5 (best)	0.20	0.11	0.35	0.15
	(0.00)	(0.01)	(0.00)	(0.00)
p-value: equality of slopes	< 0.01	< 0.01	< 0.01	< 0.01
p-value: slope rank $5 \geq 1$	< 0.01	< 0.01	< 0.01	< 0.01
p-value: slope rank $4 \geq 2$	< 0.01	< 0.01	1.000	< 0.01
Observations	130,844	130,844	130,844	100,947
R-squared	0.99	0.99	0.87	0.95
Physician type FEs:	X	X	X	X

Notes: This table reports the estimated slope coefficient per group of surgeons for alternative specifications. The slopes $\hat{\beta}^L$ are obtained from the regression $\hat{y}_{ijht} = \sum_{L=1}^{5} 1\{j \in L\}\, \beta^L \text{rank}_{k(h)} + \lambda_L + \epsilon_{ijht}$ where $\hat{y}_{ijht}$ is the predicted 30-day risk-adjusted survival from models delineated in equations (1.6) or (1.11), L is the rank of the surgeons' group, $k(h)$ is the group of hospital h, $\text{rank}_{k(h)}$ is the rank of hospital group $k(h)$ in terms of predicted 30-day risk-adjusted survival, and λ_L are surgeon group fixed effects. The predicted survival is expressed in percentage points of survival. To exclude CABG surgery potentially performed in an emergency setting, I exclude all hospital claims with non-zero emergency department amounts. For hospitals, additional covariates used for k-means clustering include the total number of beds, income per capita, and population in the hospital ZIP code. For surgeons, additional covariates used for k-means clustering include the total 2011-2017 cumulated Medicare activity, surgical activity, and CABG surgery amount in USD, income per capita, and population in the surgeon primary practice ZIP code. I average population and income per capita at the physician level across all primary practice ZIP codes they are observed at over the time frame of my sample. Primary practice locations come from the NPPES data, and ZIP code level information from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Surgeons' Medicare activity is computed using the national Medicare Provider Utilization and Payment Data. The number of beds per hospital comes from the CMS provider of services data. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included. Standard errors in parenthesis are robust standard errors.

Table 1.8: Alternative allocations of surgeons to hospitals: surgeon sorting has large consequences for patient survival

	Random		Negative assortative matching	
	National	Within HRR	National	Within HRR
$corr(\hat{\alpha}_{l(j)} + \bar{\kappa}_{l(j)}, \hat{\psi}_{k(h)} + \bar{\kappa}_{k(h)})$	-0.00	0.20	-0.84	-0.27
Change in deaths per 1,000 (reallocated - baseline)				
Aggregate	-2.99	-1.68	-9.91	-5.35
	(0.13)	(0.12)	(0.11)	(0.10)
% change from current allocation	-6	-3	-20	-11
% of national change	-	56	-	54
Standard deviation	-3.31	-2.27	-13.49	-7.21
	(0.15)	(0.14)	(0.12)	(0.13)
% change from current allocation	-7	-5	-29	-16
% of national change	-	69	-	53

Notes: This table reports results of a partial equilibrium reallocation exercise where surgeons are reallocated to alternative types of hospitals. Two types of reallocations are reported: random reallocation and negative assortative matching. Patients and surgeons are reallocated either nationally or within HRRs only. In each simulation, patients are first randomly allocated to surgeons conditional on the number of surgeries available per surgeon group. Then, in the random reallocation, patient-surgeon pairs are then randomly reallocated to hospital types conditional on the number of surgeries available per hospital type. For the negative assortative matching reallocation, surgeons from the lowest type and their patients are allocated to the best hospital type until no surgeries are available at this hospital type, and so on. For reallocations within HRRs, surgeon groups operating in an HRR are reallocated to alternative hospitals within the same HRR. 30-day mortality is predicted using parameter estimates from equation (1.6). Results are obtained using 100 simulations, and bootstrap standard errors are in parentheses (computed using 200 replications). A national random reallocation decreases the average number of deaths within 30-day as well as the dispersion in 30-day mortality for both specifications. Reallocating low-survival surgeons to high-survival hospital nationally results in negative assortative matching, leading to a decrease in average 30-day mortality and its dispersion that is more than three times larger in both specifications. Implementing reallocations within HRRs achieves more than 50% of the gains from national reallocations. The definition of hospital referral regions (HRRs) follows the definition of the Dartmouth Atlas Project. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

1.9 Appendix

1.9.1 Institutional details: coronary artery bypass graft (CABG) surgery

Processes involved during CABG surgery. CABG surgery requires team work at the center of which are the operating surgeon's skills and resources put in place by the hospital. This surgery requires an operating room and involves the operating surgeon, an anesthesiologist, a perfusionist to operate the heart-lung machine which provides blood and oxygen through the body in place of the heart and lungs, as well as surgical nurses and additional surgical staff. Aside from the operating surgeon, the rest of the team is determined by the hospital. After surgery, a team of doctors, usually called "hospitalists", and nurses monitor and care for the patient during recovery. Patients stay on average between eight and twelve days in the hospital, so that while the operating surgeon skill may be crucial to successfully restore blood flow, the hospital has a role to play in managing post-operative complications.

Cardiac surgeons. Cardiac surgeons are highly specialized physicians. In addition to medical school and residency training, cardiothoracic surgeons continue their specialization with a two to three years fellowship. They can also specialize even more within cardiothoracics by specializing in cardiac surgery. For surgeons performing CABG surgery, this surgery is their most common surgery on average on Medicare patients, followed by heart valve replacement and aortic surgery.

Since cardiac surgeons tend to be independent from hospitals, they obtain privileges and operate at multiple hospitals (Huckman and Pisano, 2006; Kolstad, 2013). While potentially costly, operating at multiple hospitals allows for more flexibility for surgeons. Operating rooms are in limited capacity: a surgeon may not always be able to operate at the same hospital. Operating at multiple hospitals may give more scheduling flexibility to the surgeon. More time sensitive surgeries may also require the first operating room available, regardless of the hospital. In addition, some surgeons may want to operate at different hospitals to access different or more patients. For example, a surgeon from the South side of Chicago may find valuable to operate in a hospital in the North side to be able to reach North side patients. Such flexibility does not come without potential costs, since surgeons have to get used to different practices and teams for example.

Limited scope for selection into treatment by hospitals and surgeons. Treatment decisions are usually made by a cardiologist and their patient prior to referral to the cardiothoracic surgeon when CABG surgery is chosen (Mukamel, Weimer, and Mushlin, 2006). Cardiologists who treat coronary artery disease manage the course of treatment for their patients. Alternative treatments include management with drugs such as beta-blockers or statins for example and percutaneous coronary intervention (PCI), a less invasive intervention that consists in inserting a stent into a narrowed artery to widen it. While less invasive, PCI may require more subsequent treatment. If a surgical treatment is chosen, the cardiologist refers the patient to an interventional cardiologist for PCI or to a cardiothoracic surgeon for CABG surgery.

There is also limited scope for selection into treatment for patients by hospitals. CABG surgery is an elective surgery rarely performed in an emergency setting since it is the most invasive treatment option. While cardiologists may refer patients to cardiothoracic surgeons within the same hospital, cardiothoracic surgeons tend to operate at multiple hospitals and to decide jointly with their patients at which hospital to operate (Wilson, Woloshin, and Schwartz, 2007). In other words, it is hard for cardiologists to select into treatment their patients based on hospitals' comparative advantages.

1.9.2 Data

Matching professional fees to hospital stays. To match an operating surgeon to the hospital where the surgery took place, I match MedPAR claims, that are at the hospital stay level and consequently identify the patient-hospital pairs, to Carrier claims, that identify the patient-operating surgeon pairs.[24]

Using the 20% sample of professional fees - the Carrier files - for years 2011 to 2017 included, I identify CABG surgery using Healthcare Common Procedure Coding System (HCPCS) codes available at the claim line level. These codes identify the task that is billed for. I use HCPCS codes 33510 to 33536 to identify claims relative to CABG surgery. Codes 33510 to 33516 indicate CABG with venous grafting only, codes 33533 to 33536 indicate CABG with arterial grafting. Codes 33533 to 33536 can be combined with add-on codes 33517 to 33523 to indicate combined arterial and venous grafting. I identify the operating surgeon as the surgeon reported as the performing physician for thsi specific claim line.

Since the identity of the hospital where the service is performed is not reported in this file, I match these claim lines to the full sample of Medicare hospital stays using the MedPAR data. I do so using claim dates and patient identifiers following Chen (2021): I match a Carrier claim line to a hospital stay when the Carrier claim date is within the admission and discharge date of the hospital stay for this patient in the MedPAR data. As indicated in Table 1.1, I am able to match the claims of more than 95% patients identified in the Carrier file.

National Plan and Provider Enumeration System (NPPES) data. The NPPES was created by CMS to assign a unique provider identifier, the National Provider Identifier (NPI), to healthcare providers, including physicians and hospitals. All healthcare providers billing Medicare are required to obtain such an identifier. These files include information at the NPI level such as physician specialties or primary practice locations.

Doctors and clinicians CMS data. This data comes from the online Medicare enrollment management system named provider, enrollment, chain, and ownership system (PECOS). It

24. The patient-hospital-operating surgeon triplets could be directly identified from the CMS Inpatient claim line files, which I did not have access to.

includes various information at the provider level; I notably use the year of graduation from medical school at the physician level in this data.

Hospital general information and complications and deaths datasets. This data contains information for all hospital registered with Medicare, including notably their ownership type and quality measures such as 30-day risk-adjusted mortality for several conditions and procedures.

CMS provider of services - hospitals files. This data is gathered as part of the CMS provider certification process. It includes additional hospital characteristics such as the number of beds, the number of operating rooms, and some measures of employment by category of worker.

Medicare provider utilization and payment data - public use files. This data at the national level contains the total amount billed to Medicare nationally or by state for each procedure (HCPCS) code. The provider-level data reports the amount billed to Medicare at the provider level for each procedure code. In both datasets, entries with 10 patients or less are redacted.

1.9.3 Risk-adjusted survival at the patient level

I compute risk-adjusted survival at the patient level using the difference between observed survival and predicted survival using a logit model. In particular, the predicted probability of survival for each patient is estimated using

$$\ln\left(\frac{Pr[Y_{ijht}=1|X_{it}]}{1-Pr[Y_{ijht}=1|X_{it}]}\right)=\alpha+\beta X_{it}$$

where X_{it} include patient covariates included in Table 1.2 - excluding outcomes - and year fixed effects.

I compute the risk-adjusted survival at the patient level such that

$$RASR_{it} = y_{ijht} - \hat{pr}_{ijht} + \bar{y}$$

where y_{ijht} is the observed survival for patient i, $\hat{pr}_{ijht}$ is the predicted survival from the logit model, and $\bar{y}$ is the average observed survival in the sample, used as scaling.

1.9.4 Empirical Bayes shrinkage of individual hospital and surgeon fixed effects

To illustrate the dispersion of the hospitals' and surgeons' fixed effect, I recover the average per provider using the 30-day risk-adjusted survival (RASR) as delineated in Appendix 1.9.3 in a simple regression using fixed effects.

Because of measurement error in these fixed effects, especially for low volume hospitals and surgeons, measuring the standard deviation across providers using these estimated fixed effects may overestimate the standard deviation in the "true" fixed effects. To address it, I use the standard empirical Bayes shrinkage technique that "shrinks" noisy fixed effects toward the mean.

Assume the estimated fixed effects are estimated with error such that

$$\hat{\psi}_h = \psi_h + e_h$$

where ψ_h is the "true" fixed effect and e_h is the measurement error of the estimated fixed effect. Note that the measurement error is assumed to be independent of the "true" fixed effect ψ_h.

Assuming e_h are independent such that

$$e_h \sim N(0, \pi_h^2)$$

where π_h^2 is the variance of the measurement error. This gives the distribution of the estimated fixed effect conditional on the true fixed effect and measurement error variance

$$\hat{\psi}_h | \psi_h, \pi_h^2 \sim N(\psi_h, \pi_h^2)$$

Assume a prior distribution for the true effect such that

$$\psi_h | x_h, \lambda, \sigma^2 \sim N(\lambda x_h, \sigma^2)$$

where σ^2 is the variance of the true fixed effect, common to all hospitals h, and λx_h is the underlying mean as a linear function of hospitals' covariates.

From Bayes' rule, we obtain

$$\psi_h | x_h, \lambda, \sigma^2, \pi_h^2, \hat{\psi}_h \sim N(b_h \hat{\psi}_h + (1 - b_h)\lambda x_h, b_h \pi_h^2)$$

with

$$b_h = \frac{\sigma^2}{\pi_h^2 + \sigma^2}$$

The empirical Bayes-adjusted fixed effects correspond to the mean of the posterior such that

$$\psi_h^{EB} = \frac{\sigma^2}{\pi_h^2 + \sigma^2} \hat{\psi}_h + \frac{\pi_h^2}{\pi_h^2 + \sigma^2} \lambda x_h$$

This last equation illustrates how the empirical Bayes shrinkage operates: the larger the variance of the measurement error for a hopistal π_h^2 is, the more weight is given to the underlying mean against the estimated fixed effect for this hospital. In other words, noisier fixed effect estimates are "shrunk" toward the underlying mean.

We need estimates for π_h^2, σ^2, and λx_h. I will assume $\lambda x_h = \lambda$, i.e., a constant for all hospitals, so that $\hat{\lambda}$ corresponds to the average survival across hospitals in the sample. I use the square of the standard errors for the estimated fixed effects as the estimate for π_h^2. Finally, I recover an estimate for $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = \frac{\sum_h w_h \left(\frac{n_h}{n_h - 1} (\hat{\psi}_h - \hat{\lambda})^2 - \hat{\pi}_h^2 \right)}{\sum_h w_h}$$

where n_h corresponds to the number of hospitals, and w_h are weights for each hospital such that $w_h = \frac{1}{\hat{\pi}_h^2 + \hat{\sigma}^2}$. More weight is given to hospitals with less measurement error. This correponds to the algorithm detailed in the Appendix of Chandra et al. (2016*a*) based on Morris (1983). $\hat{\sigma}^2$ corresponds to the estimate of the standard deviation of the "true" fixed effects, reported in Figure 1.2.

1.9.5 K-means algorithm

The k-means clustering algorithm aims at best capturing the unobserved heterogeneity across surgeons and hospitals. In particular, the k-means algorithm partitions the H hospitals in the sample into a pre-specified number of groups K by solving the following weighted k-means problem:

$$\underset{\tilde{F}, k(1), \ldots, k(H)}{\operatorname{argmin}} \sum_{h=1}^{H} n_h ||f(h) - \tilde{F}(k(h))||^2$$

where $f(h)$ is the average risk-adjusted survival at hospital h, $k(1), ..., k(H)$ is the partition of hospitals into K types, n_h the number of patients treated at hospital h, and $\tilde{F} = (\tilde{F}(1)', ..., \tilde{F}(K)')'$ are vectors where $\tilde{F}(k)$ corresponds to the mean of $f(h)$ when $k(h) = k$. The types are revealed by the clusters, such that the sum of the squared distance between hospitals' mean risk-adjusted survival in that cluster and the centroid of the cluster is minimized. The intra-type variance in mean patient risk-adjusted survival is minimized. The number of hospitals per cluster does not need to be equal.

I follow the same strategy to partition the J surgeons into a pre-specified number of

‘’“types” L such that

$$\underset{\tilde{A},l(1),...,l(J)}{\text{argmin}} \sum_{j=1}^{J} n_j ||a(j) - \tilde{A}(l(j))||^2$$

where $a(j)$ is the average risk-adjusted survival for patients treated by surgeon j, $l(1), ..., l(J)$ is the partition of surgeons into L types, n_j the number of patients treated by surgeon j, and $\tilde{A} = (\tilde{A}(1)', ..., \tilde{A}(L)')'$ are vectors.

1.9.6 Monte-Carlo simulations

I investigate the impact of using grouped fixed-effects in place of individual fixed-effects under alternative sorting regimes using Monte-Carlo simulations. Note that I maintain the assumption that the production function of survival is monotonic in surgeon and hospital quality. The monotonicity assumption requires the production function of survival to be monotonically increasing/decreasing in the hospital type conditional on a surgeon type, and vice versa. Production functions where hospitals’ and surgeons’ fixed effects enter linearly or multiplicatively are monotonically increasing.[25] This assumption is reasonable when examining output or quality measures directly, and I maintain it in the paper and in the exercises below.

Assuming a logit production function and positive assortative matching, I am able to accurately recover hospitals’ and surgeons’ types as reported in Appendix Table 1.12. For both surgeons’ and hospitals’ types, the correlation between the true and the predicted grouped fixed-effects are above 0.9, and the value of the true value of the covariance between fixed effects can be accurately recovered from the group fixed-effects.

With negative assortative matching, average risk-adjusted survival does not allow to correctly identify hospitals’ and surgeons’ types. As reported in Appendix Table 1.13, a large amount of hospitals and surgeons are misclassified so that the correlation between the

25. Relaxing the monotonicity assumption would lead to identification issues similar to the ones raised by negative assortative matching.

true and the predicted grouped fixed-effects are way below 0.9. This misclassification results in an estimated covariance of zero whether the negative assortative matching is weak or strong. Increasing the number of groups for both surgeons and hospitals allows to recover the direction of sorting, but estimates of the covariance for the fixed effects converge to the true covariance relatively slowly.

These results indicates that, assuming monotonicity of the underlying production function, positive assortative matching can be accurately identified using k-means clustering, even with a small number of k-means groups. However, I cannot separate the absence of sorting from negative assortative matching using k-means clustering with a low number of k-means groups.

1.9.7 Deriving the control function

Recall the production function of survival for patient i treated by surgeon j in hospital h as

$$Y^*_{ijht} = g(\alpha_{l(j)}, \psi_{k(h)}, X_{it}) + \epsilon_{ijht}$$

where $\alpha_{l(j)}$ and $\psi_{k(h)}$ are respectively the unobserved heterogeneity of the surgeon and hospital, X_{it} are patient observables such as age, gender, and underlying health, and ϵ_{ijht} are unobserved health shocks. I abstract away from year fixed effects in the derivations that follow.

The observed survival Y_{ijht} for patient i treated by surgeon j in hospital h is

$$Y_{ijht} = D_{ijht} Y^*_{it}$$

and

$$D_{ijht} = 1\{u_{ih} \geq u_{ih'}, \forall h'\}$$
$$\text{with } u_{ih} = \delta_h - \tau \ln(d_{ih}) + \eta_{ih}$$

where u_{ih} is the utility from patient i treated by surgeon j from getting the surgery at hospital h, δ_h is the perceived quality of hospital h, on which all patients and surgeons agree within a market, and d_{ih} is the distance between the patient ZIP code and the hospital ZIP code. I assume η_{ih} are type-I extreme value error terms.

Denote the choice of hospital by patient i as D_i which takes values $(1, ..., H)$, so that $D_i = h$ indicates that patient i treated by surgeon j goes to hospital h. Following Dubin and McFadden (1984), I impose the following linear structure to the conditional expectation of ϵ_{ijht}:

$$\mathbb{E}[\epsilon_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}, \eta_{i1}, ..., \eta_{iH}, D_i = h] = \sum_{s \in \mathcal{H}} \phi_s(\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta)$$

where μ_η is the Euler constant (mean of logit errors) and $\mathcal{H}$ the set of hospitals. Recall that ϕ_s is hospital-specific and identifies selection into hospitals, while φ is choice-specific and identifies selection on gains.

The expected survival conditional on the fixed effects, patient observables X_{it}, the choice of hospital D_i, and the unobserved logit shocks $\eta_{i1}, ..., \eta_{iH}$ can be written as

$$\begin{aligned}\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}, \eta_{i1}, ..., \eta_{iH}, D_i = h] =& \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} \\ &+ \sum_{s \in \mathcal{H}} \phi_s(\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta)\end{aligned}$$

Integrating over the unobserved logit shocks $\eta_{i1}, ..., \eta_{iH}$, we obtain

$$\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} + \sum_{s \in \mathcal{H}} \phi_s \mathbb{E}[\eta_{is} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] + \varphi \mathbb{E}[\eta_{ih} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = h]$$

To derive the control functions, note that

$$\mathbb{E}[\eta_{ih} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] - \delta_h + \lambda \ln d_{ih} - \mu_\eta$$

Using Small and Rosen (1981), we have

$$\mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] + \mu_\eta$$

so that

$$\begin{aligned} \mathbb{E}[\eta_{ih} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] &= \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] - \delta_h + \lambda \ln d_{ih} + \mu_\eta \\ &= \ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] - \delta_h + \lambda \ln d_{ih} \\ &= \ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] - \ln \Big[\exp(\delta_h + \lambda \ln d_{ih}) \Big] \\ &= \ln \Big[\frac{\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is})}{\exp(\delta_h + \lambda \ln d_{ih})} \Big] \\ &= - \ln \Big[\frac{\exp(\delta_h + \lambda \ln d_{ih})}{\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is})} \Big] \\ &= - \ln \hat{p}_{ih} \end{aligned}$$

with $\hat{p}_{ih}$ the predicted probability for i to choose hospital h obtained from the demand model.

Now, assuming the choice of hospital is $s \neq h$, we have

$$\mathbb{E}[\eta_{ih} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = s) = \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] - \delta_h + \lambda \ln d_{ih} - \mu_\eta$$

Use

$$\begin{aligned}
&\mathbb{E}[u_{ih}] = \\
&\mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] Pr(D_i = h) + \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] Pr(D_i \neq h) \\
\iff &\mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] = \frac{\mathbb{E}[u_{ih}] - \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] Pr(D_i = h)}{Pr(D_i \neq h)} \\
\iff &\mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] = \frac{\mathbb{E}[u_{ih}] - \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = h] Pr(D_i = h)}{1 - Pr(D_i = h)} \\
\iff &\mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] \\
&= \frac{\delta_h - \lambda \ln d_{ih} + \mu_\eta - (\ln \left[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \right] + \mu_\eta) Pr(D_i = h)}{1 - Pr(D_i = h)}
\end{aligned}$$

Denote $\hat{p}_{ih} = Pr(D_i = h)$ and substitute such that

$$
\begin{aligned}
&\mathbb{E}[\eta_{ih} - \mu_\eta | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] = \mathbb{E}[u_{ih} | \ln d_{i1}, ..., \ln d_{iH}, D_i = s] - \delta_h + \lambda \ln d_{ih} - \mu_\eta \\
&= \frac{\delta_h - \lambda \ln d_{ih} + \mu_\eta - \Big(\ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] + \mu_\eta \Big) \hat{p}_{ih}}{1 - \hat{p}_{ih}} - \delta_h + \lambda \ln d_{ih} - \mu_\eta \\
&= \frac{\delta_h - \lambda \ln d_{ih} + \mu_\eta - \Big(\ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] + \mu_\eta \Big) \hat{p}_{ih}}{1 - \hat{p}_{ih}} - \frac{(1 - \hat{p}_{ih})(\delta_h - \lambda \ln d_{ih} + \mu_\eta)}{(1 - \hat{p}_{ih})} \\
&= \frac{\delta_h - \lambda \ln d_{ih} + \mu_\eta - (1 - \hat{p}_{ih})(\delta_h - \lambda \ln d_{ih} + \mu_\eta) - \Big(\ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] + \mu_\eta \Big) \hat{p}_{ih}}{1 - \hat{p}_{ih}} \\
&= \frac{\hat{p}_{ih} \Big(\delta_h - \lambda \ln d_{ih} + \mu_\eta - \ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] - \mu_\eta \Big)}{1 - \hat{p}_{ih}} \\
&= \frac{\hat{p}_{ih}}{1 - \hat{p}_{ih}} \Big(\ln \Big(\exp(\delta_h - \lambda \ln d_{ih}) \Big) - \ln \Big[\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is}) \Big] \Big) \\
&= \frac{\hat{p}_{ih}}{1 - \hat{p}_{ih}} \ln \Big(\frac{\exp(\delta_h - \lambda \ln d_{ih})}{\sum_{s=1}^{H} \exp(\delta_s - \lambda \ln d_{is})} \Big) \\
&= \frac{\hat{p}_{ih}}{1 - \hat{p}_{ih}} \ln \hat{p}_{ih}
\end{aligned}
$$

Therefore, the control function can be written as:

$$
\theta_{is}(h) = \begin{cases} -\ln \hat{p}_{is} & \text{if } s = h \\ \frac{\hat{p}_{is}}{1 - \hat{p}_{is}} \ln \hat{p}_{is} & \text{if } s \neq h \end{cases}
$$

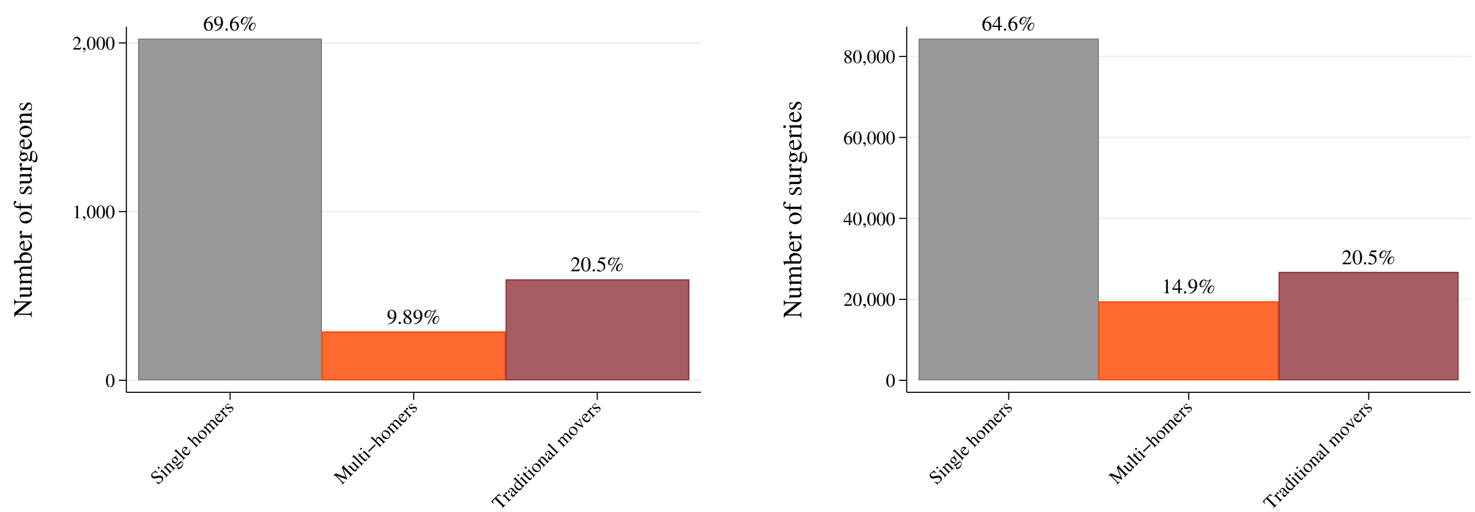
Note that the control function is positive when $s = h$ but negative otherwise since $\ln \hat{p}_{is} < 0$ with $0 < \hat{p}_{is} < 1$. This delivers the following estimating equation, with $\theta_{is}(h)$ as defined above:

$$\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] =$$
$$\alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} + \sum_{s \in \mathcal{H}} \phi_s \theta_{is}(h) + \varphi \theta_{ih}(h)$$

1.10 Additional exhibits

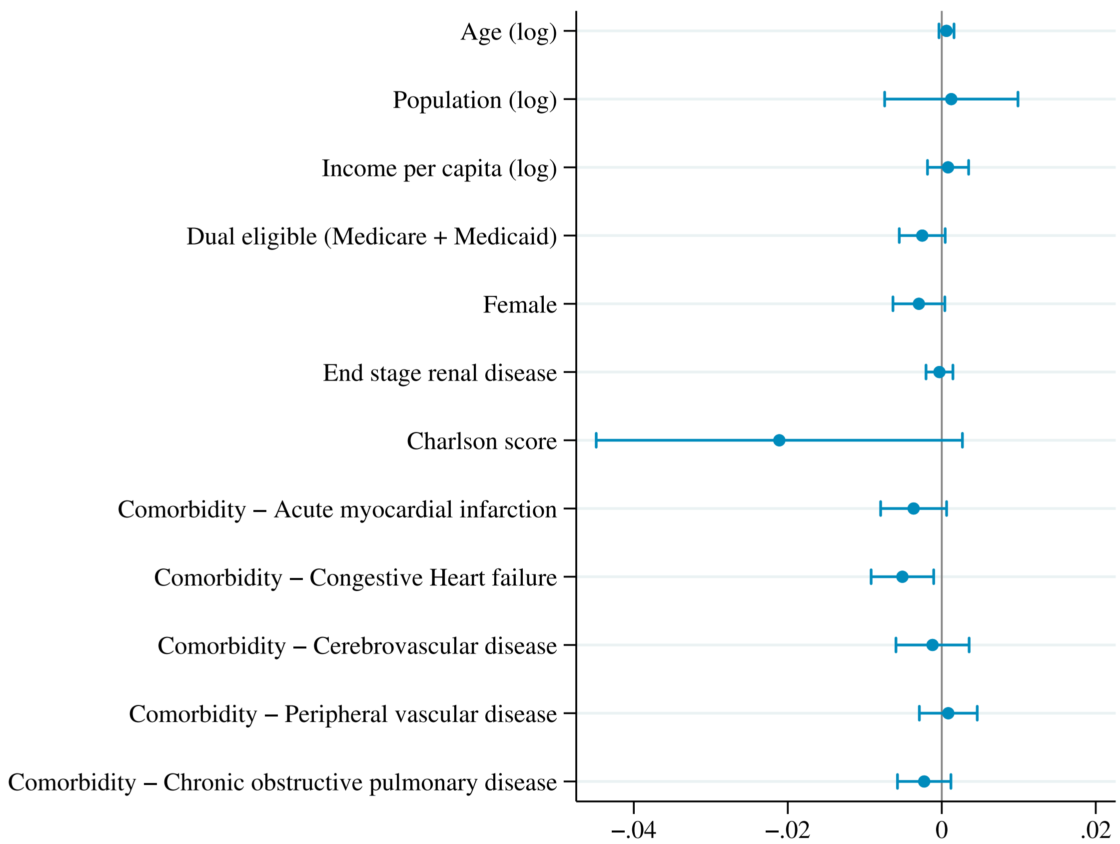
Figure 1.15: Proportion of “single-homers,” “multi-homers,” and “traditional movers” using hospital groups

(a) Proportion of surgeons (b) Proportion of surgeries



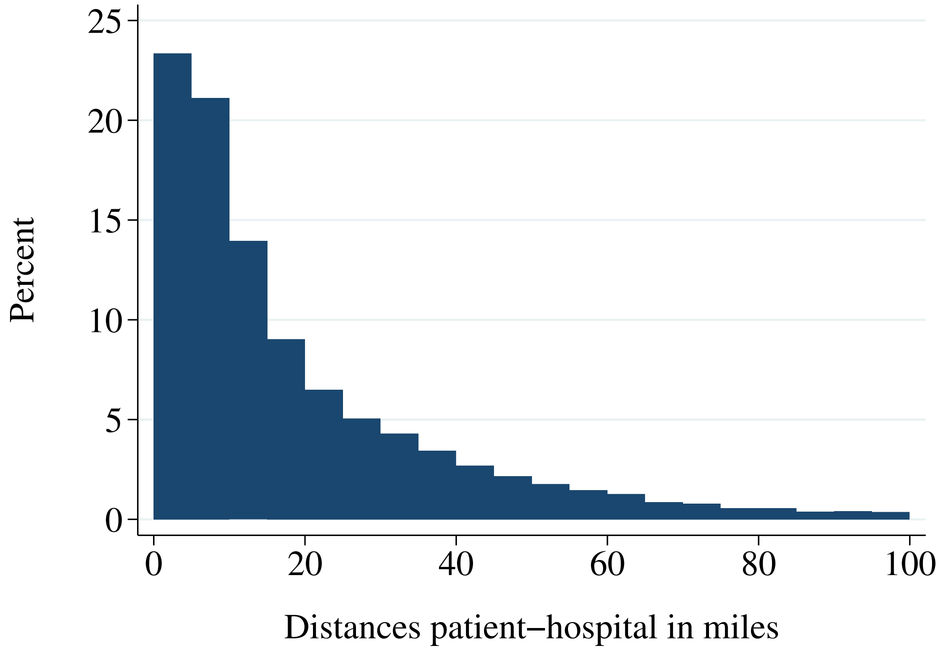
Notes: The fraction of surgeons observed at more than one hospital *group* falls to about 30%, compared to Figure 1.1 in which the fraction of surgeons observed at more than one hospital is close to 40%. “Multi-homers” are defined as surgeons who performed CABG surgeries at more than one hospital *group* within a year for four years of more in the sample. “Traditional movers” are surgeons who performed CABG surgeries at more than one hospital *group* in one, two, or three years in the sample. “Single homers” include surgeons who only performed CABG surgeries at a unique hospital *group* in the sample. K-means clustering is performed using average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.16: No evidence that hospitals systematically triage sicker patients into higher-survival surgeons



Notes: This graph examines the existence of "triaging" within hospitals, i.e., whether higher-survival surgeons tend to operate on sicker patients within a hospital using patient observables. All coefficients are close to zero and statistically insignificant, suggesting a limited role for triaging into surgeons using patient observables. Coefficients reported in this graph correspond to the estimated $\hat{\beta}$ from the regression $x_{ihj} = \alpha + \beta \text{rank}_{l(j)} + \lambda_h + \epsilon_{ijh}$. x_{ijh} correspond to the covariates of patients treated by surgeon j at hospital h, and λ_h are individual hospital fixed effects. The ranks of surgeon groups are computed as the rank in predicted risk-adjusted survival based the model from equation (1.6) assuming each hospital group is equally likely for each surgeon group. Surgeon and hospital groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Confidence intervals displayed are 95% confidence intervals constructed using clustered standard errors at the hospital level.

Figure 1.17: Distribution of distances between patients and their chosen hospital

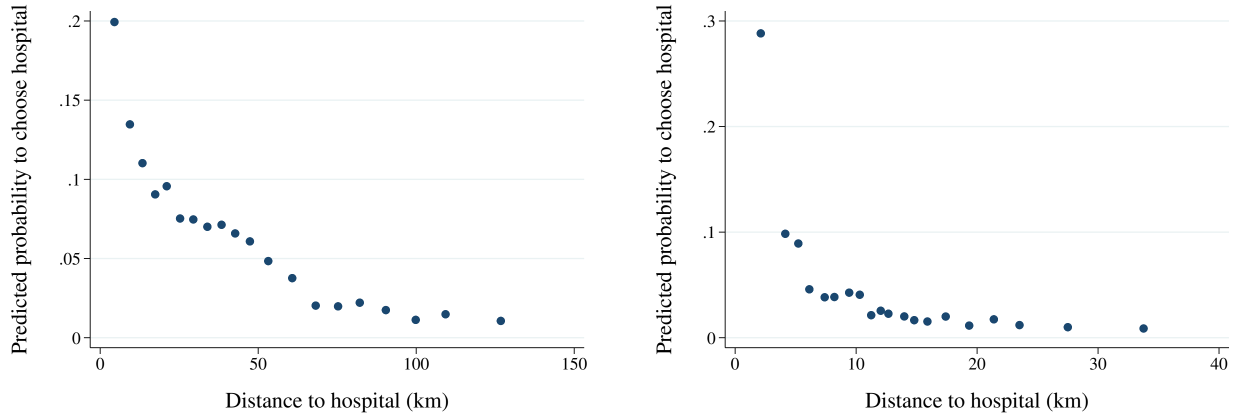


Notes: This graph depicts the distribution of distances between the patient's residential ZIP code and their chosen hospital's ZIP code, for patients treated at hospitals within their residential hospital referral region (HRR). 21% of patients get CABG surgery outside of their HRRs in the sample. The average distance to hospitals is 18.7 miles for patients treated within their residential HRR: 53% of patients are within 20 miles of the hospital and 7% of patients are within the same ZCTA as the hospital. Hospital ZIP codes come from the 2017 National Plan and Provider Enumeration System (NPPES) data, and beneficiary ZIP codes from the Medicare Beneficiary Research Identifiable Files. The definition of hospital referral regions (HRRs) follows the definition of the Dartmouth Atlas Project. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.18: Distance to the hospital is a strong predictor of hospital choice within HRRs: Boston and Chicago

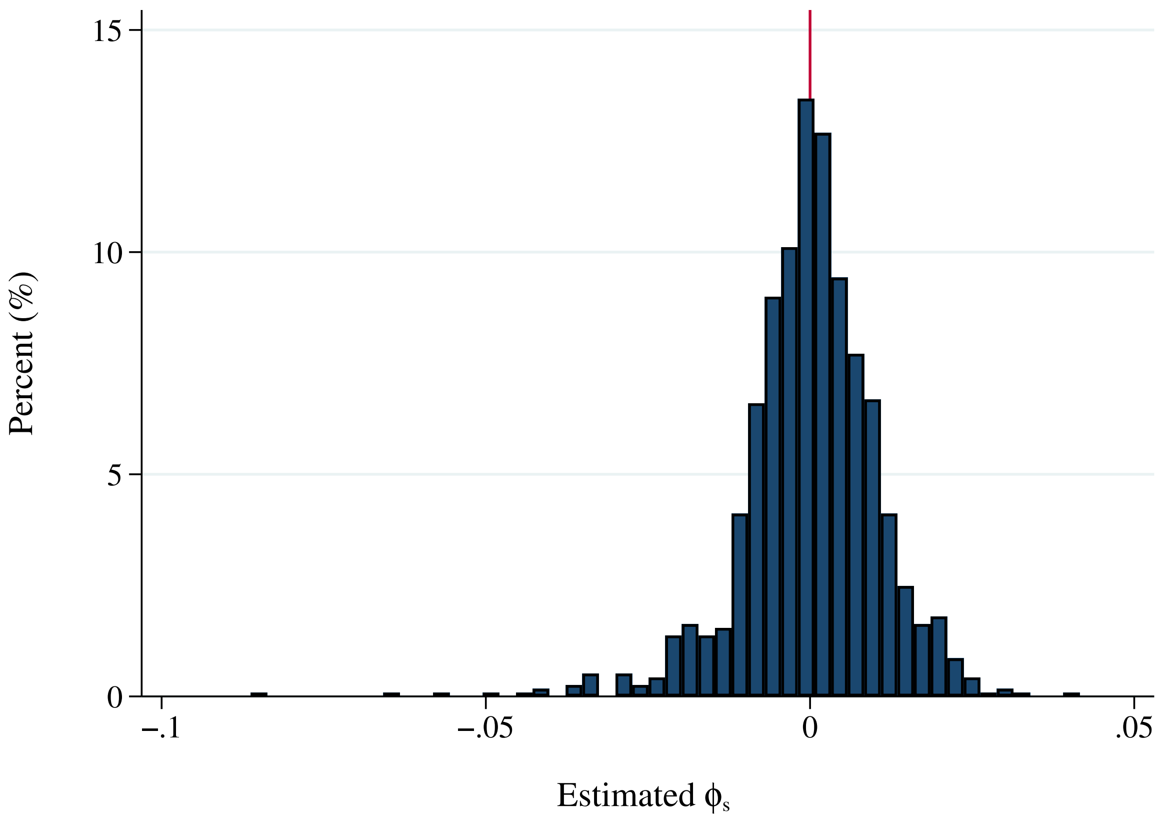
(a) Boston HRR

(b) Chicago HRR



Notes: These graphs depict the relationship between the predicted probabilities to choose a hospital using the demand model delineated in equation (1.7), estimated HRR by HRR, and distance between the patient and the hospital ZIP codes. Only predicted probabilities for hospitals within a patient's residential HRR are included. The graphs sumarize this relationship using a binned scatter plot with twenty equally sized bins. Hospital ZIP codes come from the 2017 National Plan and Provider Enumeration System (NPPES) data, and beneficiary ZIP codes from the Medicare Beneficiary Research Identifiable Files. The definition of hospital referral regions (HRRs) follows the definition of the Dartmouth Atlas Project. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

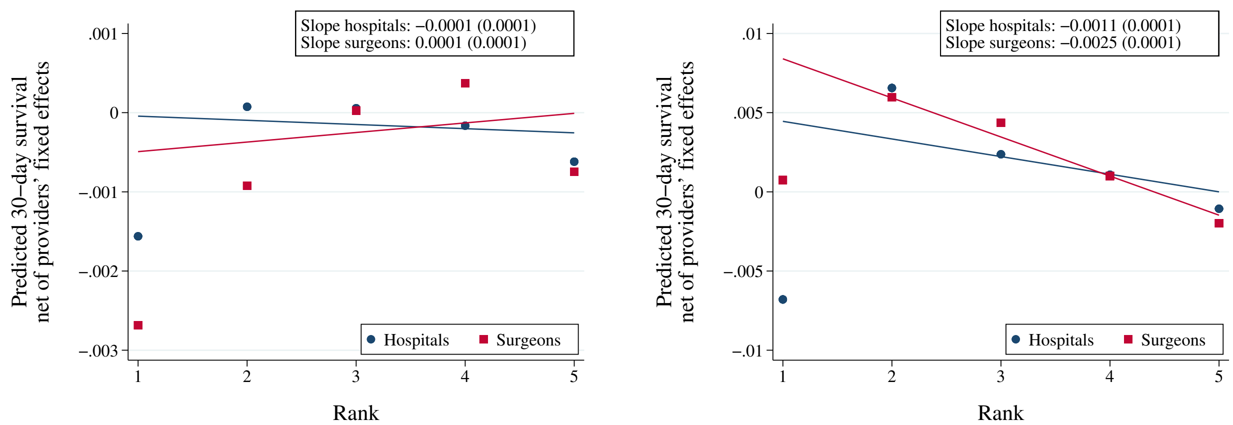
Figure 1.19: Distribution of estimated control function parameters $\hat{\phi}_s$ across hospitals



Notes: This graph shows the distribution of the estimated control function parameters $\hat{\phi}_s$ from equation (1.11), with s denoting a specific hospital. When the estimated coefficient is negative, sickest patients tend to select into that specific hospital. Conversely, when the estimated coefficient is positive, healthier patients tend to select into that specific hospital. Results suggest that some hospitals face adverse selection while other hospitals face advantageous selection. Professional fees come from the Medicare 20% carrier Research Identifiable Files, hospital stays from the Medicare MedPAR Research Identifiable Files, and beneficiary information from the Medicare Beneficiary Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.20: Relationship between predicted survival net of provider fixed effects and provider rankings

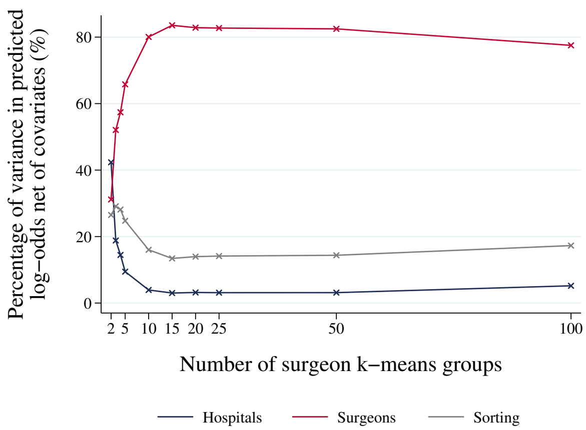
(a) Including patient covariates only (b) Including control function arguments



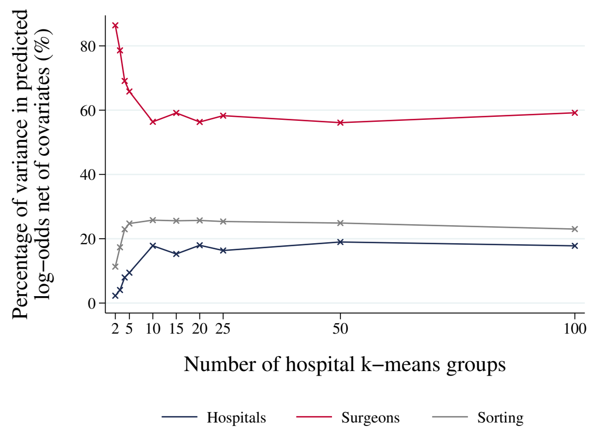
Notes: This figure reports the relationship between predicted 30-day survival net of provider fixed effects and the rank of providers from the model in equations (1.6) and (1.11). There is no systematic relationship between predicted survival net of provider fixed effects and the ranking of their provider when only including patient covariates. Leveraging distance to hospitals as an excluded instrument to identify selection on unobservables, I find evidence for systematic adverse selection into provider rankings for both surgeons and hospitals since the relationship is negative and statistically significant for surgeons and hospitals. Adverse selection appears to be stronger into surgeons. The predicted 30-day survival net of provider fixed effects is calculated as $\hat{p}_{it} = \sum_p \hat{\beta}_p X_{it,p} + \hat{\gamma}_t$ and $\hat{p}_{it} = \sum_p \hat{\beta}_p X_{it,p} + \hat{\gamma}_t + \sum_{s \in \mathcal{H}} \hat{\phi}_s \hat{\theta}_{is}(h) + \hat{\psi}\hat{\theta}_{ih}(h)$ estimated from equations (1.6) and (1.11) respectively. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. The rank of providers is calculated based on the predicted risk-adjusted survival for the provider's group when all hospitals or surgeons groups are equally likely. Standard errors displayed are robust standard errors. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Figure 1.21: Robustness of the variance decomposition to alternative numbers of k-means groups

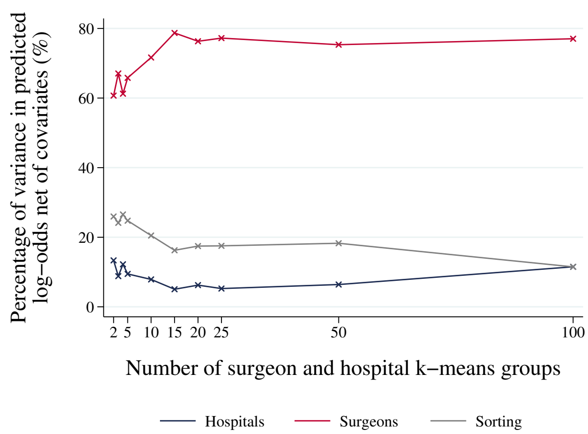
(a) Vary L, K=5



(b) Vary K, L=5



(c) Vary K and L jointly



Notes: These graphs show results from the variance decomposition of predicted log-odds when varying the number of k-means groups for hospitals, surgeons, and both jointly. Results are robust to alternative number of ex-ante groups specified for k-means clustering. K and L denote the number of ex-ante groups specified for k-means clustering for hospitals and surgeons respectively. The variance decomposition comes from the decomposition of predicted log-odds as delineated in equation (1.15). The hospital component corresponds to $Var(\hat{\psi}_{k(h)})$, the surgeon component corresponds to $Var(\hat{\alpha}_{l(j)})$, and the sorting component corresponds to $cov(\hat{\psi}_{k(h)}, \hat{\alpha}_{l(j)})$. They are expressed as a percentage of the predicted log odds of 30-day survival net of covariates $Var(\ln\left(\frac{\hat{p}_{it}}{1-\hat{p}_{it}}\right) - \sum_s \hat{\beta}_s X_{it,s})$, where $\hat{p}_{it}$ corresponds to the predicted 30-day survival from the logit model. K-means clustering is performed using average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.9: Characteristics of patients for “single-homers,” “multi-homers,” and “traditional movers”

	Single homers	Multi homers	Other movers	Differences		
	(1)	(2)	(3)	(2)-(1)	(3)-(1)	(2)-(3)
Age	72.44	72.22	72.14	-0.22***	-0.30***	0.08
	(8.23)	(8.09)	(8.28)	(0.06)	(0.05)	(0.07)
Dual eligible (Medicaid + Medicare)	0.16	0.20	0.17	0.04***	0.01***	0.02***
	(0.37)	(0.40)	(0.38)	(0.00)	(0.00)	(0.00)
Income per capita (USD, x1,000)	33.75	32.53	33.32	-1.22***	-0.43***	-0.79***
	(14.09)	(13.69)	(13.84)	(0.10)	(0.09)	(0.12)
ZIP code population (x1,000)	24.38	27.79	25.33	3.40***	0.95***	2.45***
	(18.65)	(19.84)	(18.86)	(0.14)	(0.12)	(0.16)
Female	0.30	0.31	0.30	0.01	-0.00	0.01**
	(0.46)	(0.46)	(0.46)	(0.00)	(0.00)	(0.00)
ESRD	0.04	0.06	0.05	0.01***	0.00***	0.01***
	(0.20)	(0.23)	(0.22)	(0.00)	(0.00)	(0.00)
Charlson score	3.38	3.43	3.46	0.06***	0.08***	-0.02
	(2.65)	(2.69)	(2.67)	(0.02)	(0.02)	(0.02)
30-day mortality	0.05	0.05	0.05	0.00	0.00*	-0.00
	(0.21)	(0.22)	(0.22)	(0.00)	(0.00)	(0.00)
60-day mortality	0.06	0.06	0.06	0.01***	0.00**	0.00
	(0.23)	(0.24)	(0.24)	(0.00)	(0.00)	(0.00)
Length of stay	10.35	10.21	10.34	-0.14***	-0.01	-0.12**
	(7.81)	(6.77)	(7.33)	(0.06)	(0.05)	(0.06)
Years since medical school graduation, as of 2010	23.72	24.53	21.09	0.81***	-2.63***	3.44***
	(9.15)	(9.24)	(8.36)	(0.07)	(0.06)	(0.07)
Number of patients	72,842	25,122	32,880			
Number of surgeons	1,805	369	737			

Notes: “Multi-homers” are defined as surgeons who performed CABG surgeries at more than one hospital within a year for four years of more in the sample. “Traditional movers” are surgeons who performed CABG surgeries at more than one hospital in one, two, or three years in the sample. “Single homers” include surgeons who only performed CABG surgeries at a unique hospital in the sample. “Multi-homers” and “traditional movers” tend to operate on younger, sicker, lower income patients residing in more populated ZIP codes. “Traditional movers” have on average graduated between 2 and 4 years earlier than “multi-homers” and “single-homers.” Tests for differences in means across types of surgeons are independent t-tests. Statistical significance: *** 2.5% , ** 5%, and * 10%. Medical school graduation year comes from the 2017 doctors and clinicians CMS public use dataset. Income per capita and population come from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Professional fees come from the Medicare 20% carrier Research Identifiable Files, hospital stays from the Medicare MedPAR Research Identifiable Files, and beneficiary information from the Medicare Beneficiary Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.10: Activity split across hospitals for "multi-homers"

	Percentage of surgeon's activity		
Number of hospitals in a year	2	3	4 or more
Top 1 hospital	73.1	57.4	48.4
Top 2 hospital	26.9	27.9	23.2
Top 3 hospital onward	-	14.8	28.4
Number of surgeons	352	155	35

Notes: Only "multi-homers," i.e., surgeons who performed CABG surgeries at more than one hospital within a year for four years of more in the sample, in years when they performed CABG surgeries at more than hospital are included. A surgeon's activity is measured as the total number of CABG surgeries performed by that surgeon in a given year in the sample. The share of a surgeon's activity at other hospitals than their top choice is substantial. "Multi-homers" practicing at two hospitals in a given year perform on average 73.1% of their CABG surgeries at one hospital and the remaining 26.9% CABG surgeries at a second hospital. For surgeons practicing at three or more different hospitals in a given year, more than 40% of their CABG surgeries are performed at other hospitals than their top choice. "Multi-homers" in the sample practice at two to seven different hospitals witin a year. The top 1 hospital for a surgeon is the hospital at which the surgeon performed the largest share of their CABG surgeries in a given year. The top 2 hospital is the hospital at which the surgeon performed the second largest share of their CABG surgeries in a given year. The top 3 hospital onward include hospitals at which the surgeon performed all their other CABG surgeries in a given year. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.11: Variance in 30-day survival within and across providers

		Observed		Risk-adjusted (RASR)	
		Hospitals	Surgeons	Hospitals	Surgeons
Across					
	Amount	0.00078	0.00142	0.00078	0.00141
	Percentage of total	1.6	3.0	1.6	2.9
Within					
	Amount	0.04583	0.04519	0.04691	0.04627
	Percentage of total	98.3	96.9	98.3	97.0
Total		0.04661	0.04661	0.04769	0.04769

Notes: This table decomposes the total 30-day observed and risk-adjusted patient survival variance into across versus within providers variance. Risk-adjustment is performed by predicting 30-day survival using a logit model as delineated in Appendix 1.9.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.12: Monte Carlo simulation results, assuming positive assortative matching

Sorting parameter	K	Median number of patients per hospital (mean)	Median number of patients per surgeon (mean)	Classification: $corr(\psi_h, \widehat{\psi_{k(h)}})$ (mean)	Classification: $corr(\alpha_j, \widehat{\alpha_{l(j)}})$ (mean)	Covariance: $cov(\psi_h, \alpha_j)$ (mean)	Covariance: $cov(\widehat{\psi_{k(h)}}, \widehat{\alpha_{l(j)}})$ (mean)
0.1	5	95.2	12.3	0.90	0.90	0.18	0.24
0.1	10	95.2	12.3	0.92	0.92	0.18	0.25
0.1	30	95.3	12.4	0.94	0.93	0.18	0.23
0.3	5	95.5	16.4	0.92	0.92	0.50	0.51
0.3	10	95.7	16.6	0.93	0.94	0.50	0.58
0.3	30	95.7	16.8	0.93	0.94	0.50	0.52

Notes: Assuming positive assortative matching, the correlation between true individual fixed effects and estimated group fixed effects is above 0.90. This correlation gets larger with stronger positive assortative matching, and with more ex-ante specified k-means groups. The network of surgeon-hospital pairs is simulated for 10,000 patients. The true production function is assumed to be a logit function of individual hospital and surgeon fixed effects. K-means clustering is performed on average observed survival in the simulated data with alternative number of groups, equal for hospitals and surgeons, specified by K. Estimated parameters are estimated from a two-way logit model. The means are calculated across 500 simulations.

Table 1.13: Monte Carlo simulation results, assuming negative assortative matching

				Classification		Covariance	
Sorting parameter	K	Median number of patients per hospital (mean)	Median number of patients per surgeon (mean)	$corr(\psi_h, \widehat{\psi_{k(h)}})$ (mean)	$corr(\alpha_j, \widehat{\alpha_{l(j)}})$ (mean)	$cov(\psi_h, \alpha_j)$ (mean)	$cov(\widehat{\psi_{k(h)}}, \widehat{\alpha_{l(j)}})$ (mean)
-0.1	5	95.3	12.3	0.84	0.84	-0.18	0.01
-0.1	10	95.5	12.3	0.87	0.87	-0.18	-0.00
-0.1	30	95.3	12.3	0.90	0.89	-0.18	-0.04
-0.1	40	95.4	12.4	0.91	0.90	-0.17	-0.05
-0.3	5	95.8	16.6	0.63	0.78	-0.50	0.01
-0.3	10	95.7	16.5	0.68	0.80	-0.50	-0.01
-0.3	30	95.9	16.6	0.74	0.85	-0.50	-0.05
-0.3	40	95.7	16.7	0.77	0.87	-0.50	-0.08

Notes: This table shows that classification error is larger with negative assortative matching, and that increasing the number of groups partially alleviates the negative bias on the estimated covariance. The stronger negative assortative matching is, the lower the correlation between true individual fixed effects and estimated group fixed effects. Classification error biases the estimated covariance upward toward zero, hence biasing against finding any significant sorting between surgeons and hospitals. The network of surgeon-hospital pairs is simulated for 10,000 patients. The true production function is assumed to be a logit function of individual hospital and surgeon fixed effects. K-means clustering is performed on average observed survival in the simulated data with alternative number of groups, equal for hospitals and surgeons, specified by K. Estimated parameters are estimated from a two-way logit model. The means are calculated across 500 simulations.

Table 1.14: Correlation of estimated surgeon group effects with external measures of surgeons' skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years since medical school graduation (as of 2010, log)	0.0006							-0.0008
	(0.0015)							(0.0015)
CABG volume in Medicare 2012-2017 (USD, log)		0.0017						0.0004
		(0.0008)						(0.0023)
CABG volume in Medicare 2012-2017 (frequency, log)			0.0023					0.0026
			(0.0009)					(0.0024)
Medicare volume 2012-2017 (USD, log)				0.0012				-0.0038
				(0.0008)				(0.0035)
Medicare surgical volume 2012-2017 (USD, log)					0.0015			0.0041
					(0.0007)			(0.0039)
ZIP code population (log)						-0.0004		-0.0003
						(0.0006)		(0.0007)
ZIP code median HH income (USD, log)							0.0027	0.0033
							(0.0018)	(0.0019)
Observations	2,592	2,720	2,720	2,720	2,720	2,910	2,910	2,456
R-squared	0.0001	0.0026	0.0034	0.0015	0.0022	0.0001	0.0008	0.0091

Notes: This table reports the point estimates and 95% confidence intervals from regression of the surgeon group estimates on surgeon-level covariates. Surgeon group estimates include the fixed effect with interactions as $\widehat{\alpha_l} + \frac{1}{K}\sum_k \widehat{\kappa_{lk}}$ from equation (1.6), i.e., weighting each interaction with each hospital group equally. Results are similar when using the estimated group effects from equation (1.11), but not statistically significant. Surgeon group estimates are positively correlated with surgeons experience in performing CABG within Medicare, measured in log-revenue or log-frequency, but also positively correlated with surgeons' surgical and overall experience, measured as surgical and total Medicare revenues respectively, but not statistically significant. However, the relationship with tenured experience–measured as the number of years since medical school graduation–is not statistically different from zero. The relationship with the median household income or total population in the surgeon's primary practice ZIP code is not statistically significant. Surgeons' Medicare revenues and frequency are calculated for years 2012 to 2017 from the CMS Medicare Physician & Other Practitioners file. Surgeon ZIP codes are the primary practice ZIP codes from the National Plan and Provider Enumeration System (NPPES) data in each year, except for 2013. Primary practice ZIP codes are missing for 2013. The median household income and total population is aggregated at the surgeon level as the mean across ZIP codes for years 2011-2012 and 2014-2017. Years since medical school graduation is calculated as of 2010 based on the medical school graduation in the CMS Doctors and Clinicians dataset. ZIP code level median household income and population comes from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Standard errors displayed are robust standard errors.

Table 1.15: Correlation of estimated hospital group effects with hospital-level covariates

	(1)	(2)	(3)	(4)	(5)	(6)
Number of beds (log)	-0.0008					
	(0.0011)					
Number of operating rooms (log)		-0.0024				
		(0.0013)				
Has a residency program			-0.0001			
			(0.0014)			
Is affiliated with a medical school				0.0006		
				(0.0014)		
Is non-profit					0.0025	
					(0.0014)	
Owned by gorvernment						-0.0038
						(0.0025)
Observations	1,167	555	1,167	1,167	1,167	1,167
R-squared	0.0004	0.0060	0.0000	0.0001	0.0025	0.0018

Notes: This table reports the point estimates and 95% confidence intervals from regression of the estimated hospital group effect on hospital-level covariates. Hospital group effects include the fixed effect with interactions as $\widehat{\psi_k} + \frac{1}{L}\sum_l \widehat{\kappa_{lk}}$ from equation (1.6), i.e., weighting each interaction with each surgeon group equally. Higher hospital group effects are positively correlated with larger hospitals in terms of number of beds, having a medical school affilation, being a non-profit hospital, and the number of registered nurse employed. There is statistically different from zero relationship with other available hospital covariates. The R^2 of the regression including all hospital covariates amounts to about 0.07, but reduces to less than 0.01 when including all hospitals covariates available for at least 1,000 hospitals. Results are similar when using the estimated group effects from equation (1.11). Hospital ownership is obtained from the CMS Hospital General Information dataset for 2017. ZIP code level median household income and population comes from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. All other hospital-level covariates come from the CMS provider of service dataset for 2017. Standard errors displayed are robust standard errors.

Table 1.15: Correlation of estimated hospital group effects with hospital-level covariates (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in ZIP code (log)	0.0006						
	(0.0007)						
Median income in ZIP code (USD, log)		0.0016					
		(0.0019)					
Number of physicians employed (log)			0.0001				
			(0.0004)				
Number of nurse practitioner employed (log)				-0.0015			
				(0.0006)			
Number of registered nurse employed (log)					-0.0005		
					(0.0007)		
Number of licensed nurses under contract (log)						-0.0006	
						(0.0006)	
Number of resident physicians (log)							0.0002
							(0.0006)
Observations	1,167	1,167	759	701	1,133	1,008	414
R-squared	0.0006	0.0006	0.0001	0.0078	0.0003	0.0009	0.0001

Notes: This table reports the point estimates and 95% confidence intervals from regression of the estimated hospital group effect on hospital-level covariates. Hospital group effects include the fixed effect with interactions as $\widehat{\psi_k} + \frac{1}{L}\sum_l \widehat{\kappa_{lk}}$ from equation (1.6), i.e., weighting each interaction with each surgeon group equally. Higher hospital group effects are positively correlated with larger hospitals in terms of number of beds, having a medical school affilation, being a non-profit hospital, and the number of registered nurse employed. There is statistically different from zero relationship with other available hospital covariates. The R^2 of the regression including all hospital covariates amounts to about 0.07, but reduces to less than 0.01 when including all hospitals covariates available for at least 1,000 hospitals. Results are similar when using the estimated group effects from equation (1.11). Hospital ownership is obtained from the CMS Hospital General Information dataset for 2017. ZIP code level median household income and population comes from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. All other hospital-level covariates come from the CMS provider of service dataset for 2017. Standard errors displayed are robust standard errors.

Table 1.16: Relationship between predicted survival net of provider fixed effects and provider rankings

	(1)	(2)
Predicted 30-day mortality:		
Surgeon's rank	0.000121	
	(0.000078)	
Hospital's rank		-0.000053
		(0.000075)
Observations	130,844	130,844

Notes: This table reports the relationship between predicted 30-day survival net of provider fixed effects and the rank of providers from the model in equation (1.6). There is no statistically significant relationship between predicted survival net of provider fixed effects and provider rankings. The predicted 30-day survival net of provider fixed effects is calculated as $\hat{p}_{it} = \sum_p \hat{\beta}_p X_{it,p} + \hat{\gamma}_t$ estimated from equation (1.6). Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. The rank of providers is calculated based on the predicted risk-adjusted survival for the provider's group when all hospitals or surgeons groups are equally likely. Standard errors displayed are robust standard errors. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.17: Limited evidence of "triaging" within surgeons and within hospitals using patient's predicted survival net of provider fixed effects

Predicted 30-day mortality:	(1)	(2)
Surgeon's rank	0.000297	
	(0.000113)	
Hospital's rank		0.000141
		(0.000186)
Observations	130,844	130,844
Hospital FE	X	
Surgeon FE		X

Notes: This table investigates the existence of "triaging" of patients across surgeon groups within hospitals in column (1) and "triaging" of patients across hospital groups within surgeons in column (2). There is no evidence of systematic adverse selection into higher-survival providers within surgeons and hospitals. If anything, the positive relationship between predicted survival and surgeon rankings within hospitals suggests advantageous selection into surgeons within hospitals, which is more consistent with surgeons bringing in their own patients rather than the hospital assigning surgeons to patients. The relationship is indistinguishable from zero within surgeons, suggesting that surgeons do not systematically "triage" their patients into higher-survival hospitals. This table reports the coefficients $\hat{\delta}$ from regressions $\hat{p}_{ijht} = \delta_1 \text{rank}_{l(j)} + \lambda_h + \epsilon_{ijht}$ for hospitals in column (1) and $\hat{p}_{ijht} = \delta_2 \text{rank}_{k(h)} + \lambda_j + \epsilon_{ijht}$ for surgeons in column (2). $\hat{p}_{ijht}$ is the predicted 30-day survival net of provider fixed effects as $\hat{p}_{ijht} = \sum_p \hat{\beta}_p X_{it,p} + \hat{\gamma}_t$ from equation (1.6). λ_h and λ_j are individual hospital and surgeon fixed effects respectively. The rank of providers is calculated based on the predicted risk-adjusted survival for the provider's group when all hospitals or surgeons groups are equally likely. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Standard errors displayed are clustered as the hospital level in column (1) and the surgeon level in column (2). Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.18: Variance decomposition for 30-day survival

	Percentage of variance (%)	
	Selection on observables	Control function
Hospitals $Var(\psi_{k(h)})$	0.24	0.18
Surgeons $Var(\alpha_{l(j)})$	1.81	2.23
Sorting $2 \times cov(\alpha_{l(j)}, \psi_{k(h)})$	0.68	0.66
Patients covariates $Var(\beta X_{it})$	1.85	2.39
Year $Var(\lambda_t)$	0.20	0.20
Covariance FEs-patients covariates $2 \times cov(\alpha_{l(j)} + \psi_{k(h)} + \lambda_t, \beta X_{it})$	-0.33	-0.70
Covariance surgeon and hospital-year $2 \times cov(\alpha_{l(j)} + \psi_{k(h)}, \lambda_t)$	-0.01	-0.01
Residuals $Var(\epsilon_{ijht})$	95.56	95.06
N patients	111,059	111,059
N surgeons	2,911	2,911
N hospitals	1,167	1,167

Notes: This table shows the total variance decomposition of patients 30-day survival. Fixed effects are estimated following equation (1.12). The contribution of surgeons is large, larger than the contribution of hospitals, and comparable to the contribution of included patient observables. The covariance between estimated surgeon and hospital group fixed effects is positive, revealing positive assortative matching of surgeons across hospitals. The fraction of the variance explained remains small, at about 5%, which is consistent with the literature (Hull, 2018). Elements in each column sum to 100%. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included.

Table 1.19: The relationship between patient outcomes and distance to the chosen hospital is similar when including patient observables

30-day survival	(1)	(2)	(3)
Log distance (km)	0.00150	0.00159	0.00172
	(0.00070)	(0.00070)	(0.00079)
Observations	103,152	103,152	103,152
R-squared	0.00621	0.01560	0.02254
Patient's HRR FE	Yes	Yes	Yes
Patients' observables		Health-income-age	All

Notes: This table illustrates the stability of the relationship between 30-day survival and the logarithm of distance when including different set of patient observables in X_{it}. The estimated regression is $Y_i = \alpha_0 + \alpha_1 \ln d_{ih} + \alpha_3 X_{it} + \lambda_{HRR(i)} + \epsilon_i$ where Y_i is 30-day survival, d_{ih} is the distance between the patient's and the chosen hospital's ZIP codes, $\lambda_{HRR(i)}$ are patient HRR fixed effects, and X_{it} includes different sets of patient observables. Column (1) includes no patient covariate, column (2) includes patient age bins, Charlson score, and ZIP code log income per capita, and column (3) includes all available patient observables depicted in Table 1.2. The stability of the logarithm of distance parameter across specifications lends support for the exclusion restriction assumption. Hospital ZIP codes come from the 2017 National Plan and Provider Enumeration System (NPPES) data, and beneficiary ZIP codes from the Medicare Beneficiary Research Identifiable Files. Distances are calculated using ZCTA-to-ZCTA distances for distances below 100 miles, using HSA-to-HSA distances when above 100 miles and when patient and provider HSAs differ, and capped at 100 miles when patients and providers are in the same HSA but with ZCTAs distant over 100 miles. Patients' residential ZIP codes are mapped to income per capita and total population using the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. The Charlson score and comorbidities are obtained using all diagnoses appearing in inpatient, outpatient, and professional fee claims up to the twelve months prior to the surgery. The definition of hospital referral regions (HRRs) follows the definition of the Dartmouth Atlas Project. Years 2011 to 2017 are included. Standard errors in parenthesis are clustered at the patient's HRR level.

Table 1.20: Robustness of the imperfect substitutability result to alternative number of groups

	(1) Baseline	(2) K = 10; L = 5	(3) K = 5; L = 10	(4) K = L = 10
Slope surgeon rank 1 (worst)	2.31	1.47	1.26	1.31
	(0.08)	(0.05)	(0.31)	(0.12)
Slope surgeon rank 2	1.64	0.98	2.14	1.29
	(0.01)	(0.01)	(0.04)	(0.04)
Slope surgeon rank 3	1.33	0.73	1.35	0.75
	(0.01)	(0.01)	(0.01)	(0.02)
Slope surgeon rank 4	0.81	0.44	1.36	0.72
	(0.00)	(0.00)	(0.02)	(0.01)
Slope surgeon rank 5	0.20	0.13	1.14	0.44
	(0.00)	(0.00)	(0.01)	(0.01)
Slope surgeon rank 6			0.80	0.29
			(0.01)	(0.00)
Slope surgeon rank 7			0.64	0.44
			(0.01)	(0.01)
Slope surgeon rank 8			1.10	0.31
			(0.01)	(0.00)
Slope surgeon rank 9			0.36	0.20
			(0.00)	(0.00)
Slope surgeon rank 10 (best)			-0.05	-0.03
			(0.00)	(0.00)
p-value: equality of slopes	< 0.01	< 0.01	< 0.01	< 0.01
p-value: slope rank $5 \geq 1$	< 0.01	< 0.01		
p-value: slope rank $4 \geq 2$	< 0.01	< 0.01		
p-value: slope rank $10 \geq 1$			< 0.01	< 0.01
p-value: slope rank $8 \geq 3$			< 0.01	< 0.01
Observations	130,844	130,844	130,844	130,844
R-squared	0.99	0.96	0.98	0.95
Physician type FEs:	X	X	X	X

Notes: This table reports the estimated slope coefficient per surgeon group for alternative specifications. The slopes $\hat{\beta}^L$ are obtained from the regression $\hat{y}_{ijht} = \sum_{L=1}^{5} 1\{j \in L\}\, \beta^L \text{rank}_{k(h)} + \lambda_L + \epsilon_{ijht}$ where $\hat{y}_{ijht}$ is the predicted 30-day risk-adjusted survival from models delineated in equation (1.6), L is the rank of the surgeon group, $k(h)$ is the group of hospital h, $\text{rank}_{k(h)}$ is the rank of hospital group $k(h)$ in terms of predicted 30-day risk-adjusted survival, and λ_L are surgeon group fixed effects. The predicted survival is expressed in percentage points of survival. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 1.3. Professional fees come from the Medicare 20% carrier Research Identifiable Files, and hospital stays from the Medicare MedPAR Research Identifiable Files. Years 2011 to 2017 are included. Standard errors displayed are robust standard errors.

CHAPTER 2

MARKET SIZE AND TRADE IN MEDICAL SERVICES

Abstract: We document substantial interregional trade in medical services and investigate whether regional economies of scale explain it. In Medicare data, one-fifth of production involves a doctor treating a patient from another region. Larger regions produce greater quantity, quality, and variety of medical services, which they "export" to patients from smaller regions. These patterns reflect scale economies: greater demand enables larger regions to improve quality, so they attract patients from elsewhere. Contrary to concerns that production is too concentrated, we estimate that larger regions have higher marginal returns. We study counterfactual policies that would lower travel costs rather than relocating production.

2.1 Introduction

Rural Americans have worse health outcomes (Deryugina and Molitor, 2021; Finkelstein, Gentzkow, and Williams, 2021), but America's doctors are disproportionately located in big cities (Rosenblatt and Hart, 2000).[1] This contrast might suggest a spatial mismatch between consumers and producers of medical services, and arguments about whether physicians are geographically "maldistributed" go back decades (Newhouse et al., 1982*b*; Skinner et al., 2019). To evaluate this concern, we must consider two economic mechanisms: economies of scale and patients' travel costs. We find that both are key to understanding spatial patterns of healthcare within the United States.

When medical services exhibit increasing returns to scale, there are benefits to geographically concentrating production. Indeed, medicine has long been suggested as an industry in which the division of labor is limited by the extent of the market (Arrow, 1963; Baumgardner, 1988*b*). But if healthcare markets are geographically segmented, the only way to serve

1. This chapter is co-authored with Jonathan Dingel, Joshua Gottlieb, and Maya Lozinski.

patients in smaller regions is to disperse production across space, foregoing the benefits of scale.[2] For time-sensitive emergency care, this assumption is plausible. But the vast majority of medical spending is not for such emergencies. For example, if patients with cancer can travel across regions in search of the ideal oncologist—one specialized in their particular type of cancer, one with a better reputation, or simply a better personal match—the economic geography of medical care may resemble other tradable industries. Society would face a proximity-concentration tradeoff: patients who import medical services produced elsewhere incur trade costs but benefit from higher quality generated by scale economies.

We quantify the roles of local increasing returns and trade costs in medical services. Using millions of Medicare claims, we find that "imported" medical procedures—defined as a patient's consumption of a service produced by a medical provider in a different region—constitute about one-fifth of US healthcare consumption. Imports are a larger share of consumption for patients in smaller markets. "Exported" medical services are disproportionately produced in large markets. Larger regions specialize in producing less common procedures, and these procedures are traded more. These patterns are attributable to local increasing returns to scale: larger regions produce higher-quality services because they serve more patients. We estimate a model and use it to quantify how production or travel subsidies would affect patients' access to care and the quality produced in each region. Spatially neutral policies affect regions differently depending on their size and trade patterns.

Section 2.2 develops a model of trade in medical services to guide our analysis. We adapt standard models of agglomeration and trade to a setting in which the government sets prices, so endogenous quality and travel patterns clear markets. If there are local increasing

2. Many economists assume trade costs for medical services are prohibitively high. Hsieh and Rossi-Hansberg (2021): "Producing many cups of coffee, retail services, or health services in the same location is of no value, since it is impractical to bring them to their final consumers." Jensen and Kletzer (2005): "Outside of education and healthcare occupations, the typical 'white-collar' occupation involves a potentially tradable activity." Bartik and Erickcek (2007): "An industry can bring in new dollars by selling its goods or services to persons or businesses from outside the local economy ('export-base production')...For health care institutions, demand for services tends to be more local."

returns, larger markets produce higher-quality care and export it. When economies of scale are sufficiently strong relative to market size, the model predicts that larger markets will be net exporters of medical services. Market size matters more at smaller scales, so less common medical procedures respond more to differences in market size.

Section 2.3 describes our Medicare claims data. Medicare is the federal government's insurance program for the elderly and disabled and the largest insurer in the United States. Medical service providers submit claims that report the treatment location, where the patient lives, and distinguish among thousands of distinct medical procedures.

Section 2.4 begins our empirical investigation by examining how production and consumption vary with market size. Production is geographically concentrated in larger markets, while consumption is much less so. This contrast implies that larger markets are net exporters of medical services to smaller markets. To test whether this pattern reflects a home-market effect—that is, larger demand causes larger regions to export medical services—we estimate a gravity model of bilateral gross trade flows (Costinot et al., 2019). Controlling for the geographic distribution of demand and travel distances, regions with larger residential populations export more medical care. Local increasing returns to scale are so strong that greater demand induces a larger increase in exports than imports, making larger markets net exporters of medical care. We show that these scale effects cannot be attributed to larger markets having lower input costs or medical production raising population size.

Section 2.5 shows that trade and market size play a larger role in less common procedures. The imported share of consumption is 22% for above-median-frequency procedures and 35% for those below the median. Doctors performing rare procedures export their services more often and across a broader geographic scope, sometimes serving patients who reside thousands of kilometers away. For example, half of the patients having left ventricular assist devices (LVADs) inserted to restore their heart function come from outside the surgeon's region, while only 15% of screening colonoscopies are imported. Consistent with the model, the home-market effect is substantially stronger for less common procedures: a larger residential population drives a greater increase in net exports for rarer services.

Section 2.6 shows that larger markets produce higher-quality services thanks to economies of scale. We recover revealed-preference estimates of regional service quality by estimating patients' willingness to travel to each exporting region for medical services.[3] These estimates are positively related to external quality measures, such as hospital rankings published by *U.S. News and World Report*. Inferred quality rises considerably with the regional volume of production. We estimate the scale elasticity of production to be about 0.6: a region producing 10% more because of greater demand produces about 6% higher quality.

A variety of mechanisms could generate these local increasing returns to scale: finer specialization among physicians, sharing of lumpy capital equipment, knowledge diffusion, learning by doing, and greater availability of complementary inputs (Marshall, 1890). While we cannot test all these hypotheses, we find that physicians in larger markets are more specialized and more experienced in the procedures they perform. Trade enables patients from across regions to share in these benefits of scale: imports are more likely to be provided by a specialist—and the appropriate specialist—than locally produced services. Specialization and learning by doing likely contribute to the local increasing returns that produce higher-quality medical care in larger markets.

We use our estimates of scale economies and trade costs to quantitatively explore the proximity-concentration tradeoff. Section 2.7 shows that policies affect regions differently depending on their size and trade patterns. A nationwide increase in reimbursements raises local output quality more in smaller regions, but these regions experience smaller increases in patients' market access because fewer of their patients consume local services. We then examine the implications of increasing access to care in one region by either increasing reimbursements or reducing travel costs. Increasing reimbursements has a higher return in

3. Regional quality estimates and other results may be downloaded at `http://jdingel.com`.

more populous regions: the nationwide improvement in patient market access is about 15% higher per dollar of spending when raising reimbursements in the largest regions instead of the smallest regions. Increasing reimbursements in one region reduces output quality in neighboring regions, while improving patients' market access to the extent they import from the treated region. Reducing travel costs for one region increases its import demand, which improves both output quality and market access in neighboring regions. The rich pattern of consequences when subsidizing patients in low-output regions highlights the importance of trade and agglomeration for the incidence of these policies on patients and producers.

The higher-quality care available in larger markets may not benefit all patients equally. Patients of lower socioeconomic status are less likely to travel for better medical care. Gravity regressions show that patients from the lowest neighborhood-income decile exhibit a distance elasticity of -2.1, while those in the highest decile have a distance elasticity of -1.7. This finding is not driven by differences in the composition of care needed: these patients are more sensitive to distance even when we examine travel patterns within specific billing codes. Thus, the gains generated by local increasing returns do not benefit all patients equally.

This paper builds on research in urban, trade, and health economics. Urban economists have documented skill-biased agglomeration in production as knowledge workers have become more numerous and concentrated in skilled cities (Berry and Glaeser, 2005; Moretti, 2011; Diamond, 2016; Davis and Dingel, 2020; Eckert, Ganapati, and Walsh, 2020). Connecting this to the production and trade of services has been more difficult. Most studies of the geography of services analyze restaurants and retailers (Davis et al., 2019; Agarwal, Jensen, and Monte, 2020; Allen et al., 2021; Miyauchi, Nakajima, and Redding, 2021; Burstein, Lein, and Vogel, 2022). We show that—even in a service-based economy—the sizes of both local and potential export markets influence production and quality. This suggests that healthcare can serve as an export base for large markets (Bartik and Erickcek, 2007).

The trade literature has examined market-size effects in manufacturing but investigated services much less. Davis and Weinstein (2003), Hanson and Xiang (2004), and Bartelme et al. (2019) link manufactures' market size to export patterns, in line with the home-market effect of Krugman (1980) and Helpman and Krugman (1985). Dingel (2017) shows that market-size effects drive quality specialization across US cities. Market-size effects for pharmaceuticals have been estimated using demographic variation over time (Acemoglu and Linn, 2004) and across countries (Costinot et al., 2019). Services are much less studied, in part because of the paucity of reliable trade data (Lipsey, 2009; Muñoz, 2022). We advance this literature using the detailed procedure and location information in medical claims data.

The importance of medical care for health, life expectancy, and welfare generates substantial public-policy interest. Rural locations have worse health outcomes but fewer doctors per capita. An important series of papers by Newhouse et al. (1982*b*,*d*,*e*), Newhouse (1990), and Rosenthal, Zaslavsky, and Newhouse (2005*b*) considered this issue and argued against targeting a uniform geographic distribution of physicians. Building on these studies, we measure interregional trade in medical services, estimate the impact of geography on patient access, and connect this trade to economies of scale. Importantly, we use modern trade theory to guide our modeling, estimation strategy, and counterfactual policy analysis.

2.2 Theoretical framework

This section develops a model of trade in medical services tailored to our empirical analysis of US healthcare. Patients select quality-differentiated services and face trade costs. Regional increasing returns cause the quality-adjusted cost of producing a service to decline with scale. The distinction between lower costs and higher quality is important in our empirical context. The US government plays a unique role in healthcare, purchasing a large share of all output and imposing substantial regulations. We focus on Medicare, the large federal program that purchases healthcare for the elderly and disabled at regulated prices. In this context, prices do not play their traditional role in clearing markets. Instead, quality of care and patients' distance from care bring this market towards equilibrium.

For brevity, we present a competitive model, but the consequences of regional increasing returns for trade flows in a fixed-price environment do not hinge on this assumption. Appendix 2.10.1 shows that a monopolistic-competition model with one medical provider in each region delivers the same predictions. As in flexible-price models, many market structures can give rise to a home-market effect (Costinot et al., 2019).

Beyond healthcare, this model speaks to agglomeration effects in other markets subject to price controls. We show that such circumstances can be captured by a modest modification to conventional trade models. Our model continues to deliver a gravity equation for trade flows and to predict home-market effects. This framework delivers testable predictions about spatial variation in services' quality and trade patterns when prices are fixed.

2.2.1 Demand

We use a logit model of individuals choosing providers for a given service. Providers and patients are in regions indexed by i or j, with $\mathcal{I}$ denoting the set of regions. Let N_j denote the number of patients residing in region j who make a choice.[4] All providers in a region are identical. Utility has a provider-region-specific component, a region-pair component, and an idiosyncratic component: patient k in region j choosing a provider in region i obtains utility

$$U_{ik} = \ln \delta_i + \ln \rho_{ij(k)} + \epsilon_{ik}.$$

The provider-region-specific component δ_i would usually include a product's characteristics and price. Since Medicare pays reimbursement rates that it sets administratively,[5] the δ_i relevant for the patient is the quality of the providers in region i. The region-pair component

4. Appendix 2.10.2 extends the model to have multiple patient types.

5. Patients pay a share of these reimbursements through copayments and deductibles. But note that these cost-sharing rules are constant nationally, and most Medicare patients have a supplemental insurance (Medigap or Medicaid) which covers most or all of this cost-sharing.

ρ_{ij} represents bilateral inverse trade costs (proximity). The idiosyncratic component ϵ_{ik} is independently and identically drawn from a standard Gumbel distribution, so the probability that patient k selects a provider in region i is

$$\Pr\left(U_{ik} > U_{i'k}\ \forall i' \neq i\right) = \frac{\exp\left(\ln \delta_i + \ln \rho_{ij(k)}\right)}{\sum_{i' \in 0 \cup \mathcal{I}} \exp\left(\ln \delta_{i'} + \ln \rho_{i'j(k)}\right)}.$$

There is an outside option denoted by $i = 0$, which represents individuals choosing to forgo care, and we normalize its common component to zero, $\ln \delta_0 = \ln \rho_{0j(k)} = 0\ \forall k$.[6]

This choice probability implies a gravity equation for the quantity of trade between any two regions when we aggregate patients' decisions. Let Q_{ij} denote the quantity of procedures supplied by providers in i to patients residing in j, and let Q_{0j} denote the number of patients in j selecting the outside option. Because each patient selects at most one provider, $N_j = \sum_{i \in \mathcal{I} \cup \{0\}} Q_{ij}$. The demand by patients in j for procedures performed in i is

$$Q_{ij} = \delta_i \frac{N_j}{\Phi_j} \rho_{ij}, \tag{2.1}$$

where $\Phi_j \equiv \sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}$ is the expected value of the choice set for patients in region j. We call this Φ_j "patient market access." Equation (2.1) is a gravity equation with an origin i component, a destination j component, and an ij pair component. Total demand

6. This formulation of demand is familiar from the hospital competition literature, which has studied competition among hospitals on price and quality. The literature tends to assume competition occurs within a specified geographic radius (*e.g.,* Kessler and McClellan, 2000; Cooper et al., 2018) or within a metropolitan area or similar geographic unit (*e.g.,* Ho, 2009; Gowrisankaran, Nevo, and Town, 2015; Clemens and Gottlieb, 2017; Lewis and Pflum, 2017; Ho and Lee, 2019; Dafny, Ho, and Lee, 2019; Garthwaite, Ody, and Starc, 2022). Data in this literature are often limited to certain states (*e.g.,* Town and Vistnes, 2001; Gaynor and Vogt, 2003; Capps, Dranove, and Satterthwaite, 2003; Lewis and Pflum, 2015; Ericson and Starc, 2015; Ho and Lee, 2017). Patients who are treated outside their home region may be dropped from the data or treated as choosing the outside option (as in Gaynor and Vogt, 2003). These definitions may be appropriate for modeling competition within specified markets (though they have been questioned by Gaynor, Kleiner, and Vogt, 2013; Dranove and Ody, 2016) and are natural if one assumes healthcare demand is local—as has been standard (see footnote 2). We assume all regions are in each patient's choice set, so there are no "control" markets and modeling strategic interactions would be very computationally costly.

for procedures produced in i is

$$Q_i = \delta_i \sum_j \frac{N_j}{\Phi_j} \rho_{ij}. \tag{2.2}$$

2.2.2 Production

We assume competitive production of services with free entry and local increasing returns that are external to the firm. That is, each price-taking provider chooses its output quality and quantity given total regional production, an exogenous factor price, and an exogenous productivity shifter. A provider in region i that employs L units of the composite input to produce service of quality δ produces the following output quantity:

$$A_i \frac{H(Q_i)}{K(\delta)} L.$$

Improving quality is costly so $K(\delta)$ is increasing. Regional increasing returns to scale are a weakly increasing, concave function $H(Q_i)$ of total regional production, Q_i, which competitive firms take as given (Chipman, 1970). The regional productivity shifter A_i captures any other influences, such as historical investments. Provider size L is indeterminate (and unimportant) given the linear production function, external economies of scale, and price-taking behavior. The composite input is supplied to region i at factor price w_i.[7] Thus, the unit cost of producing quality δ in region i is

$$C(Q_i, \delta_i; w_i, A_i) \equiv \frac{w_i K(\delta_i)}{A_i H(Q_i)}.$$

In our institutional setting, output prices are not an equilibrium object determined solely by the intersection of supply and demand. Instead Medicare sets "reimbursement rates"

7. If the regional factor supply were upward-sloping rather than perfectly elastic, we would estimate increasing returns net of the cost of hiring additional inputs. That is, if the factor supply elasticity were β, our estimate of the scale elasticity α from equation (2.4) below would instead be an estimate of the effective scale elasticity $\tilde{\alpha} \equiv \alpha - \frac{\beta}{1+\beta}$.

largely independent of quality, quantity, or region,[8] which we denote $\overline{R}$. Each provider that produces output of the highest quality produced in region i earns revenue $\overline{R}$ per unit.

Provider optimization and free entry make the unit cost equal to the reimbursement rate in each region. Given the factor price w_i and productivity shifter A_i, the free-entry condition

$$C(Q_i, \delta_i; w_i, A_i) = \overline{R} \tag{2.3}$$

defines a regional isocost curve: the set of quantity-quality combinations for which the average cost of production equals the reimbursement rate. This isocost curve is the set of potential equilibrium production outcomes in region i. Regional increasing returns make the isocost curve upward-sloping in (Q, δ) space. With free entry and fixed prices, the benefits of scale are realized as higher-quality services in higher-output regions.

While our assumptions thus far suffice for qualitative results, we later specify functional forms for additional predictions and empirical quantification; specifically, $K(\delta_i) = \delta_i$ and $H(Q_i) = Q_i^{\alpha}$, with a scale elasticity of $\alpha \in (0, 1)$. In this case, the free-entry condition (2.3) is

$$\overline{R} = \frac{w_i \delta_i}{A_i Q_i^{\alpha}}. \tag{2.4}$$

8. While Medicare does have some quality incentive programs, the money at stake is a small share of Medicare's overall spending (Gupta, 2021). Medicare has some spatial variation in physician reimbursements, but it is not very large and has diminished over time (Clemens and Gottlieb, 2014).

2.2.3 Equilibrium

Equilibrium equates supply and demand in each region, $Q_i = \sum_j Q_{ij}$. Given exogenous parameters $\overline{R}$, $\{w_i, A_i, N_i\}_{i\in\mathcal{I}}$, and $\{\rho_{ij}\}_{(i,j)\in(\mathcal{I},\mathcal{I})}$, an equilibrium is a set of quantities and qualities $\{Q_i, \delta_i\}_{i\in\mathcal{I}}$ that simultaneously satisfy equations (2.2) and (2.3).

2.2.4 Scale effects in autarky

We first consider equilibrium in autarky: patients can choose whether to receive care, but they cannot travel between regions ($\rho_{ij} = 0$ for $i \notin \{0, j\}$). In this case, all demand is local and equation (2.2) simplifies to

$$Q_{jj} = \frac{\delta_j \rho_{jj}}{1 + \delta_j \rho_{jj}} N_j. \tag{2.5}$$

The autarkic equilibrium is at the intersection of the demand curve given by equation (2.5) and the free-entry isocost curve given by equation (2.3).[9] An increase in population size, $\Delta N_j > 0$, affects equilibrium outcomes by shifting the demand curve.

Figure 2.1 illustrates how greater demand affects quality in autarky. Panel 2.1(a) shows the role of increasing returns to scale. The vertical axis shows quality δ_i and the horizontal axis shows quantity Q_i (on logarithmic scales). Higher quality attracts more patients, so demand is upward-sloping.[10] We draw two cases of the free-entry isocost curve defined by equation (2.4): the horizontal line depicts constant returns ($\alpha = 0$) and the upward-sloping line depicts increasing returns ($\alpha > 0$). With constant returns, a rightward shift in demand ($\Delta N_j > 0$) causes a proportional increase in quantity produced and no change in output quality. With increasing returns, the demand shift elicits higher quality because producers move up the isocost curve and thus implies a more-than-proportional increase in quantity

9. For the equilibrium to be Marshallian stable, the demand curve must be steeper than the isocost curve at the intersection. There is a stable equilibrium because equation (2.5) means $Q_{jj} \to N_j$ as $\delta_j \to \infty$.

10. For visual clarity, we draw a log-linear demand curve. The logit demand function (2.5) is in fact log-convex, which is consistent with all the comparative statics illustrated in Figure 2.1.

produced because the share of patients receiving care rises.

Panel 2.1(b) shows that an increase in demand raises quality more as the demand curve is increasingly elastic. The panel depicts two demand curves: the one on the left is more elastic, as we would expect for a less-common procedure.[11] Shifting each demand curve to the right raises the equilibrium quality of each procedure because of increasing returns to scale. This market-size effect is larger for the less common procedure with more elastic demand because the demand shift is amplified by a larger increase in quantity demanded.[12]

2.2.5 Market-size effects on trade flows

We now consider trade. With multiple regions and finite trade costs ($\rho_{ij} > 0$), some patients will engage in trade—*i.e.*, select a provider located in another region. This trade stems from two sources. First, in the logit demand system with finite trade costs, patients have idiosyncratic preferences that yield a strictly positive probability of choosing every region. Second, when quality varies, regions producing higher-quality services attract more patients.

Fixing the qualities produced in other regions, an increase in one region's demand affects its trade flows through three mechanisms. First, greater demand for services directly raises a region's demand for imports through the N_j term in equation (2.1). A larger population translates proportionally to a greater demand for imports. Second, with increasing returns, an increase in N_i elicits an increase in quality δ_i, which raises region i's *gross* exports to each region. Costinot et al. (2019) call this the "weak home-market effect." Third, if increasing returns are sufficiently strong, the increase in quality δ_i improves region i's patient market access Φ_i so much that $\ln \delta_i$ rises more than $\ln \left(\frac{N_i}{\Phi_i}\right)$ does. That is, the increase in region i's gross exports exceeds any increase in its gross imports. This is the "strong" home-market

11. The demand function (2.5) is log-convex, so demand is indeed more elastic at lower quality. This is a fixed-price counterpart of Marshall's second law that demand is more elastic at higher prices.

12. Alternatively, one could obtain this prediction by assuming that demand is log-linear and the isocost curve is log-concave. A rightward shift in demand would cause a larger (log) difference in quality for the low-volume procedure on the steeper part of the isocost curve.

effect: an increase in local demand raises a region's *net* exports.

Figures 2.1(c) and 2.1(d) introduce trade and illustrate the distinction between weak and strong home-market effects.[13] Panel 2.1(c) depicts the quality and quantity produced in one region under two scale elasticities. Comparing points B and C, we see that a given increase in demand elicits a larger quality improvement when increasing returns are stronger. Panel 2.1(d) depicts equilibrium exports and imports as a function of the region's demand shifter N_j. The import curves are upward-sloping because an increase in local demand raises demand for imports. The export curves are upward-sloping because of increasing returns: an increase in local demand causes an increase in quality, which causes an increase in gross exports. This is the weak home-market effect. When the scale elasticity α is larger—the free-entry isocost curve in Figure 2.1(c) is steeper—greater demand elicits a larger increase in output quality, which steepens the export curve and flattens the import curve in Figure 2.1(d). When the export curve is steeper than the import curve, there is a strong home-market effect: the increase in demand raises exports more than imports.

We predict larger effects of market size for less common procedures. When two procedures have the same production function and trade costs, demand is more elastic at the rare procedure's equilibrium quantity. As Figure 2.1(b) shows, an increase in demand raises quality more when the demand curve is more elastic, leading to a stronger home-market effect for the rarer procedure.

If rare procedures also have greater economies of scale (higher α)—for example, because they require specialized equipment—that would amplify this contrast. This result motivates a difference-in-differences research design: we compare the market-size effects of common and rare procedures.

These results continue to hold when an increase in demand in one region affects equilib-

13. These diagrams are fixed-price analogues of Figures II and III in Costinot et al. (2019). See their discussion of the assumption that one region is large enough to affect its own quality but too small to affect the quality produced in other regions. This assumption is only made for this figure.

rium outcomes in all other regions. To demonstrate this, we consider the isoelastic special case with scale elasticity $\alpha \in (0,1)$ and examine the home-market effect in the neighborhood of a symmetric equilibrium. Suppose all regions are the same size, $N_i = \bar{N}\ \forall i$, and trade costs are symmetric: $\rho_{ii} = 1$ and $\rho_{ij} = \rho \in (0,1)\ \forall i \notin \{0, j\}$. There is a symmetric equilibrium, which has quality $\bar{\delta}$ and patient market access $\bar{\Phi}$ in each region. As detailed in Appendix 2.10.3, we totally differentiate the system of equations in terms of $\{\mathrm{d}\delta_i, \mathrm{d}N_i\}_{i=1}^{\mathcal{I}}$ and evaluate this system with $\mathrm{d}N_1 > 0$ and $\mathrm{d}N_j = 0\ \forall j \neq 1$ at the symmetric equilibrium.

With increasing returns of any magnitude, there is a weak home-market effect; with sufficiently strong increasing returns, there is a strong home-market effect. When $\alpha > 0$, an increase in the population size of region 1 elicits an increase in the quality of service produced in region 1 relative to the other regions:

$$\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1} = \left[\frac{1-\alpha}{\alpha}\frac{(\bar{\Phi}-1)}{(1-\rho)\bar{\delta}} + \frac{(1-\rho)\bar{\delta}}{\bar{\Phi}}\right]^{-1} \mathrm{d}\ln N_1 > 0.$$

This higher quality causes region 1 to export more to every other region: $\frac{\mathrm{d}\ln Q_{1j}}{\mathrm{d}\ln N_1} > 0$. The effect on the region's net exports is

$$\mathrm{d}\ln Q_{1,j\neq 1} - \mathrm{d}\ln Q_{j\neq 1,1} = \left[\frac{1 - \frac{1-\alpha}{\alpha}\frac{1+(\mathcal{I}-1)\rho}{1-\rho}}{\frac{1-\alpha}{\alpha}\frac{(1+(\mathcal{I}-1)\rho)}{(1-\rho)} + \frac{(1-\rho)\bar{\delta}}{1+(1+(\mathcal{I}-1)\rho)\bar{\delta}}}\right] \mathrm{d}\ln N_1. \tag{2.6}$$

Net exports increase if and only if

$$\frac{\alpha}{1-\alpha} > \frac{1+(\mathcal{I}-1)\rho}{1-\rho}.$$

When this inequality holds, the larger population size of region 1 makes it a *net* exporter of the medical procedure; *i.e.*, the procedure exhibits a strong home-market effect around the symmetric equilibrium. This occurs if increasing returns are sufficiently strong (α is large enough) and trade costs are sufficiently large (ρ is small enough). Otherwise, there is a weak home-market effect but not a strong one. Given a strong home-market effect, the effect in equation (2.6) is diminishing in the number of potential patients $\bar{N}$, so we predict a stronger home-market effect for less common procedures.

While the existence of increasing returns seems likely—at least for some types of medical care—there is no guarantee they are sufficiently large to generate a strong home-market effect. When larger markets are net exporters, they produce care that smaller regions need. This trade can also support the larger markets' economies: rather than exporting manufactured goods, as in decades past, larger cities can reinvent themselves (Glaeser, 2005) and export medical services. Absent a strong effect, healthcare would be a net import, not an economic base, for larger regions.

2.3 Data description

Our primary dataset is 2017 claims data from Medicare, the US federal government's insurance program for the elderly and disabled. Medicare is the largest health insurer in the United States. It does not directly employ physicians or run its own hospitals. Instead, it pays bills submitted by independent physicians, physician groups, hospitals, and other medical service providers. These bills—called "claims" in industry terminology—report the specific services provided using 5-digit codes from the Healthcare Common Procedure Coding System (HCPCS). There are over 12,000 distinct HCPCS codes, which identify individual procedures at a granular level.[14] Federal regulation determines the payment for each claim, rather than physicians' or hospitals' pricing decisions. In alternative analyses we use groupings of patient *diagnoses* to account for potential substitution between treatments.[15]

The claims data report the geographic location of both the physician providing the care and the patient receiving it, allowing us to construct a trade matrix for medical services. We study all medical care provided by physicians outside an emergency room, whether in an office or hospital facility.[16] Because Medicare rarely reimbursed telehealth in 2017, this trade involves traveling to receive a service delivered in-person.[17] We aggregate the ZIP-code-level information up to 306 hospital referral regions (HRRs), which are geographic units defined by the Dartmouth Atlas Project to represent regional health care markets for tertiary

14. For instance, there are distinct codes for providing flu vaccines based on patient age, whether the vaccine protects against three or four strains of flu, and whether administration is intramuscular or intranasal. There are distinct codes for chest X-rays based on whether the images are of ribs, the breastbone, or the full chest, both sides or one side of the body, and the number of images taken (1, 2, 3, or 4+).

15. We use the Clinical Classifications Software Refined (CCSR) diagnosis categories produced by the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project. CCSR aggregates over 70,000 ICD-10-CM diagnosis codes into "clinical categories," of which 482 have at least 20 patients each in our data. We split these categories at the median frequency to separate common from rare diagnoses.

16. Our results are robust to adding the value of hospital facility fees on top of physicians' professional fees.

17. In 2012, Medicare spent only \$5 million—less than 0.001% of its expenditures—on telehealth services (Neufeld and Doarn, 2015), lagging other insurers (Dorsey and Topol, 2016).

medical care based on 1992–93 data. We construct HRR-to-HRR trade flows by interpreting the patient's residential HRR as the importing region and the service location's HRR as the exporting region.[18] The Dartmouth Atlas Project defines HRRs by aggregating residential areas based on where patients were referred for major cardiovascular surgical procedures and for neurosurgery and requires each HRR to have at least one city where both major cardiovascular surgical procedures and neurosurgery were performed. Thus, the construction of these geographic units should tend to minimize trade between different HRRs.[19]

Physicians, hospitals, pharmacies, and other healthcare providers submit different types of claims. We use a random 20% sample of all physician claims paid by Traditional (fee-for-service) Medicare in 2017, selected randomly by patient.[20,21] One year of data from this sample contains 229 million services, representing $19 billion in spending. The Medicare claims are not perfectly representative of all US healthcare, since Medicare beneficiaries are elderly or disabled. But the geographic distribution of Medicare beneficiaries is quite similar to the overall population, and Medicare alone finances one-fifth of medical spending. So it is likely to capture the key features of overall healthcare production and consumption.

Since we only see a sample of Medicare data—and hence an even smaller share of overall medical care—we might completely miss physicians or procedures so rare that a 20% sample includes none of them in a particular location. We use two other sources to address this concern. First, we use a less-detailed but more comprehensive extract of Medicare data

18. The Medicare claims are US patients receiving care at US service facilities. These data do not report any international transactions. Throughout this paper, "imports" and "exports" refer to domestic transactions between regions of the United States.

19. We have also used alternative geographies, including core-based statistical areas (CBSAs) and metropolitan statistical areas, a subset of CBSAs that excludes the smaller micropolitan areas. Because these yield consistent findings, we do not report all such estimates.

20. We also use data from 2011 to 2016 to investigate trade patterns over time in Appendix Figure 2.13.

21. One-third of Medicare patients opt out of the traditional version of Medicare, where care is paid directly by the government, in favor of a private insurance scheme ("Medicare Advantage"). In these private schemes, the government pays the insurer a fixed amount per patient and the insurers are responsible for the patient's care. Because Medicare does not pay claim-level bills in these private insurance schemes, the availability and quality of data for the privately insured patients is lower. We exclude these patients from our analysis.

(based on all Traditional Medicare patients) to replicate some of our analyses and obtain extremely similar findings.[22] Second, we use physician registry data to study the geographic patterns of production by specialty. These data provide the ZIP code and specialty of all physicians registered to practice in the United States. Physician specialty is conceptually distinct from medical service—and there is not a one-to-many mapping of specialties to services, since many services can be provided by physicians of different specialties—but we expect many of the same economic forces to apply at the level of physician specialties.

2.4 Is there a home market effect in medical services?

This section estimates how scale economies and trade costs shape the geography of aggregate healthcare production and consumption. Section 2.4.1 documents size-related spatial variation in both production and consumption. Section 2.4.2 shows that bilateral trade declines with distance. Section 2.4.3 describes our empirical strategy, which identifies the consequences of market size using gravity equations to model bilateral trade flows of medical services. Section 2.4.4 reports the empirical estimates, which demonstrate a strong home-market effect.

2.4.1 Spatial variation in production and consumption

Figure 2.2 shows maps of healthcare production and consumption across regions. The consumption map shows the substantial geographic variation that has been well-documented by the Dartmouth Atlas and related literature on geographic variation in healthcare (Fisher et al., 2003*a*,*b*; Finkelstein, Gentzkow, and Williams, 2016). The production map shows even more pronounced variation: more production in large urban agglomerations and less in rural

22. Appendix 2.11.1 explains why we must use the 20% sample and uses the 100% data to confirm some of our measures. It also shows that the relative frequencies of services purchased by private insurance are similar to those in Medicare.

areas. There is substantial variation in production even between neighboring regions, while spatial variation in consumption is smoother.

The subsequent panels show patterns of trade, which constitutes the difference between production and consumption. Nationally, 22.4% of production is exported to a patient in another region.[23] Panel 2.2(c) shows the ratio of production to consumption; a value larger than one means an HRR is a net exporter. Net-exporting regions tend to be major urban agglomerations, plus places such as Rochester, Minn. and Hanover, N.H. that specialize in healthcare. Panel 2.2(d) shows gross exports as a share of local production for each HRR. Three-quarters of services produced in the Rochester metropolitan area, home to the top-rated Mayo Clinic, are provided to patients from other regions, who travel an average of 545 km to Rochester. As a major healthcare exporter with a population of merely 220,000, Rochester is an outlier: larger regions are responsible for a disproportionate share of medical services production.

Figure 2.3 plots the average production and consumption per capita across HRRs of different sizes. Both rise monotonically with population. Production rises about twice as steeply, with a population elasticity of 0.13 versus 0.06 for consumption. The difference between production and consumption is net trade: larger markets are net exporters and smaller markets are net importers. Gross trade flows exceed net trade flows, with imports comprising about one-third of consumption in the smallest regions. Exports per capita are approximately flat, which means total exports are increasing with local population. Imports per capita decline with an elasticity of -0.25 with respect to population.

23. This value is nearly identical whether measured across HRRs or across CBSAs. Appendix Figure 2.13 shows that the exported share rose steadily from 18.6% in 2011 to its 2017 level of 22.4%. For manufactured goods, the export share across CBSAs is about 68%.

2.4.2 Bilateral trade and bilateral distance

Despite the clear patterns in Figure 2.3, geographic variation in trade is far from entirely explained by market size. The four regions with the lowest export shares are Anchorage, Honolulu, and Yakima and Spokane, Wash., likely reflecting their isolated geographic locations. The highest export shares are in Rochester, Minn., Ridgewood, N.J. (just outside of New York City), and Hinsdale, Ill. (just west of Chicago). Other than Rochester—home to the Mayo Clinic—these exporting regions are all on the edge of major metropolitan areas and serve patients from those metros' hinterlands. To ensure our analysis captures these geographic patterns, we next examine bilateral trade flows.

Figure 2.4 depicts how trade varies with the distance between the patient and place of service. Figure 2.4(a) shows the distribution of distances patients travel for care, distinguishing between care provided in the patient's home region and other regions.[24] Within HRRs, there is a narrow distribution of distances that peaks around 10 km. When visiting providers in a different HRR, patients travel a great variety of distances. There is a local plateau between approximately 30–100 km, suggesting a fair amount of travel to nearby HRRs, perhaps indicating regional medical centers. There is another substantial peak at thousands of kilometers, demonstrating substantial long-distance travel for care.[25] Patients' willingness to travel these distance underpin our revealed-preference estimates of regional service quality.

Figure 2.4(b) shows that trade declines with distance. The blue curve depicts trade volume against distance (for pairs of HRRs with positive trade flows) after removing fixed

24. For travel within an HRR, we use the distance between the centroids of the patient's residential ZIP code and the ZIP code of the service location. We obtain the centroid coordinates from the Census Bureau's corresponding ZIP code tabulation areas (ZCTAs). For travel across HRRs, we use ZCTA-to-ZCTA distances when they are within 160 km, and (for computational ease) use HRR-to-HRR distances beyond 160 km.

25. The average patient travels 500 km to Chicago and 605 km to New York City, compared with less than 135 km to Urbana-Champaign, Ill. or Charlottesville, Va. An older literature cited in Dranove and Satterthwaite (2000) finds that patients who travel farther to hospitals tend to incur higher treatment costs.

effects for each exporter and each importer.[26] This intensive-margin relationship is roughly log-linear. The red curve shows the extensive margin: the share of pairs with positive trade as a function of distance. This is 100% for nearby pairs and under 60% for the most distant pairs. These patterns motivate the inclusion of distance covariates in our gravity-based analysis.

Patients may vary in their ability or willingness to travel, especially by socioeconomic status. We quantify it here, to the extent feasible in our data, for use in counterfactual scenarios and interpreting welfare implications. Figure 2.4(c) depicts distance elasticities estimated separately by neighborhood income decile.[27] We find a strong, nearly monotonic relationship between socioeconomic status and the distance elasticity: patients from the highest neighborhood-income decile exhibit a distance elasticity 25% smaller than those in the lowest decile.[28] This means patients from higher-income neighborhoods are more amenable to travel for medical care. Thus, the benefits of agglomeration—higher-quality rare care produced in major centers—may not be shared evenly. This is especially notable given the empirical setting: Medicare insures the near-universe of elderly and disabled Americans.

2.4.3 Gravity-based empirical strategy

We base our empirical examination of trade flows on a gravity equation that summarizes the geography of demand. We obtain this equation from the model by assuming the region-pair component in equation (2.1) satisfies $\ln \rho_{ij} = \gamma X_{ij} + \upsilon_{ij}$, where X_{ij} is a vector of observed trade-cost shifters and υ_{ij} is an orthogonal unobserved component. Taking expectations and

26. This application of the Frisch-Waugh-Lovell theorem is only feasible for positive trade volumes.

27. Our data do not contain patients' wealth or income, so we use their residential ZIP code. We split ZIP codes into deciles by median household income and estimate equation (2.12) separately by decile.

28. These estimates are consistent with the interaction that Silver and Zhang (2022) estimate between income and distance to care. These differences in distance elasticities are not driven by differences in the composition of procedures. When we estimate elasticities separately for rare and common services—or even for individual procedures (see Appendix Table 2.7)—the income gradient of distance elasticities persists.

then logs yields gross bilateral trade flows:

$$\ln \mathbb{E}\left(\overline{R}Q_{ij}\right) = \ln \delta_i + \ln\left(\frac{N_j}{\Phi_j}\right) + \gamma X_{ij}. \tag{2.7}$$

The left side of (2.7) is the value of procedures exported from region i to patients residing in j. We specify the first two right-side regressors as either observable demand shifters or fixed effects in different specifications described below. We generally parameterize observed trade-cost shifters as containing log distance and a same-region dummy, so that $\gamma X_{ij} = \gamma_1 \ln \text{distance}_{ij} + \gamma_0 \mathbf{1}(i = j)$. Alternative specifications include $(\ln \text{distance}_{ij})^2$ or replace these continuous distance covariates with indicators for distance deciles.

When using the total value of bilateral exports as the dependent variable in (2.7), we aggregate quantities across thousands of distinct medical procedures using the average national Medicare reimbursement rate for each procedure. This produces an expenditure measure independent of any spatial variation in reimbursement rates.[29] We also estimate procedure-level versions of (2.7) for selected procedures, such as LVAD insertion and screening colonoscopy. The dependent variable in this case is the procedure count and no aggregation is required. Since observed bilateral trade is zero for many pairs of regions, especially when looking at trade in individual procedures, we estimate (2.7) using Poisson pseudo-maximum-likelihood (PPML; Santos Silva and Tenreyro, 2006).

We test for a home-market effect in medical services using population as an observed demand shifter. Following Costinot et al. (2019), we differentiate the system of equations (2.2) and (2.3). around the symmetric equilibrium. This delivers the local relationship between trade and population, independent of market access Φ_j. The estimating equation is

$$\ln \mathbb{E}\left[\overline{R}Q_{ij}\right] = \lambda_X \ln \text{population}_i + \lambda_M \ln \text{population}_j + \gamma X_{ij}. \tag{2.8}$$

29. Mechanically, we multiply the quantity of each procedure by the national average price for that procedure and denote the sum across all procedures by $\overline{R}Q_{ij}$.

Relative to (2.7), equation (2.8) replaces $\ln \delta_i$ and $\ln\left(\frac{N_j}{\Phi_j}\right)$ by log population in the producing and consuming regions, respectively. A positive coefficient $\lambda_X > 0$ implies a weak home-market effect as defined in Costinot et al. (2019): *gross* exports increase with market size. If $\lambda_X > \lambda_M > 0$, the home-market effect is strong: *net* exports increase with market size.

One potential concern with estimating (2.8) directly is reverse causality. Suppose that success in exporting medical services serves as an employment base that raises current population size, as epitomized by "anchor institutions." For example, William Worrall Mayo settling in Rochester, Minn. in the 1860s, and subsequent investment in medical care and reputation, helps explain Rochester's current population (Clapesattle, 1969).

We use two instrumental variables to address this concern. First, we use historical population. Medicine was a far smaller industry in 1940, and it is implausible that it could have driven local population in the way it might today. Since population is persistent over time, population in 1940 predicts contemporary population, and we are interested in capturing any effects of historical population that operate through current population. We therefore instrument for both the exporting region's and importing region's contemporaneous log populations with the respective log populations in 1940.

Our second instrument goes farther back than 1940 and uses local geology to predict population. Rosenthal and Strange (2008) and Levy and Moscona (2020) show that shallower subterranean bedrock makes construction easier, leading to higher population density. Bedrock depth also predicts population size, so we use this as a second instrument for local demand, again for both the importing and exporting regions.[30]

30. This instrument is currently only available for CBSAs, but not for HRRs. We demonstrate that our main results are robust to defining markets based on CBSAs and to using both instruments at this level. Levy and Moscona (2020) show that the instrument has ample first-stage power for predicting population density; the same is true for our endogenous variables (population levels).

2.4.4 A strong home-market effect in medical services

Table 2.1 reports the results of estimating (2.8). The first column shows significant, positive coefficients on both patient and provider market population. The coefficient on provider-market population is two-thirds greater than that on patient-market population. This demonstrates what Costinot et al. (2019) term a *strong home-market effect.* Not only does a larger population increase gross exports, but it does so more than it increases gross imports by local patients. The distance elasticity of medical services trade between hospital referral regions is -1.7. This is substantially larger than the distance elasticity of -0.95 estimated for trade in manufactures between CBSAs (Dingel, 2017).[31] This suggests that trade in personal services incurs greater distance-related costs, relative to the degree of product differentiation across regions, than trade in manufactured goods. The most obvious difference is that patients themselves must travel to the provider.

The next two columns of Table 2.1 demonstrate that more flexible distance-covariate specifications do not alter the result. Column 2 introduces the square of log distance as an additional covariate. Column 3 replaces the parametric distance controls with dummies for deciles of distance. The result is stable across the columns: gross and net exports both increase with market size. The magnitudes are stable in columns 2 and 3, and the magnitude of gross (though not net) exports increases when excluding zeros.

The last column of Table 2.1 uses the historical population instrument to address concerns about reverse causality. We obtain similar home-market-effect estimates to our baseline results. Appendix Table 2.8 reports similar results estimated using CBSAs rather than HRRs as our geographic unit. It also shows the CBSA-based results are robust to instrumenting with either historical population or bedrock depth. Appendix Table 2.9 reports similar results when adding facility payments on top of physician fees.

31. We find a distance elasticity of medical services trade between CBSAs of -2.3. The analogous elasticity of health care and social assistance services trade between Canadian provinces is -1.42 (Anderson, Milot, and Yotov, 2014). The distance elasticity of international trade is typically near -0.9 (Disdier and Head, 2008).

The primary competing explanation for these results is other factors that reduce the cost of production w_i in larger markets. If doctors prefer to live in big cities (Lee, 2010), as college graduates generally do (Diamond, 2016), they could accept lower nominal wages and thus reduce healthcare production costs in such cities.

We investigate whether this mechanism is sufficiently large quantitatively to drive a net cost reduction in larger markets. We use data from Gottlieb et al. (2020) to measure the population elasticities of doctors' earnings and the American Community Survey (Ruggles et al., 2022) to examine other healthcare workers' earnings and real estate costs.[32] We confirm that doctors are cheaper in larger markets (Gottlieb et al., 2020), but other costs rise with population size. Appendix Figure 2.14 shows that the population elasticity of doctors' earnings is -0.01, but that for non-physicians is 0.045. To compute the population elasticity of labor costs, we use ACS data to estimate that non-physician labor's share of healthcare production is three times as much as physician labor's share. The population elasticity of labor costs is thus positive. The higher cost of real estate in larger markets reinforces these higher labor costs. This spatial variation in costs undercuts the idea that amenities make production cheaper in larger markets.

A number of related phenomena do not threaten our results. If doctors accept lower wages because they prefer the sort of work available in healthcare agglomerations, this is not a confound. Rather, it is a mechanism increasing profitability in healthcare agglomerations: greater scale lowers the cost of an input. Similarly, teaching hospitals are not a confounder. Teaching hospitals tend to be large, suggesting an agglomeration benefit of combining training with treatment at scale. Indeed, medical training exposes trainees to a large volume of patients so that they learn clinical skills by practicing them. The most salient example is Cornell University: after an abortive attempt to have medical training in both Ithaca and New York City, the Cornell Trustees quickly closed down the Ithaca location

32. Appendix 2.11.2 discusses subtleties of the income data.

and centered the medical school in New York—where the patients and doctors were more abundant—in the early 20th century (Flexner, 1910; Gotto and Moon, 2016). As this history illustrates, the potential local demand for care can drive the location of medical training.[33] If academic hospitals attract doctors, and their location is driven by market size, they are part of the agglomeration mechanism, not a confounder.

One final concern is measurement error in Medicare's records of patients' residences. To address this, Appendix 2.11.4 first demonstrates our results' robustness to excluding states with large seasonal populations. Second, we examine how far dialysis patients appear to travel. We find that residential measurement error is limited and does not drive our results.

2.5 Comparing rare and common services

Because our model predicts larger home-market effects for rarer procedures, comparing market-size effects by service frequency is a finer test of our theory. Section 2.5.1 examines how spatial variation in the production and consumption of each procedure relates to market size. Section 2.5.2 generalizes our gravity-based regression analysis to estimate home-market effects separately for rare and common procedures.

2.5.1 Spatial variation in production and consumption by frequency

We estimate the population elasticity of production and consumption per Medicare beneficiary for each procedure.[34] We find that production rises with market size more than consumption, especially for less common procedures.

33. In general education, in contrast, university placement induces economic growth (Moretti, 2004).

34. Davis and Dingel (2020) relate population elasticities to other measures of geographic concentration, such as location quotients, and estimate population elasticities of employment for various skills and sectors.

Method

We first estimate the population elasticity of production per Medicare beneficiary for each procedure. Let Q_{pi} denote the count of procedure p produced in region i and its national volume be $Q_p = \sum_i Q_{pi}$. Let M_i denote the number of Medicare beneficiaries residing in i. For each procedure p, we estimate the following relationship across regions:

$$\ln \mathbb{E}\left[\frac{Q_{pi}}{M_i}\right] = \zeta_p + \beta_p \ln \text{population}_i. \tag{2.9}$$

The estimated population elasticity of production per beneficiary, $\hat{\beta}_p$, describes how production varies with market size, and we estimate it using Poisson pseudo-maximum-likelihood.[35] If the quantity produced were simply proportional to population, β_p would be zero.

Our model suggests that scale effects play a larger role for rarer procedures. It predicts less common services will have higher population elasticities of production. We therefore estimate a linear regression relating $\hat{\beta}_p$ to the total national volume of service p, $\ln Q_p$.

To summarize size-linked variation in consumption patterns, we separately estimate the population elasticity of *consumption* per beneficiary for each procedure. That is, we estimate a Poisson model in which the outcome variable is the count of procedure p consumed by patients residing in region i, G_{pi}, per Medicare beneficiary residing there:

$$\ln \mathbb{E}\left[\frac{G_{pi}}{M_i}\right] = \zeta_p^C + \beta_p^C \ln \text{population}_i. \tag{2.10}$$

If $\beta_p^C \neq \beta_p$, there is size-predicted net trade in procedure p. Our model predicts that procedure frequency influences the pattern of trade, a prediction we test in Section 2.5.2.

35. In a robustness check, we have also estimated a zero-inflated Poisson model, to account for the possibility that fixed costs are especially important for the decision of whether to provide the first instance of a service in a region. These results (not reported here) are quite similar.

Results

Production per beneficiary rises with market size, especially for less common procedures. Figure 2.5(a) relates the population elasticity of production per beneficiary $\hat{\beta}_p$ for each procedure to its national volume $\ln Q_p$. Across all volumes, procedure output per beneficiary increases with market size. Less common procedures have higher elasticities, consistent with economies of scale that decline with quantity.

This finding raises questions about patients' access to care. What happens to patients who live in smaller markets but need rare services? To investigate this question, we estimate equation (2.10), the population elasticity of *consumption* per beneficiary of each procedure.

The population elasticity of consumption per beneficiary is smaller for the vast majority of procedures and less steeply related to a procedure's national frequency. Figure 2.5(a) also plots the population elasticity of consumption per beneficiary $\hat{\beta}_p^C$ for each procedure against its national volume $\ln Q_p$. While the relationship is negative, the slope for consumption is only one third that for production. Appendix Table 2.13 reports the production, consumption and trade patterns for two exemplar procedures: screening colonoscopy and LVAD implantation. Colonoscopies are common and geographically dispersed, while LVAD procedures are rare, geographically concentrated, and traded over longer distances.

We have thus far modeled patients as demanding (and providers as producing) specific service codes. An alternative view is that patients have a particular medical condition that requires treatment, but the patients may not know what particular care they need; they simply know they require care. As physicians might use different treatments across regions for the same condition, our estimates thus far could reflect substitution among procedures. We address this by conducting a similar analysis at the level of clinical condition.

Figure 2.5(b) shows production and consumption elasticities by diagnosis, rather than by procedure. The key patterns remain similar: production elasticities are higher than consumption and decline more rapidly with national patient volume. Both consumption and production elasticities have less steep relationships with national volume than for procedures. This could reflect measurement error within each category: the 482 diagnosis categories we use are far coarser than the 8,253 procedures in Figure 2.5(a). Alternatively, it could indicate true substitution among procedures within a condition that varies with location.

The contrasting population elasticities of production and consumption summarized in Figure 2.5 imply trade in medical services between markets of different sizes. Just as theories of trade with scale effects would predict, larger markets export rare procedures and smaller markets import them. For almost all procedures, production increases more than proportionately with market size. Consumption also increases more than proportionately with market size, but much less so than production. The differences between these elasticities mean net exports vary with market size. The implied net trade between markets of different sizes is particularly large for procedures that have small national volumes.

2.5.2 Market-size effects are stronger for less common procedures

Procedure-level variation in bilateral trade provides a finer test of how market-size effects depend on a procedure's frequency. Appendix Figure 2.15(a) shows a wide distribution of imports as a share of consumption by procedure.[36] We divide procedures into two equal-sized groups, common and rare, based on the quantity produced nationally and show each group's distribution of import shares across regions in Panel 2.15(b). The difference is dramatic: rare procedures (those with national frequency below the median) have much higher import shares, while the common procedures are overwhelmingly lower.[37] To formally test for differences in home-market effects, we again employ gravity models.

36. This kernel density plot exhibits a spike at just above 20%, indicating that trade is, quite common in most procedures. There is a long tail reaching all the way to 1 and also many procedures with few or even zero imports.

37. Nationally, the imported share of consumption is 22% for below-median-frequency procedures and 35% for those above the median. Within both groups of procedures, there is substantial variation in import shares across hospital referral regions.

Empirical strategy

To test the model's difference-in-differences prediction for trade volumes, we estimate market-size effects separately for common and rare services. We compute trade flows between each HRR pair $\overline{R}Q_{ijc}$ separately for these two categories of care, $c \in \{\text{common}, \text{rare}\}$. We thus have two observations for each ij pair, allowing us to estimate:

$$\begin{aligned} \ln \mathbb{E}\left[\overline{R}Q_{ijc}\right] = {} & \lambda_X \ln \text{population}_i + \lambda_M \ln \text{population}_j + \gamma X_{ij} \\ & + \left(\mu_X \ln \text{population}_i + \mu_M \ln \text{population}_j + \psi X_{ij}\right) \cdot \mathbf{1}(c = \text{rare}). \end{aligned} \tag{2.11}$$

An alternate specification introduces ij-pair fixed effects, which absorb all the covariates not interacted with $\mathbf{1}(c = \text{rare})$. The theory from Section 2.2.5 predicts stronger market-size effects for rare procedures, $\mu_X > 0$.

Results

Table 2.2 reports estimates for a gravity regression in which each pair of location has two observations: one for rare services and one for common. Column 1 repeats our baseline regression from Table 2.1 but with this new structure and obtains identical results. Column 2 limits the sample to pairs of location that have positive trade in at least one of the two procedure groups, which is the estimation sample used in the remainder of the table. In columns 3 and following, we interact both provider-market and patient-market population with an indicator for rare services. We find significant and robust evidence that the home-market effect is stronger for rare services. The coefficient on provider-market population increases by about 50% relative to common services. The coefficient on patient-market population shrinks by nearly half. Column 4 introduces location-pair fixed effects. Columns 5 and 6 are analogues of the previous two, but add a quadratic distance control. These results are statistically indistinguishable from the previous columns.

Table 2.3 shows that these results are robust to instrumenting for market size with either historical population or depth to bedrock. Columns 1 and 2 show estimates for common and rare services, respectively, when instrumenting for population in each region by its 1940 population. Columns 3 and 4 repeat the exercise using CBSAs rather than HRRs, and columns 5 and 6 switch to the bedrock-depth instrument. The results are consistent regardless of geographic unit or instrument. The estimates' stability suggests that neither the aggregate result nor the variation with procedure frequency is driven by anchor institutions or similar omitted variables.

The finding that less common procedures exhibit stronger home-market effects is robust to different ways of defining rare and common care. Table 2.4 demonstrates that our result holds when we look across diagnoses rather than procedures, and Appendix Table 2.14 shows the same when including facility spending. As with the production and consumption elasticities in Figure 2.5(b), the magnitude of the difference between rare and common care shrinks. This could reflect substitution across care within a diagnosis or a less precise classification of diagnoses than of procedures. But the qualitative pattern holds and remains significant, consistent with the model's difference-in-difference prediction.

These findings reflect each procedure's national frequency, not how often an individual patient receives the same procedure. We call the latter concept the procedure's "engagement". If patients are less willing to travel for high-engagement services and these services are more common, higher engagement could drive the stronger home-market effect we observe for rare procedures. In fact, the national frequency of a service has a very low correlation with various measures of engagement for that service, so it does not confound this result.[38] While the distance elasticity is more negative for high-engagement procedures, Appendix Table 2.15 shows that separating high- from low-engagement procedures does not meaningfully alter the estimated differential impacts of population size for rare procedures.

38. The correlation between the share of patients who had more than one claim for the procedure in a given year and the procedure's frequency is 0.14.

Figure 2.6 returns to categorizing services by frequency, reporting estimates of (2.8) separately for each national frequency decile. The blue circles show estimated provider-market population elasticities, which decline monotonically from the least common to most common procedures. The red squares show patient-market population elasticities, which increase across the frequency distribution. The difference between the respective coefficients demonstrates a strong home-market effect for all deciles. This effect is stronger the less common the procedure. Appendix Table 2.16 shows the same pattern among illustrative procedures.[39]

The potential concern about omitted cost shifters from Section 2.4.4 has an analogue here: Do the doctors who provide rare services benefit more from urban amenities than those providing common ones, lowering the cost of producing rare services in larger markets? This has facial plausibility if rare services are produced by elite specialists, who are higher-earning and more willing to pay for urban amenities through lower compensation.

Examining the population elasticities of physician earnings for each specialty alleviates this concern. If urban amenities drive specialists' locations, earnings elasticities should be negative, especially for rare specialties. But Appendix Figure 2.16 shows that the income elasticities are close to zero on average and uncorrelated with the specialty's national abundance. However urban amenities affect physicians' choices, they do not exhibit the compensating differentials necessary to explain the relationship between market size and specialization.

39. We show two common procedures—screening colonoscopy and cataract surgery—along with four rare ones: two treatments for brain cancer, implantation of a left ventricular assist device (LVAD), and total colectomy. All six procedures exhibit strong home-market effects, but the differences between $\hat{\lambda}_X$ and $\hat{\lambda}_M$ are smaller for the common procedures than the rare ones.

2.6 Estimating the scale elasticity of quality

To estimate the scale elasticity of regional medical services production, we first estimate each region's quality in Section 2.6.1. Section 2.6.2 describes our empirical strategy for estimating the scale elasticity, which Section 2.6.3 reports to be around 0.6 for aggregate medical services. Section 2.6.4 documents one mechanism linking these increasing returns and interregional trade: larger markets support a finer division of labor, and traded services are performed by more specialized and more experienced physicians.

2.6.1 Quality estimates

We use a two-step procedure, which begins by estimating a fixed-effects version of the gravity equation. In equation (2.7), the exporting region i component of the bilateral trade flow is its perceived service quality. We can thus estimate $\ln \delta_i$ as the origin fixed effect in this gravity equation. Similarly, $\ln\left(\frac{N_j}{\Phi_j}\right)$ can be estimated as a destination fixed effect, denoted $\ln \theta_j$. This implies the following estimating equation:

$$\ln \mathbb{E}\left(\overline{R}Q_{ij}\right) = \underbrace{\ln \delta_i}_{\text{exporter FE}} + \underbrace{\ln \theta_j}_{\text{importer FE}} + \gamma X_{ij}. \tag{2.12}$$

We interpret the exporter fixed effects $\widehat{\ln \delta_i}$ as a revealed-preference measure of quality, an interpretation we validate using hospital rankings and measures of physician specialization. The importer fixed effects $\widehat{\ln \theta_j}$, plus an assumption about potential market size, enable us to compute $\Phi_j = N_j/\hat{\theta}_j$, a measure of patient market access for those who reside in location j. We also estimate (2.12) separately by service frequency, yielding $\widehat{\ln \delta_i}^{\text{rare}}$ and $\widehat{\ln \delta_i}^{\text{common}}$.

To test whether $\widehat{\ln \delta_i}$ reflects quality, the first three panels of Figure 2.7 compare the estimated exporter fixed effects to external measures of regional hospital quality. We count the number of times each region's hospitals appear on *U.S. News* Best Hospitals.[40] We also obtain Hospital Safety Grades from the Leapfrog Group and average them by HRR. The significant positive slopes in both Figures 2.7(a) and 2.7(b) show that patients prefer to obtain care from HRRs with better *U.S. News* rankings. There is also a positive relationship with Hospital Safety Grades, shown in Figure 2.7(c).[41] The positive relationships with both measures suggest that our estimates capture a meaningful measure of hospital quality.

The *U.S. News* rankings are intended to capture the "Best Hospitals," a concept associated with providing highly specialized care. So it is natural that there is a stronger relationship between the *U.S. News* rankings and exporter fixed effects for rare services; the slope in Figure 2.7(b) is twice as large as that for common services in Figure 2.7(a).[42]

2.6.2 Empirical approach

We use the estimated exporter fixed effects $\widehat{\ln \delta_i}$ to examine the determinants of regional service quality, in particular the scale elasticity, α. In the free-entry condition (2.4), service quality in region i is an isoelastic function of the quantity produced, conditional on revenue, cost, and productivity shifters. Taking the log of (2.4) and rearranging terms yields an estimating equation for the quality-quantity relationship across locations:

$$\ln \delta_i = \alpha \ln Q_i + \ln \overline{R} - \ln w_i + \ln A_i. \tag{2.13}$$

Replacing $\ln \delta_i$ with its estimate $\widehat{\ln \delta_i}$ from (2.12) yields an estimating equation for $\hat{\alpha}$.[43]

40. Appendix 2.11.2 explains how we use these rankings.

41. The distance elasticity does not meaningfully vary with procedure frequency. This suggests that patients' preference for a particular region loads onto the region fixed effects, consistent with our interpretation.

42. In contrast, safety grades are not differentially relevant for rare services: Appendix Figures 2.18(a) and 2.18(b) show virtually identical slopes.

43. Appendix 2.11.5 quantifies the potential bias resulting from our observing only the quantity produced for Traditional Medicare beneficiaries, rather than the total quantity produced for all patients. It shows that the bias is small: the estimates in Table 2.5 should be deflated by about 5%.

One potential concern with estimating equation (2.13) by ordinary least squares is reverse causality. Shifts of the isocost curve would cause movements along the upward-sloping demand curve, biasing the estimated scale elasticity upwards. We address this with three instruments, starting with current population. Population is relevant for healthcare output and is valid if not correlated with healthcare quality other than by driving local demand. The "anchor institutions" concern discussed in Section 2.4.3 could violate this exclusion restriction, so we also use the historical population and bedrock-depth instruments.

Despite our instruments, other channels related to population size could generate the same relationship as the market-size effect we estimate. Most significantly, physicians might prefer to live in cities (Lee, 2010), regardless of patient demand. This could drive up quality in large markets, but through a different mechanism than the one we emphasize.

Before we address this problem, first note what is *not* a problem: physicians preferring to work in larger regions for job-related reasons. A larger population of patients allows physicians to specialize, conduct research, and train medical students. As discussed in Section 2.4.4, these forces operate through the scale of healthcare production in the region. Academic medical centers are often an important part of a region's medical industry. If their scale attracts workers, this is an agglomeration benefit α ought to capture.

The challenge to our interpretation arises if physicians prefer larger markets for non-professional reasons, and this labor supply shift increases quality. If urban amenities attract physicians—and higher-quality physicians in particular—this would represent variation in w_i or A_i that is correlated with population size and hence local output in equation (2.13). The analysis of local costs in Section 2.4.4 and Appendix Figure 2.14 mitigates this concern.

2.6.3 Scale improves quality

Estimated service quality $\widehat{\ln \delta_i}$ rises substantially with the regional volume of production $\ln Q_i$. Figure 2.7(d) depicts this relationship and Table 2.5 reports regression estimates.

The estimated scale elasticity is around 0.6 and stable under various estimation approaches. The first row uses OLS, while subsequent rows instrument for output using contemporaneous or historical population. The first and third columns omit the diagonal Q_{ii} observations when estimating the gravity equation (2.12), to avoid any bias from having a region's own local consumption influence both the quality measures and output. The third and fourth columns control for spatial variation in reimbursements. Across twelve estimates, the lowest elasticity is 0.53 and the highest is 0.97. Instrumenting for output tends to reduce the estimated scale elasticity. Excluding the diagonal of the trade matrix when estimating quality tends to raise it. The results for CBSAs, reported in Appendix Table 2.17, are also stable across specifications and when using the alternative bedrock instrument. While the existence of home-market effects implied local increasing returns, these estimates quantify their magnitude.[44] These estimates are central to our counterfactual calculations in Section 2.7.

The second panel of Table 2.5 estimates the scale elasticity for rare services. Section 2.2.5 shows that market-size effects are larger for rarer procedures even if all procedures have the same scale elasticity. These differences are amplified if rarer procedures have a larger scale elasticity than more common procedures. The scale elasticity is indeed substantially larger for rare services, with estimates centered around 0.9.

2.6.4 Scale facilitates the division of labor

One source of increasing returns—though certainly not the only one—could be division of labor among physicians. In particular, the specialized labor required to produce rare services could drive the patterns we found in Section 2.5 across treatments and diagnoses. Specialized services may require physicians with specific training, whom low demand in smaller HRRs may not support (Dranove, Shanley, and Simon, 1992).

44. These estimates lie in the middle of other estimated agglomeration elasticities, Kline and Moretti (2013) estimate an elasticity of 0.4–0.47 from the Tennessee Valley Authority's investments. In manufacturing, Greenstone, Hornbeck, and Moretti (2010) report an analogous elasticity above 1 (a 12% increase in total factor productivity caused by adding a plant representing 8.6% of the county's prior output).

Specialization as a source of local increasing returns

To study this mechanism, we estimate the population elasticity of physicians per capita for each specialization and relate it to the number of physicians in the specialization. Let Y_{si} be the number of doctors of specialty s in location i.[45] We estimate a Poisson model,

$$\ln \mathbb{E}\left[\frac{Y_{si}}{\text{population}_i}\right] = \zeta_s^S + \beta_s^S \ln \text{population}_i, \tag{2.14}$$

for each specialty s by maximum likelihood.

Figure 2.8(a) shows a clear negative relationship between a specialty's per capita population elasticity $\hat{\beta}_s^S$ and the national number of physicians in that specialization.[46] A natural explanation for rare procedures and rare specializations both being geographically concentrated in larger regions is that the size of the market limits the division of labor. To the extent that producing rare procedures requires specialized physicians, a larger volume of patients makes production economically viable.

Consistent with this idea, Appendix Figure 2.17 shows the number of distinct procedures produced as a function of market size by procedure type. We group procedures into seven categories, count the number of procedures produced in each region, and project these onto regional population.[47] Larger regions produce a greater variety of procedures in all seven categories. If physicians specialize in particular procedures, this makes sense: larger markets

45. Data come from the National Plan and Provider Enumeration System (NPPES) data, which cover all physicians, not just those serving Medicare patients. These data only report the number of doctors/specialists and their location, but contain no further information about procedures performed. We restrict attention to the 223 specializations within Allopathic & Osteopathic Physicians. We restrict attention to national provider identifiers of the "individual" entity type (as opposed to "organization"). We consider each physician's primary specialty, as indicated in the NPPES file. Results (unreported) are similar when we allow for multiple specialties per physician, a common occurrence in the NPPES data.

46. This pattern is not attributable to spatial sorting driven by rare specialties commanding higher earnings. In fact, a specialty's number of physicians and mean earnings are uncorrelated. Appendix Table 2.19 shows that controlling for a specialty's earnings has no effect on the negative relationship between population elasticity and number of physicians across specialties.

47. Appendix Table 2.20 reports regression estimates for these relationships.

have more specialties of physicians and thus a greater ability to provide rare procedures.

This evidence on specialization does not preclude other agglomeration mechanisms from also playing a role. Lumpy capital, knowledge diffusion (Baicker and Chandra, 2010), and thicker input markets could also be important productivity benefits of scale. We focus on specialization and physician experience because of their close link to the procedure-level agglomeration we analyze and we can observe them in claims data.

Imports are specialist-intensive

We next ask whether the distribution of specialties helps explain trade. Figure 2.8(b) shows the share of imports and of local consumption that are provided by specialists as a function of regional population.[48] Imports are significantly more specialist-intensive than local production. This difference is especially pronounced in the smallest regions, and it remains true throughout the population distribution.

Does trade match patients with the appropriate specialist? Among all specialty care, we determine the two most common specialties to provide each unique service and label these the "standard" specialties for that care. We then determine whether each instance of the treatment was provided by a standard or non-standard specialty.

Figure 2.8(c) shows the share of imports and of local care provided by the standard specialties. Imports are more likely to come from the standard specialist than local care, and the distinction is especially pronounced in the smallest regions. The difference is substantial: Local care in the smallest regions is 40% more likely to be provided by a non-standard specialist than in the largest regions (7.0% vs. 5%). When importing medical services, this share falls to 5%—indistinguishable from the largest regions' local care.

We conduct a similar analysis based on provider experience. Using the public Medicare provider data (based on all Traditional Medicare patients), we count the number of times

48. We define "specialist" to mean all physicians except those whose primary specialty is internal medicine, general practice, or family practice.

the physician billed for the specific service in the previous year. We divide this experience measure by the procedure's national mean and average it across all procedures provided to patients in an HRR. Figure 2.8(d) shows that, at all population sizes, care imported from other regions is produced by more experienced providers than locally produced care.[49] Patients in larger regions see more experienced providers for both imported and locally produced care.

Specialists are disproportionately located in larger markets, as are physicians with more experience in any given procedure. Since imported care is predominantly specialty care, and provides patients access to this higher experience, we conclude that visiting the appropriate specialist based on training or experience is part of the value proposition for trade in medical care. This provides a second validation of our interpretation that trade reflects quality variation. Patients travel to regions with highly-ranked hospitals, which larger markets tend to have—along with the ability to provide rare services. This market-size effect strongly predicts gross and net exports. Together, this suggests that economies of scale play an important role in increasing the quality of care, and trade between regions enables patients from many regions to share the benefits of this agglomeration.

2.7 Tradeoffs and counterfactual scenarios

Given the estimated strength of local increasing returns, geographically concentrating healthcare production has substantial benefits. Larger regions support specialists, house experienced physicians, and produce more specialized procedures. But this geographic concentration implies that patients in smaller regions may suffer from limited access to care. We use observed trade flows and our estimates of the scale elasticity α and region-specific qualities δ_i to quantify how various counterfactual policy scenarios would change each region's patient

49. This comparison restricts attention to procedures that are performed in all hospital referral regions (143 procedures). Thus, regional variation does not reflect the fact that larger markets produce a greater number of distinct codes (Appendix Figure 2.17).

market access. Our results underline the importance of distinguishing between the quality of locally produced services and the quality of services to which local residents have access.

We compute counterfactual equilibrium outcomes relative to the baseline equilibrium. For the baseline equilibrium, define export shares $x_{ij} \equiv \frac{Q_{ij}}{\sum_{j'} Q_{ij'}}$ and import shares $m_{ij} \equiv \frac{Q_{ij}}{N_j}$. For every variable or parameter y, denote the ratio of its counterfactual value y' to its baseline value y by $\hat{y} \equiv \frac{y'}{y}$. Appendix 2.12.1 shows how we solve for the relative counterfactual endogenous qualities ($\hat{\delta}$) using baseline equilibrium shares (x_{ij}, m_{ij}), the scale elasticity (α), and relative counterfactual exogenous parameters ($\hat{A}, \hat{R}, \hat{w}, \hat{\rho}, \hat{N}$). In particular, counterfactual qualities are given by a system of $\mathcal{I}$ equations with unknowns $\{\hat{\delta}_i\}_{=1}^{\mathcal{I}}$:

$$\hat{\delta}_i = \left(\hat{R}_i \hat{A}_i / \hat{w}_i\right)^{\frac{1}{1-\alpha}} \left(\sum_{j \in \mathcal{I}} \frac{x_{ij} \hat{\rho}_{ij} \hat{N}_j}{m_{0j} + \sum_{i' \in \mathcal{I}} m_{i'j} \hat{\delta}_{i'} \hat{\rho}_{i'j}}\right)^{\frac{\alpha}{1-\alpha}}.$$

The first term of this expression, $\left(\hat{R}_i \hat{A}_i / \hat{w}_i\right)^{\frac{1}{1-\alpha}}$, shows that the scale elasticity α governs the effect of exogenous supplier shifters, including reimbursements $\hat{R}_i$, on quality produced in a region. Reimbursement rates shift the scale of production, and stronger scale economies (higher α) amplify these shifts. The second term shows how changes in other regions influence local outcomes through trade, combined with scale. Thus, our counterfactual scenarios rely on both our estimates of the scale elasticity α and observed trade patterns.[50]

We first consider the impact of a nationwide change in reimbursements. Increasing reimbursements uniformly by 10% has heterogeneous effects. Figure 2.9(a) depicts the change in output quality in each region. Remote, rural areas tend to experience the largest increases in output quality δ_i. Large regions such as Boston, New York, Atlanta, and Florida have the smallest increases, because they produce more care at baseline.

50. In order to compute import shares, we assume that the number of potential patients is proportional to the number of enrolled Traditional Medicare beneficiaries. See Appendix 2.12.2 for details. The qualitative and spatial patterns of counterfactual outcomes do not depend on what share of potential patients we assume choose the outside option. Appendices 2.12.3 and 2.12.4 generalize this method of computing counterfactual outcomes to the model with multiple types of patients introduced in Appendix 2.10.2.

Figure 2.9(b) shows the impact on patient market access is nearly opposite: regions with the largest increase in output quality have the smallest improvements in market access. Their residents already had high import shares, so the least reliance on local production. The increase in local quality thus has limited impact on their overall market access. For patients who switch to consuming local care, the gains are modest as local production is still lower-quality than the care they otherwise import. In contrast, patients in Houston, Dallas, or Florida had limited reason to travel. The increase in δ_i due to higher local reimbursements, even if modest, improves their access relatively more.

Figure 2.9(c) summarizes these contrasting changes in output quality and patient access. Regions with the lowest initial patient market access Φ_i have the biggest increase in local production quantity and quality, $\hat{Q}_i$ and $\hat{\delta}_i$, but the smallest increase in patient market access $\hat{\Phi}_i$. Appendix Figure 2.19 conducts this exercise separately for rare and common services. The patterns are qualitatively similar, but the impacts on quality are much larger for rare services because of their larger scale elasticity ($\alpha = 0.9$ rather than $\alpha = 0.6$), more concentrated baseline production, and higher baseline trade shares.

These results help reconcile two notable aspects of US healthcare policy. First, a range of recent studies find medical outcomes that match our predictions: patients who travel farther for care in larger markets tend to have better outcomes (Battaglia, forthcoming; Fischer, Royer, and White, 2022; Petek, 2022). Second, there is nevertheless a major political and policy effort to subsidize production in rural areas.[51] Our contrasting results for output and access rationalize this pattern: *producers* in rural areas are especially dependent on high reimbursements. This naturally leads to political pressure to subsidize production in such places. But *patients* do not necessarily benefit. They would often benefit from traveling to

51. These policies include Critical Access Hospitals, Health Professional Shortage Areas, rural-biased adjustments to Medicare's Geographic Practice Cost Index for physician work, hospital geographic reclassification for Medicare reimbursements, increasing residency slots in rural areas, and more federal and state programs. The effectiveness of these policies is not always clear (Khoury, Leganza, and Masucci, forthcoming; Falcettoni, 2021).

larger markets for better care, suggesting that the emphasis on local production may not be efficient—even from the perspective of rural patients.

We next consider the impact of this nationwide reimbursement increase on different income groups, indexed by κ. We compute changes in market access for each region and income group, and rescale them into percentage changes, $100(\hat{\Phi}_{j\kappa} - 1)$. At the region-by-income-tercile level, Table 2.6 regresses these changes on income-group dummies (columns 1–3) and HRR fixed effects (columns 2–3). Column 1 shows that the market access gain for the highest income tercile is nearly 20% larger than for the lowest tercile. The lowest tercile experiences an 8.8% increase in patient market access (the constant in the regression). The highest tercile gains this plus an additional 1.6 percentage points. The difference is explained by differences in the groups' outside option shares, $m_{0j\kappa}$, as column 3 shows. Patients in the highest tercile are more likely to seek care, so benefit more from quality improvements.

Policies often target specific regions so we now examine how targeted production subsidies affect output quality and patient access. Figure 2.10 contrasts the consequences of raising reimbursements by 30% in Boston and in Paducah, Ky. Figure 2.10(a) depicts the impact on quality of care in each region relative to its baseline value in the Boston scenario. Free entry means that higher reimbursements translate to higher-quality care produced in Boston. Quality declines in the rest of New England as patients substitute away and scale economies translate lower volumes into lower quality (an "agglomeration shadow", as in Fujita and Krugman, 1995). These effects diminish with distance to Boston.

Regions that experience larger declines in output quality due to Boston's expansion simultaneously experience larger improvements in patient market access. Figure 2.10(b) depicts the change in the value of patients' market access, $\hat{\Phi}_i$. Patients in Boston benefit the most from the higher reimbursement of their local production. Outside Boston, regional changes in patient market access are nearly opposite the changes in local output quality. The nearest regions import sufficient volumes that the benefits of improved quality in Boston exceed the declines in the quality of local production, causing their patient market access to improve. Regions closer to Boston experience larger declines in the quality of local production precisely because their residents' choice sets improve more, spurring more substitution. In more distant regions, the welfare impacts are neutral to ever-so-slightly negative.

We again see disproportionate gains for patients who live in higher-income neighborhoods, shown in columns 4–6 of Table 2.6. The value of market access increases by 0.098% for first-tercile patients nationwide; this is orders of magnitude lower than in columns 1–3 because only one region's reimbursement is increasing. Gains are 70% larger for third-tercile patients. Once again this can be explained by baseline trade shares, as column 3 shows.

The consequences of higher reimbursement rates in Paducah, Ky. exhibit very different spatial patterns than in Boston. Figures 2.10(c) and 2.10(d) depict the regional changes in output quality and patient market access, respectively, caused by a 30% reimbursement increase in Paducah. Unlike Boston, Paducah is a net importer: its consumption of medical services exceeds local production by more than one-third. Higher reimbursements that improve output quality in Paducah cause Paducahans to reduce their imports from neighboring regions. This reduces the quantity produced in neighboring regions, lowering their output quality, similar to the regional spillovers in the Boston scenario. But Figure 2.10(d) shows that those regions where output quality declines more are the regions where patient market access declines more, contrary to the pattern of outcomes in the Boston scenario.

The contrasting outcomes reflect trade flows in the baseline equilibrium: Boston is a net exporter of medical services and Paducah is a net importer. Paducah imports one-third of its consumption, and Boston imports only six percent. Higher reimbursements in Boston cause output quality declines in nearby regions—largely because residents of those regions import more when Boston's quality improves. In contrast, higher reimbursements in Paducah reduce neighboring regions' output quality largely because Paducah residents demand fewer exports from these regions when Paducah's quality improves. Nearby regions import little from Paducah, so they benefit little from its improved quality. Appendix Figure 2.20 shows that the lessons from Boston and Paducah generalize: the pattern of spillovers from increasing reimbursements in one region is driven by that region's net trade in medical care. To summarize, the spillover consequences of subsidizing production in one region depend on the pattern of trade; changes in regional output quality need not align with changes in regional patient market access.

The distributional consequences of region-specific subsidies depend on which region is subsidized. We compute the nationwide gains in market access from subsidizing production in each region, one at a time. Figure 2.11 shows this gain, scaled by the increase in total spending, as a function of region size. The aggregate gain in market access per dollar spent is higher in larger markets: further concentration of production has larger benefits. The graph also shows the gains per dollar separately by income tercile. Unlike the Boston scenario, in which benefits accrue more to higher-income ZIP codes, subsidizing production in less populous regions benefits lower-income ZIP codes more. These contrasts reflect geographic divides in incomes: lower-income patients are more likely to live in and near smaller regions.

Rather than subsidizing local production, policies might improve patient market access in a particular region by facilitating trade. We examine the consequences of a policy that reduces travel costs for Paducahans obtaining care elsewhere (specifically, $\hat{\rho}_{i,\text{Paducah}} = 1.3$ when $i \neq \text{Paducah}$).[52] Figure 2.12 shows that, unlike an increase in Paducah reimbursements, this policy has positive spillovers on neighboring regions. These regions increase their exports to Paducah, and thus their own scale and quality. This improves their residents' market access.[53] Because lower-income patients are more sensitive to distance and are less likely to import care from other regions, a larger travel subsidy is necessary to achieve the

52. The impact of this change on Paducah residents' market access $\hat{\Phi}_{\text{Paducah}}$ is similar to an 8% increase in reimbursements in Paducah.

53. Recall that our model assumes elastic supply. The short-run impact on exporters may be more complex if there are short-term diseconomies of scale due to crowding or queuing.

same percentage improvement in their patient market access. To increase each income tercile's patient market access in Paducah by 7%, one would need to reduce trade costs by 40% for the first income tercile and by 37% for the third income tercile.

So this policy benefits both Paducah and its neighbors—though we do not estimate the costs of this travel subsidy. But facilitating travel reduces the quantity—and thus the quality—produced in Paducah. Analysts looking at the impact of travel subsidies on the quantity or quality of care provided in Paducah itself would reach very different conclusions than those looking at the impact on patient market access.

These counterfactual scenarios are subject to significant caveats, and we have not attempted to identify the optimal policy. Even so, this simple model rationalizes important aspects of the economics and politics of US healthcare policy. The counterfactual scenarios highlight our main findings: Healthcare production has substantial local increasing returns, and patient travel plays a meaningful role in enabling access to higher-quality care. Given these economic mechanisms, regional spillovers are larger when economies of scale are stronger, depend on the pattern of trade flows, and differ depending on whether policies subsidize production or travel. This shows the importance of distinguishing between regional output quality and regional patient access when evaluating healthcare policies.

2.8 Conclusion

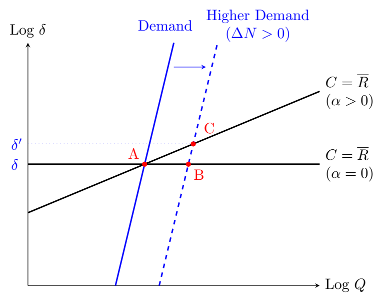
Smaller markets have fewer specialized physicians, produce less medical care per capita, and have worse health outcomes than larger markets. Thanks to trade in medical services, less production does not translate one for one into less consumption of medical services. Instead, trade affords patients who live in smaller markets access to higher-quality care. This higher quality comes in part from consuming services that would otherwise be unavailable, visiting appropriate specialists, and accessing experienced physicians.

This trade amplifies the scale advantages of large markets and hence the quality of care they produce. This means the healthcare industry can serve as an export base for large cities. Substantial scale economies also imply that policies to reallocate care across regions may impact the quality of care available. We simulate policies that aim to improve care access in "under-served" markets. The rich and varied patterns of welfare consequences when subsidizing production or travel highlight the importance of trade and agglomeration for the incidence of these policies on patients and producers.

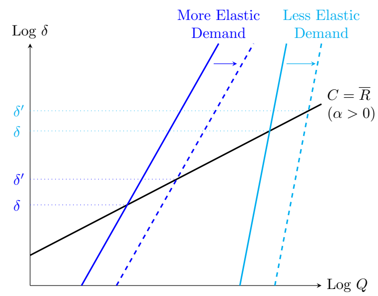
2.9 Figures and Tables

Figure 2.1: Illustrative model diagrams

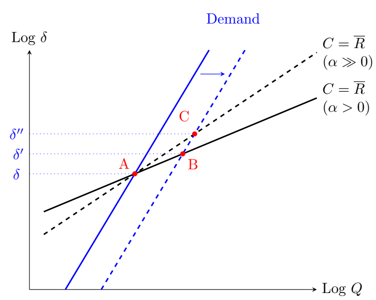
(a) Autarky: Constant vs. increasing returns



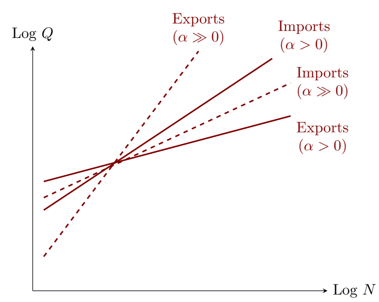
(b) Autarky: Market size and demand elasticity



(c) Quality and quantity depend on scale elasticity



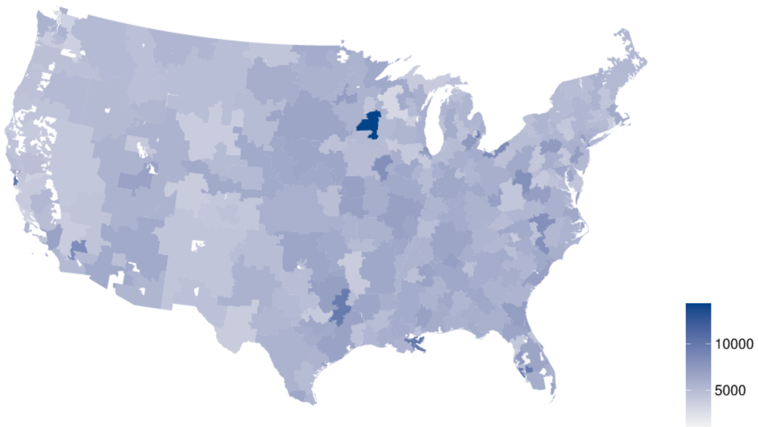
(d) Exports and imports



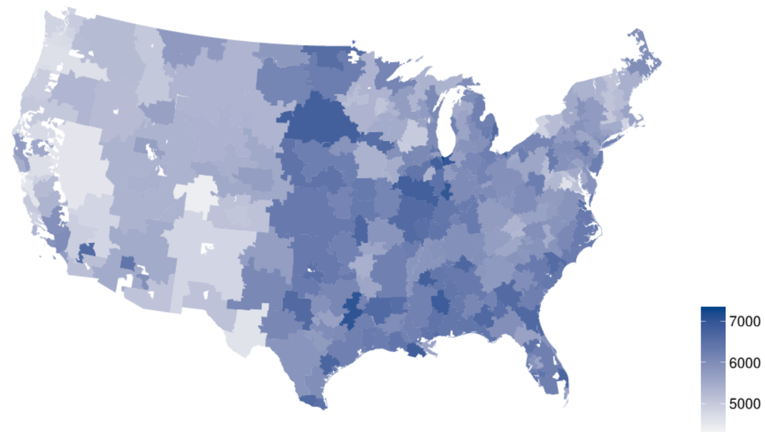
Notes: This figure depicts how increasing demand in one region affects its equilibrium outcomes. In panels (a)–(c), quantity produced Q is on the horizontal axis and service quality δ is on the vertical axis. The black lines depict the free-entry isocost curve, $C = \overline{R}$, given by equation (2.3). The blue and cyan lines depict demand for the region's service, which we depict as log-linear for visual clarity. (The logit demand function is actually log-convex, which is consistent with all the depicted comparative statistics.) Equilibrium is the intersection of the demand and isocost curves. An increase in demand is the rightward shift from the solid to the dashed demand curve. This shift increases equilibrium quality from δ to δ'. Panel (a) shows that higher demand elicits higher quality if there are increasing returns to scale. Panel (b) shows that this quality improvement is larger when demand is more elastic. Panels (c) and (d) introduce trade and compare the extent of quality improvement under two different magnitudes of increasing returns ($\alpha > 0$ and $\alpha \gg 0$). These magnitudes govern the patterns of interregional trade, shown in panel (d) as a function of the number of potential patients N. Imports from other regions rise with N. With increasing returns to scale ($\alpha > 0$), exports to other regions also rise with N (a weak home-market effect). When the scale elasticity α is larger ($\alpha \gg 0$), the import curve is flatter and the export curve is steeper. With sufficiently strong increasing returns, an increase in local demand causes a greater increase in exports than imports (a strong home-market effect).

Figure 2.2: Production, consumption, and trade across regions

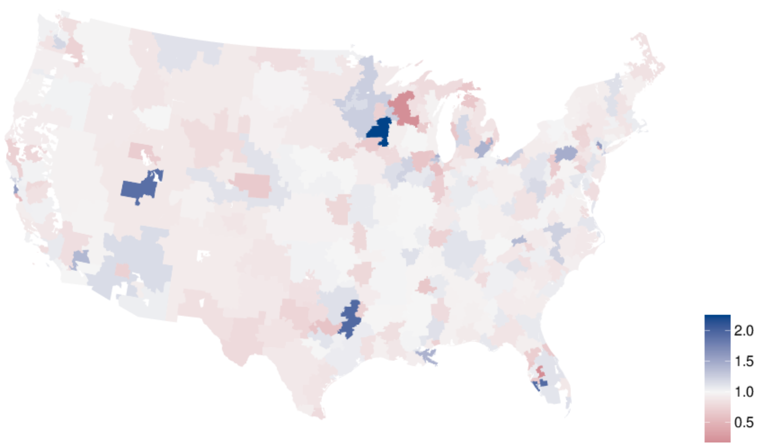
(a) Production per capita



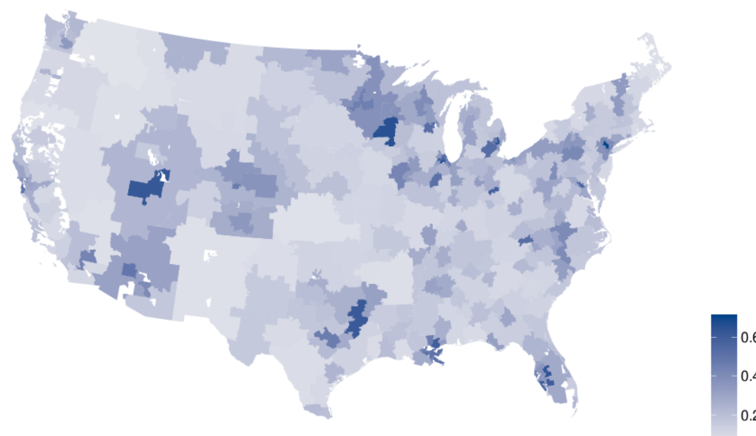
(b) Consumption per capita



(c) Production divided by consumption

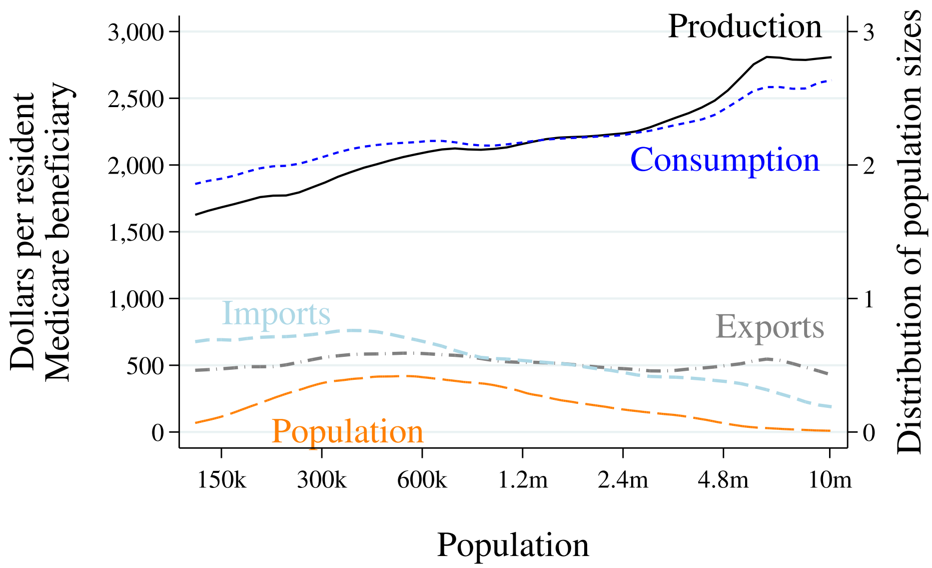


(d) Gross exports relative to production



Notes: Panel (a) shows production per capita, including professional and facility fees. The hospital referral region (HRR) of production is the location where the service is provided. Panel (b) shows consumption per capita, including professional and facility fees. The HRR of consumption is based on the patient's residential address. Colors depict deciles of production per capita in both Panels (a) and (b). Panel (c) shows the ratio of production per capita to consumption per capita for professional services. Panel (d) shows gross exports as a share of total production by HRR for professional services. Data come from the Medicare 20% carrier Research Identifiable Files. All calculations exclude emergency-room care and skilled nursing facilities. Expenditures are computed by assigning each procedure its national average price. HRR definitions are from the Dartmouth Atlas Project.

Figure 2.3: Production and consumption of medical care across regions

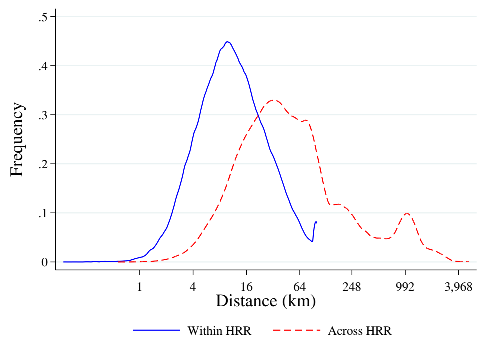


Population elasticity (log−log regression slope) of transactions per resident Medicare beneficiary:
Production: 0.13 (0.02), Consumption: 0.06 (0.01)
Exports: −0.00 (0.05), Imports: −0.25 (0.03)

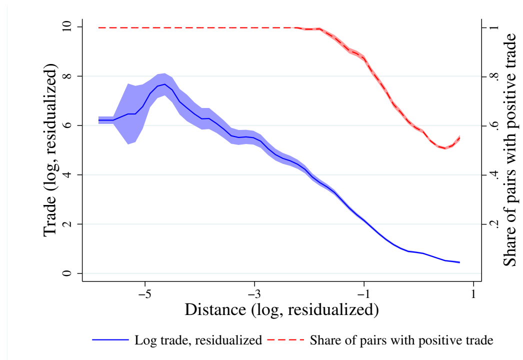
Notes: This figure shows production, consumption, and trade per capita of Medicare services across hospital referral regions (HRRs) of different sizes, all smoothed via local averages. We use the Medicare 20% carrier Research Identifiable Files to compute the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. The black series shows production of medical care per Medicare beneficiary residing in the HRR of production. The blue series shows consumption of medical care per Medicare beneficiary residing in the HRR of consumption. The dashed dark-gray series shows interregional "exports" of medical care and the dashed light-blue series shows interregional "imports" of medical care, again per Medicare beneficiary. The orange series depicts the distribution of HRR population sizes. HRR definitions are from the Dartmouth Atlas Project.

Figure 2.4: Patients travel between regions and trade declines with distance, more so for lower-income patients

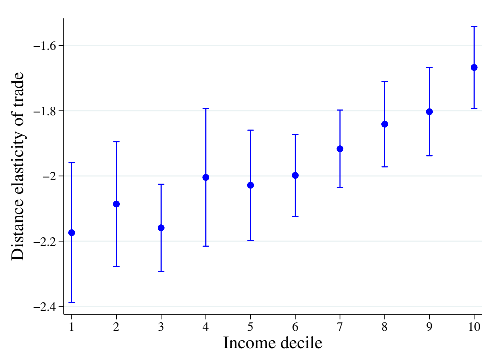
(a) Distribution of travel distances within and across HRRs



(b) Trade volume and extensive margin by distance



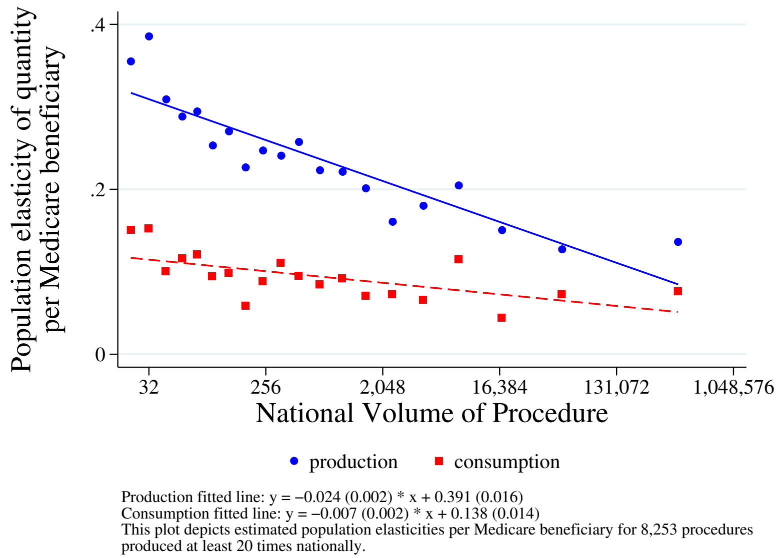
(c) Higher-income patients are less sensitive to distance



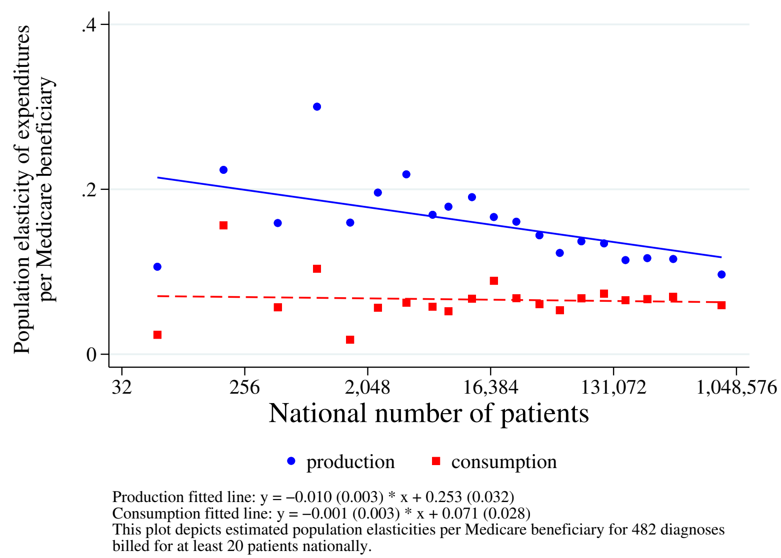
Notes: Panel (a) shows the distribution of patients' travel distances when patients obtain care within their home HRR (blue distribution) and when they travel across HRRs (red distribution). Travel distances measure the distance between home and treatment locations. For travel within a hospital referral region, the distance measure reflects the distance between the centroid of the patient's residential ZIP code and the ZIP code of the service location. We use ZCTA-to-ZCTA distances downloaded from the National Bureau of Economic Research; those exceeding 160 kilometers are winsorized at 160 kilometers. For travel across HRRs, we use ZCTA-to-ZCTA distances when they are within 160 kilometers and (for computational ease) use HRR-to-HRR distances beyond 160 kilometers. In Panel (b), the blue series depicts the volume of trade against distance, after conditioning out the fixed effects in equation (2.12), for positive-trade pairs of locations. The red series shows the share of HRR pairs with positive trade as a function of the distance between them, after conditioning out the importer fixed effects and exporter fixed effects, as in equation (2.12). Panel (c) depicts the coefficient on log distance obtained by estimating equation (2.12) separately for each decile of the national ZIP-level median-household-income distribution. The 95% confidence intervals are computed using standard errors two-way clustered by both patient HRR and provider HRR. Patients from higher-income ZIP codes are less sensitive to distance.

Figure 2.5: Population elasticities of production and consumption

(a) Population elasticities by procedure frequency

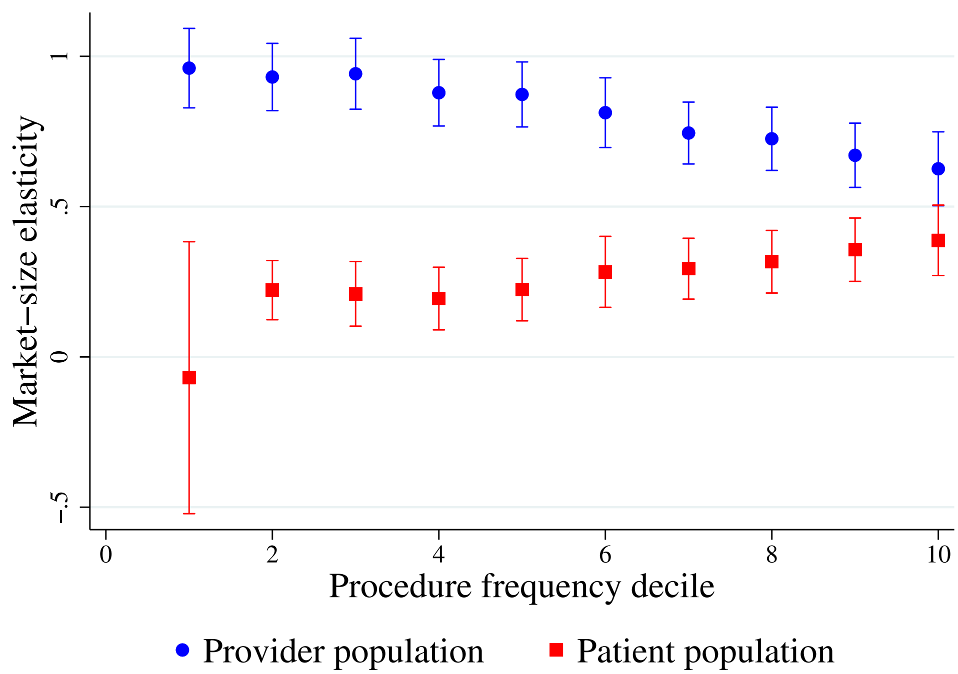


(b) Population elasticities by diagnosis frequency



Notes: The vertical axis of both panels plots the population elasticities of quantity of medical care produced and consumed per local Medicare beneficiary. The elasticities are computed using the Poisson models in equations (2.9) and (2.10) based on production location and patients' residential location, respectively. Panel (a) estimates these elasticities for each of the procedures provided at least 20 times nationally in the Medicare data. The horizontal axis shows the total national volume of physician services for the procedure. Panel (b) estimates the elasticities for care provided to treat each of the Clinical Classifications Software Refined (CCSR) diagnoses billed for at least 20 patients nationally in the Medicare data. Expenditures are computed from the Medicare 20% carrier Research Identifiable Files using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. The horizontal axis shows the total number of patients nationally with the diagnosis. In both panels, the blue dots are a binned scatterplot of the estimated population elasticity of production per beneficiary as a function of the national volume. The red dots are the same for consumption (residential location)-based estimates. There is a significant negative relationship for production, indicating that production elasticities are highest for rare services and rare diseases. The relationship for consumption is much more modest. The difference between these estimates must be driven by trade between locations.

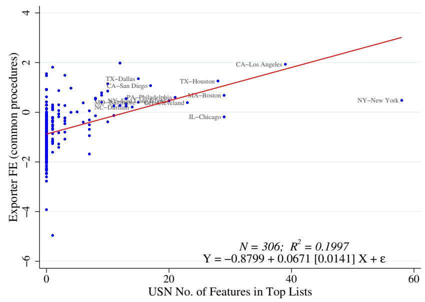
Figure 2.6: The home-market effect is stronger for rarer procedures



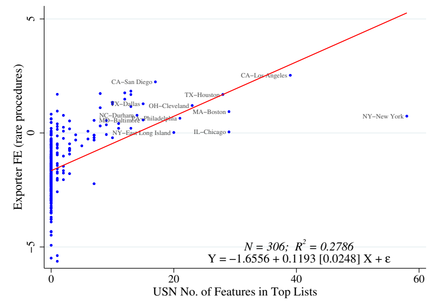
Notes: This figure groups non-emergency physician-provided services in the Medicare claims data into deciles based on the national frequency of each procedure. For each decile, we estimate equation (2.8), testing for a home market effect, and plot the estimated coefficients on provider and patient market log population with their 95% confidence intervals. The coefficients on provider-market size always exceed the respective coefficients on patient-market size, indicating a strong home-market effect. The coefficients on provider-market size monotonically decrease across the deciles. The coefficients on patient-market size monotonically increase across the deciles. Together, these two patterns show that the home-market effect is stronger the less common the procedure is, in line with the theoretical difference-in-difference prediction.

Figure 2.7: Estimated quality is positively correlated with total output and external quality metrics

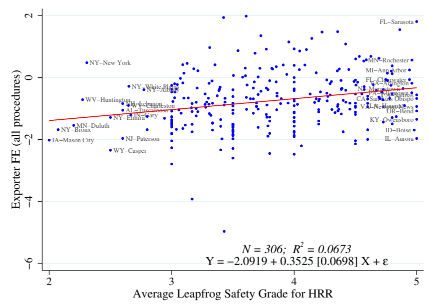
(a) *U.S. News* vs. quality estimated for common services



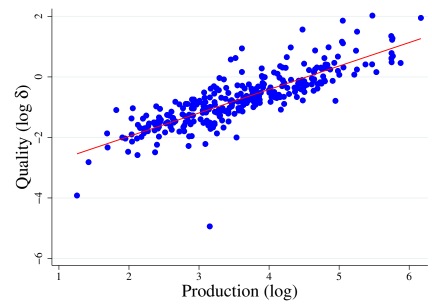
(b) *U.S. News* vs. quality estimated for rare services



(c) Leapfrog Safety Grade vs. estimated quality



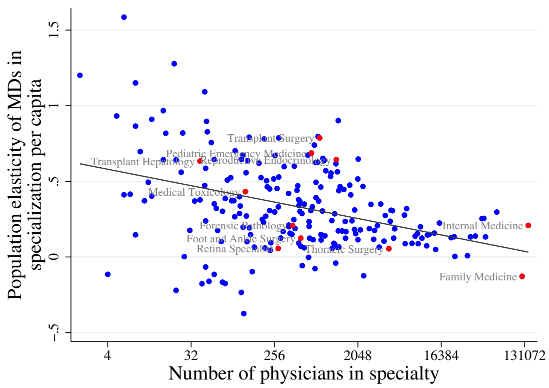
(d) Quality is higher in regions producing more output



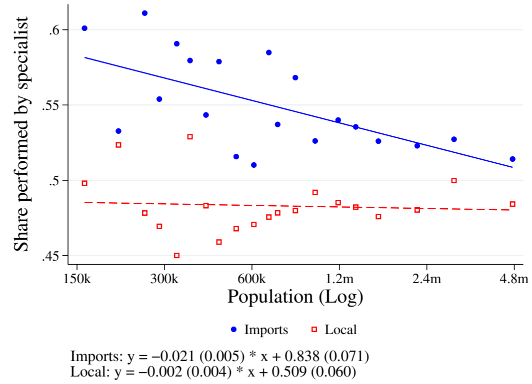
Notes: The first three panels show the relationship between the exporter fixed effects (our revealed-preference measure of quality) and external quality measures. The vertical axis shows the exporter fixed effects for each HRR estimated using trade in common services in Panel (a), using trade in rare services in Panel (b), and for all services in Panels (c) and (d). The horizontal axis in Panels (a) and (b) is a count of the number of times each region's hospitals appear on the *U.S. News* list of best hospitals. *U.S. News* produces an overall ranking as well as rankings for 12 particular specialties. We count the number of times each HRR's hospitals appear on any of these 13 lists. Both panels show a positive relationship, indicating that patients travel farther to obtain care from regions highly ranked by *U.S. News*. The relationship is stronger for rare services, as the slope is nearly double that for common services. The horizontal axis in Panel (c) is the average safety grade for hospitals in an HRR (mapping A=5, B=4, etc.), for grades determined by the Leapfrog Group. These are positively correlated with exporter fixed effects. Panel (d) shows the relationship between production and the exporter fixed effects from equation (2.12), across HRRs. HRR production is measured as Medicare output produced (in millions US dollars) for non-emergency physician services in the 20% carrier file.

Figure 2.8: Imports are specialist-intensive, especially in smaller regions

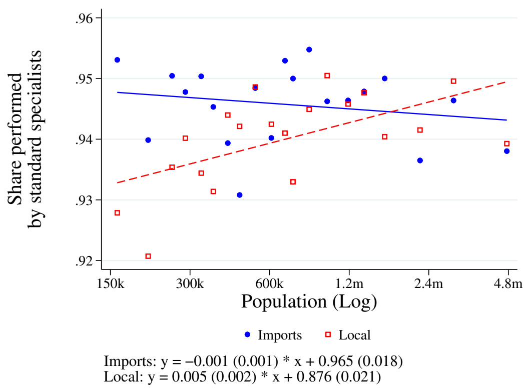
(a) Population elasticities of physician specializations



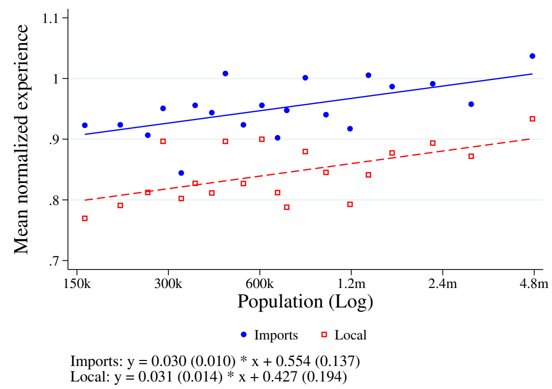
(b) Specialty care imports



(c) "Standard" specialty care



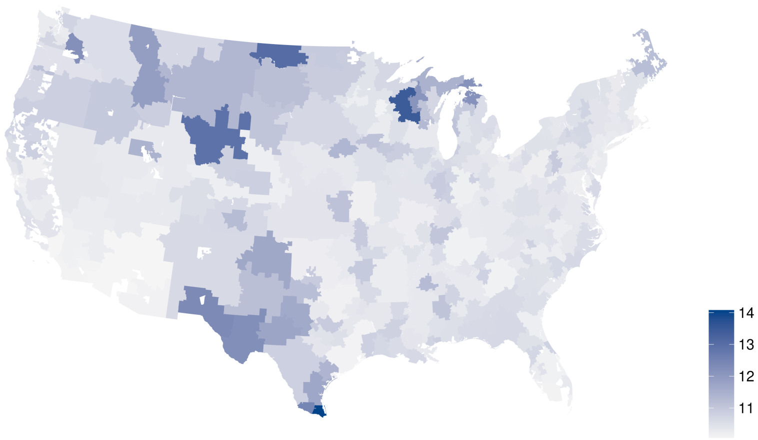
(d) Provider experience



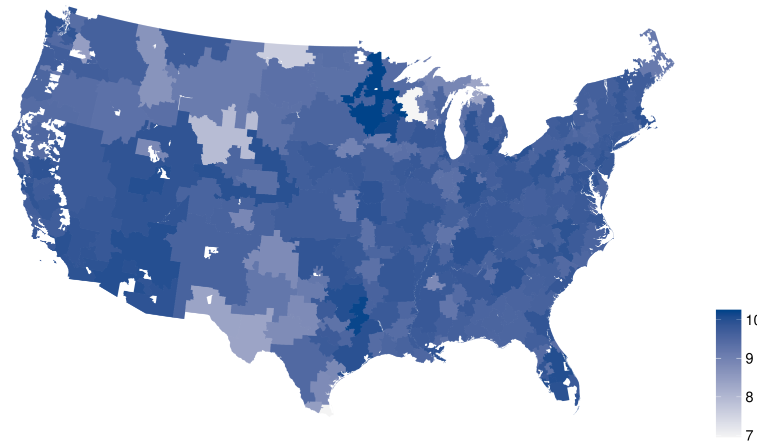
Notes: The vertical axis of Panel (a) depicts the population elasticities of quantity of physicians in an HRR. The population elasticities are computed for each specialty using the Poisson model in equation (2.14). The horizontal axis shows the nationwide number of physicians in each specialty. The negative relationship indicates that rare specialties are disproportionately concentrated in high-population regions. Panel (b) shows the share of procedures that are performed by a specialist, for imports and locally produced procedures, by market size. We define generalists as internal-medicine, general-practice, and family-practice physicians and define specialists as all other physicians. Imports are more likely to be performed by a specialist, and smaller markets' imports especially so. Panel (c) examines procedures that are typically performed by specialists, and classifies the "standard" specialists as the top two specialties performing the procedure nationally. It shows the shares of procedures performed by the "standard" specialties in imported specialty care and locally produced specialty care as a function of local population size. Imports are more likely to be performed by "standard" specialties, especially for smaller regions. Panel (d) shows the mean relative experience of providers for care produced locally and imported by population size of the patient's region. This panel describes only procedures that are performed in all hospital referral regions (143 procedures). In public-use Medicare data, we define a provider's experience for a given procedure as the number of times they performed the procedure for Traditional Medicare patients in the prior calendar year. Before aggregating to the regional level, we rescale experience in each procedure so that its mean is one. On average, patients in larger markets obtain treatment from more experienced providers. At all population levels, imported care is produced by more experienced providers than local care.

Figure 2.9: Counterfactual outcomes when reimbursements increase 10% everywhere

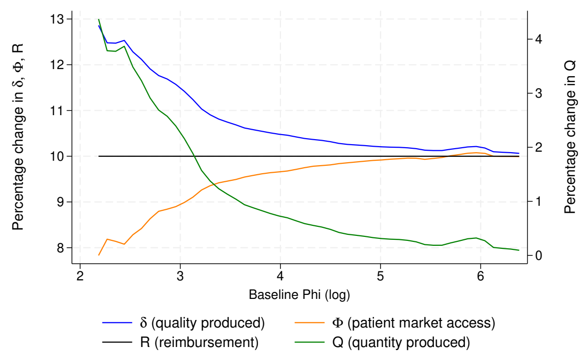
(a) Change (%) in output quality δ_i



(b) Change (%) in patient market access Φ_i



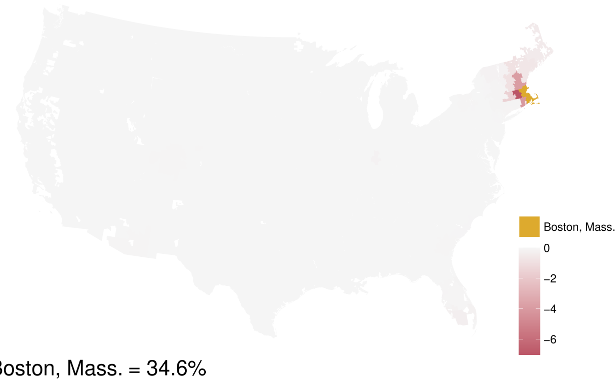
(c) Outcomes as a function of baseline patient market access Φ_i



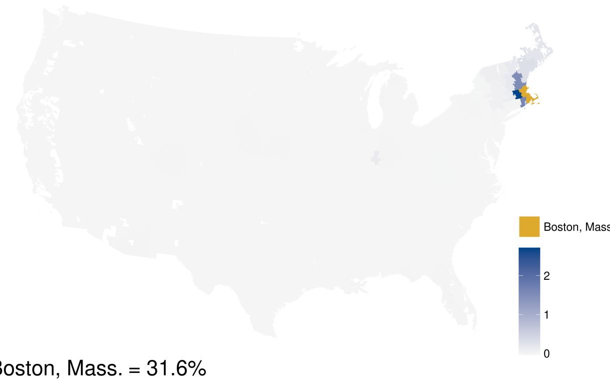
Notes: Panels (a) and (b) show the impacts of increasing reimbursements by 10% everywhere ($\hat{R}_i = 1.1 \ \forall i$) based on our estimated model. Panel (a) depicts the percentage change in quality of care δ_i provided in each region. Panel (b) depicts the percentage change in the value of market access Φ_i for patients who live in a region. Panel (c) shows local linear regressions of the percentage changes in δ_i, Φ_i, and Q_i against the region's initial patient market access, Φ_i. There is a negative relationship between the percentage changes in δ and Φ across regions. Patients who live in the regions with the largest quality increases in δ tend to have the lowest gains in patients' market access, Φ. The exercise is described in detail in Section 2.7.

Figure 2.10: Counterfactual outcomes for higher reimbursements in one region

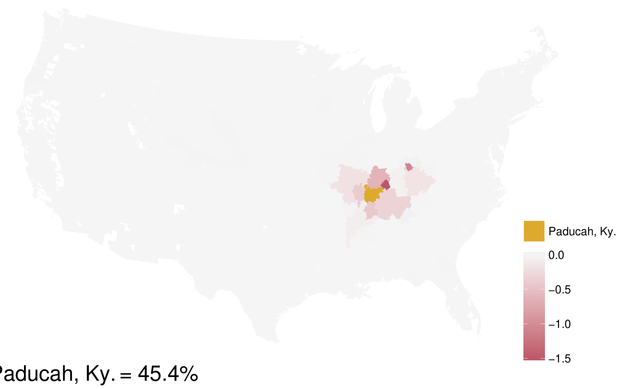
(a) Change (%) in output quality δ_i: higher reimbursement in Boston, Mass.



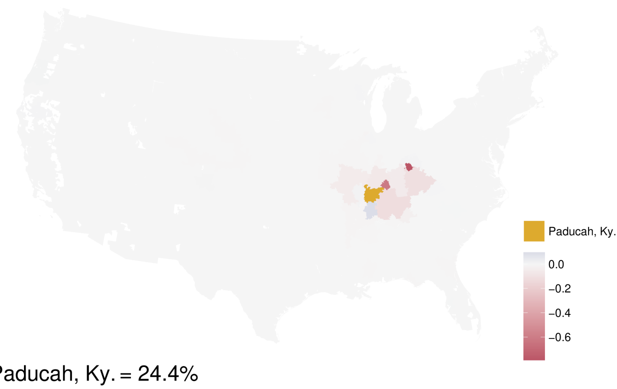
(b) Change (%) in market access Φ_i: higher reimbursement in Boston, Mass.



(c) Change (%) in output quality δ_i: higher reimbursement in Paducah, Ky.

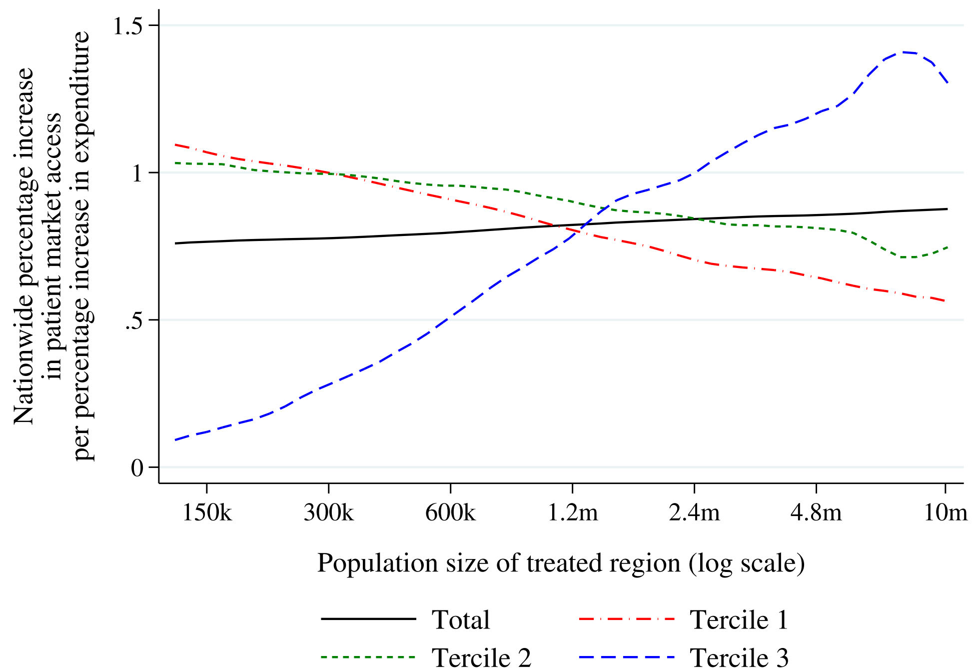


(d) Change (%) in market access Φ_i: higher reimbursement in Paducah, Ky.



Notes: Panels (a) and (b) show the impacts of increasing reimbursements by 30% in the Boston, Mass. HRR ($\hat{R}_i = 1.3$) based on our estimated model. Panel (a) illustrates the percentage change in quality of care δ_i provided in each region. Panel (b) illustrates the percentage change in the value of market access Φ_i for patients who live in an region. Panels (c) and (d) are analogous, but for a 30% increase in reimbursements in Paducah, Ky., a net importer. In all panels, the predicted change for the region whose reimbursement changes ("treated region") is listed on the map itself. In both cases, the quality produced in neighboring regions declines (Panels (a) and (c)). Patients in regions near Boston benefit from increased access to the treated region (Panel (b)), so there is a negative relationship between the percentage changes in δ and Φ across regions. In contrast, patients in regions near Paducah suffer a decrease in access (Panel (d)). The contrasting outcomes stem from Boston being a net exporter and Paducah being a net importer in the baseline equilibrium. The exercise is described in detail in Section 2.7.

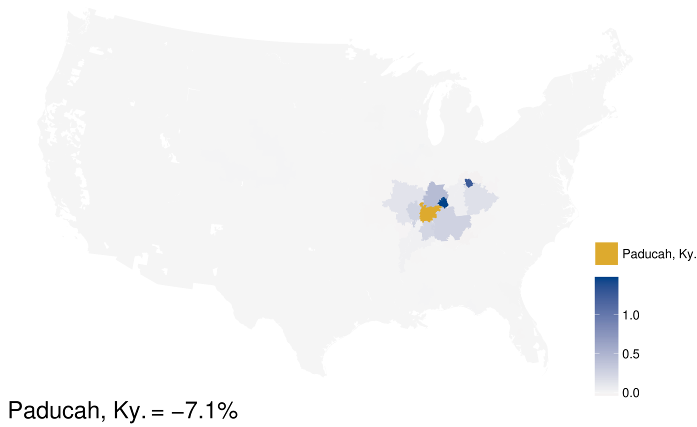
Figure 2.11: Changes in access $\hat{\Phi}_{j\kappa}$ by income when increasing reimbursements



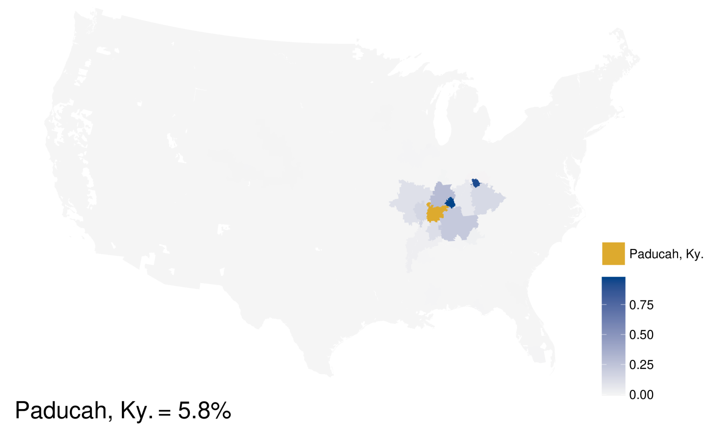
Notes: This figure summarizes the counterfactual outcomes of 30% higher reimbursements in one HRR as a function that HRR's population size. The nationwide return is the percentage increase in patient market access $\sum_{\kappa}\sum_{j} N_{j\kappa}\Phi_{j\kappa}$ per percentage increase in nationwide expenditures $\sum_{i} Q_i R_i$. The tercile-specific return is the increase in tercile-specific patient market access $\sum_{j} N_{j\kappa}\Phi_{j\kappa}$. Increasing reimbursements in more populous HRRs has the highest return when measured as impact on aggregate market access. Subsidies in less populous regions favor lower-income patients, primarily because there are more low-income patients living in and close to smaller regions.

Figure 2.12: Counterfactual outcomes when changing travel costs for Paducah, Ky. residents

(a) Change (%) in quality δ_i: reducing Paducah residents' travel costs by 30%



(b) Change (%) in access Φ_i: reducing Paducah residents' travel costs by 30%



Notes: Both panels show the impact of a 30% fall in travel costs for Paducah residents ($\hat{\rho}_{ij} = 1.3\ \forall i \neq$ Paducah). Panel (a) illustrates the percentage change in quality of care δ_i provided in each region. Panel (b) illustrates the percentage change in the value of market access Φ_i for patients who live in a region. The note shows the change for Paducah itself. Reduced travel costs for Paducah residents improves their market access but reduces the quality of care produced in Paducah itself. The increase in imports by Paducah residents causes service quality in neighboring regions to increase because of scale effects. This higher quality in turn attracts additional patients from the ring surrounding them, reducing quality slightly in that distant ring.

Table 2.1: Aggregate medical services exhibit a strong home-market effect

	(1)	(2)	(3)	(4)
Estimation method:	PPML	PPML	PPML	IV
Provider-market population (log)	0.635	0.642	0.644	0.597
	(0.0622)	(0.0605)	(0.0453)	(0.0730)
Patient-market population (log)	0.380	0.376	0.405	0.360
	(0.0605)	(0.0581)	(0.0421)	(0.0517)
Distance (log)	-1.654	0.124		0.106
	(0.0497)	(0.289)		(0.255)
Distance (log, squared)		-0.181		-0.179
		(0.0283)		(0.0250)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.017	0.011	0.002	0.017
Observations	93,636	93,636	93,636	93,636
Distance elasticity at mean		-2.46		-2.46
Distance deciles			Yes	

Notes: This table reports estimates of equation (2.8), which estimates the presence of weak or strong home-market effects. The sample is all HRR pairs ($N = 306^2$), and the dependent variable in all regressions is the value of trade. The independent variables are patient- and provider-market log population, log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments (GMM). Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.2: The home-market effect is stronger for rare procedures

	(1)	(2)	(3)	(4)	(5)	(6)
λ_X Provider-market population (log)	0.635	0.622	0.621		0.629	
	(0.0622)	(0.0601)	(0.0602)		(0.0592)	
λ_M Patient-market population (log)	0.380	0.381	0.382		0.379	
	(0.0605)	(0.0580)	(0.0581)		(0.0566)	
μ_X Provider-market population (log) × rare			0.302	0.291	0.316	0.288
			(0.0468)	(0.0453)	(0.0477)	(0.0455)
μ_M Patient-market population (log) × rare			-0.225	-0.220	-0.232	-0.212
			(0.0686)	(0.0669)	(0.0703)	(0.0657)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.017	0.019	0.020		0.014	
p-value for H_0: $\mu_X \leq \mu_M$			<0.001	<0.001	<0.001	<0.001
Observations	187,272	113,468	113,468	113,468	113,468	113,468
Distance controls	Yes	Yes	Yes	Yes		
Distance [quadratic] controls					Yes	Yes
Patient-provider-market-pair FEs				Yes		Yes

Notes: This table reports estimates of equation (2.11), which introduces interactions with an indicator for whether a procedure is "rare" (provided less often than the median procedure, when adding up all procedures provided nationally). The interactions with patient- and provider-market population reveal whether the home-market effect is larger for rare procedures. The unit of observation is {rare indicator, exporting HRR, importing HRR} so the number of observations is 2×306^2 in column 1, and the dependent variable in all regressions is the value of trade. Columns 2 onwards drop HRR pairs with zero trade in both procedure groups, and column 2 shows that this restriction has a negligible impact on the estimated log population coefficients. Columns 3 onwards include the rare indicator interacted with patient- and provider-market populations and distance covariates. Columns 1–4 control for distance using the log of distance between HRRs. Columns 5 and 6 add a control for the square of log distance. Columns 4 and 6 introduce a fixed effect for each ij pair of patient market and provider market, so these omit all covariates that are not interacted with the rare indicator. The positive coefficient on provider-market population × rare across all columns indicates that the home-market effect is stronger for rare than for common services. The negative coefficient on patient-market population × rare across all columns indicates that the *strong* home-market effect has a larger magnitude for rare services. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.3: The stronger home-market effect for rare procedures is robust to instrumenting for population

	(1)	(2)	(3)	(4)	(5)	(6)
Geography:	HRR	HRR	CBSA	CBSA	CBSA	CBSA
Instrument:	1940 pop	1940 pop	1940 pop	1940 pop	Bedrock	Bedrock
Procedure Sample:	Common	Rare	Common	Rare	Common	Rare
Provider-market population (log)	0.595	1.081	0.716	0.895	1.157	1.753
	(0.0731)	(0.0914)	(0.0249)	(0.0388)	(0.307)	(0.524)
Patient-market population (log)	0.362	0.0477	0.396	0.328	0.182	-0.582
	(0.0518)	(0.115)	(0.0261)	(0.0344)	(0.373)	(0.580)
Distance (log)	0.102	0.992	-3.412	-1.378	-4.678	-4.631
	(0.255)	(0.442)	(0.294)	(0.989)	(1.049)	(2.520)
Distance (log, squared)	-0.179	-0.263	0.105	-0.0742	0.210	0.181
	(0.0251)	(0.0497)	(0.0287)	(0.0935)	(0.0845)	(0.199)
Observations	93,636	93,636	857,476	857,476	781,456	781,456
Distance elasticity at mean	-2.46	-2.77	-1.91	-2.43	-1.68	-2.05

Notes: This table reports estimates of equation (2.8), when separating procedures into those above- and below-median frequency and instrumenting for log population. The dependent variable in all regressions is the value of trade. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. We report coefficients on provider market population, patient market population, log distance, and log distance squared. Every specification also includes a same-market ($i = j$) indicator variable. The odd-numbered columns are trade in above-median-frequency procedures; the even-numbered columns are trade in below-median-frequency procedures. In columns 1 and 2, the sample is all HRR pairs ($N = 306^2$). In columns 3 and 4, the sample is all CBSA pairs ($N = 926^2$). In columns 5 and 6, the sample is all CBSA pairs for which the bedrock-depth instrumental variable is available ($N = 844^2$). We use 1940 population counts to produce two instrumental variables: 1940 population in the patient market and 1940 population in the provider market are instruments for log population in the patient market and log population in the provider market, respectively. Similarly, we use bedrock depth to produce two instrumental variables for CBSAs. Both the strong home-market effect and its larger magnitude for rare procedures are robust to instrumenting for population, estimating by GMM. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.4: The home-market effect is stronger for rarer diagnoses

	(1)	(2)	(3)	(4)	(5)	(6)
λ_X Provider-market population (log)	0.635	0.622	0.616		0.624	
	(0.0625)	(0.0604)	(0.0588)		(0.0578)	
λ_M Patient-market population (log)	0.382	0.383	0.386		0.383	
	(0.0606)	(0.0580)	(0.0569)		(0.0555)	
μ_X Provider-market population (log) $\times$ rare			0.0719	0.0687	0.0763	0.0683
			(0.0547)	(0.0519)	(0.0561)	(0.0506)
μ_M Patient-market population (log) $\times$ rare			-0.0422	-0.0409	-0.0429	-0.0380
			(0.0419)	(0.0403)	(0.0440)	(0.0395)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.018	0.020	0.021		0.015	
p-value for H_0: $\mu_X \leq \mu_M$			0.114	0.113	0.113	0.115
Observations	187,272	112,626	112,626	112,626	112,626	112,626
Distance controls	Yes	Yes	Yes	Yes		
Distance [quadratic] controls					Yes	Yes
Patient-provider-market-pair FEs				Yes		Yes

Notes: This table augments equation (2.8) by adding interactions with an indicator for whether a diagnosis is "rare" (provided less often than the median diagnosis, when adding up all patients receiving the diagnosis nationally) or "common" (more often than median). The interactions with patient- and provider-market population reveal whether the home-market effect is larger for rare diagnoses. The unit of observation is {rare indicator, exporting HRR, importing HRR} so the number of observations is 2×306^2 in column 1, and the dependent variable in all regressions is the value of trade. Valid primary diagnoses observed in 1,000 distinct claims or more nationally in the professional fees 20% sample are included. Columns 2 onwards drop HRR pairs with zero trade, and column 2 shows that this restriction has a negligible impact on the estimated log population coefficients. Columns 1–4 control for distance using the log of distance between HRRs. Columns 5 and 6 add a control for the square of log distance. Columns 4 and 6 introduce a fixed effect for each ij pair of patient market and provider market, so these omit the patient- and provider-market population covariates. The positive coefficient on provider-market population $\times$ rare across all columns indicates that the home-market effect is stronger for rare than for common diagnoses. The negative coefficient on patient-market population $\times$ rare across all columns indicates that the *strong* home-market effect is especially true for rare diagnoses. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.5: Scale elasticity estimates

Panel A: All services	Baseline	No Diagonal	Controls
OLS	0.776	0.803	0.810
	(0.031)	(0.045)	(0.041)
2SLS: population (log)	0.712	0.798	0.724
	(0.031)	(0.050)	(0.039)
2SLS: population (1940, log)	0.533	0.663	0.545
	(0.072)	(0.098)	(0.067)
Panel B: Rare services			
OLS	0.947	1.083	0.927
	(0.030)	(0.045)	(0.035)
2SLS: population (log)	0.912	1.026	0.881
	(0.037)	(0.049)	(0.049)
2SLS: population (1940, log)	0.835	0.951	0.789
	(0.061)	(0.084)	(0.070)

Notes: This table reports estimates of $\hat{\alpha}$ from ordinary least squares (OLS) or two-stage least squares (2SLS) regressions of the form $\widehat{\ln \delta_i} = \alpha \ln Q_i + \ln R_i + u_i$, where $\widehat{\ln \delta_i}$ is estimated in equation (2.12), Q_i is region i's total production of non-emergency-room physician services for Medicare beneficiaries, R_i is Medicare's Geographic Adjustment Factor, and u_i is an error term. In the rows labeled "2SLS" we instrument for $\ln Q_i$ using the specified instruments. The $\ln R_i$ control is omitted in the columns labeled "no controls". Appendix Table 2.17 reports analogous estimates at the CBSA level, which allows us to also control for input costs (as input cost data are more reliable for CBSAs than for HRRs). In the columns labeled "no diag", Q_{ii} observations were omitted when estimating $\widehat{\ln \delta_i}$ in equation (2.12). Standard errors are robust to heteroskedasticity. Across all of the permutations of our method, we estimate substantial scale economies.

Table 2.6: Regression of $\hat{\Phi}_{j\mathfrak{k}}$ on tercile dummies and trade shares

	Nationwide Reimbursement Increase			Boston Reimbursement Increase		
	(1)	(2)	(3)	(4)	(5)	(6)
Income tercile = 2	1.080	1.085	-0.143	2.00e-05	0.00298	-0.00494
	(0.0554)	(0.0555)	(0.0318)	(0.00938)	(0.00966)	(0.00214)
Income tercile = 3	1.568	1.553	-0.240	0.0697	0.0649	-0.00732
	(0.0712)	(0.0698)	(0.0549)	(0.0349)	(0.0347)	(0.00266)
Imported share $(1 - m_{0j\kappa} - m_{jj\kappa})$			-0.519			
			(0.129)			
$m_{0j\kappa}$			-12.65			
			(0.422)			
$m_{\text{Boston},j\kappa}$						36.95
						(0.435)
Constant	8.763	8.767	11.10	0.0984	0.0989	-0.0691
	(0.0594)	(0.0403)	(0.0769)	(0.105)	(0.0127)	(0.00265)
Observations	885	885	885	885	885	885
R-squared	0.498	0.675	0.988	0.000	0.980	1.000
HRR fixed effects	No	Yes	Yes	No	Yes	Yes

Notes: This table uses linear regressions to summarize how market access changes across HRRs j and income terciles κ in response to two different counterfactual policies. The dependent variable in all columns is the percentage change in market access, $100 \times (\hat{\Phi}_{j\mathfrak{k}} - 1)$. Standard errors (in parentheses) are clustered by market. Columns 1, 2, and 3 consider a 10% reimbursement increase nationwide. Columns 4, 5, and 6 consider a 30% reimbursement increase in Boston only. The constant in the first regression reports the percentage change for the lowest income terciles, and the coefficients on the other terciles are the additional percentage point gain for those terciles relative to the lowest. Other controls include the outside option market share $m_{0j\kappa}$, imported share $1 - m_{0j\kappa} - m_{jj\kappa}$ (where $m_{jj\kappa}$ is local production), and Boston's market share $m_{\text{Boston},j\kappa}$. The coefficients are much smaller in columns 4, 5, and 6 because only Boston is treated, so most of the country is hardly affected. When we add market share controls, the coefficients indicating tercile differences become much smaller, indicating that baseline trade patterns drive the distributional impacts.

APPENDIX

2.10 Theory appendix

2.10.1 Monopolistic competition with one firm per region

Suppose that there is a single firm in each region that offers fixed-price services to patients under monopolistic competition with the firms in other regions. Assume $K(\delta_i) = \delta_i$ and $H(Q_i) = Q_i^{\alpha}$. The profit-maximizing choice of quality δ_i by the firm in region i is

$$\max_{\delta_i} \pi_i = \left(\bar{R} - \frac{w_i \delta_i}{A_i Q_i^{\alpha}}\right) Q_i \quad \text{where } Q_i = \sum_j Q_{ij} = \delta_i \sum_j \frac{N_j}{\Phi_j} \rho_{ij}$$

$$\frac{\partial \pi_i}{\partial \delta_i} = 0 \implies \frac{\bar{R}}{2-\alpha} = \frac{w_i \delta_i}{A_i Q_i^{\alpha}} = C(Q_i, \delta_i; w_i, A_i)$$

This expression replaces the free-entry condition (2.4) in the definition of equilibrium. Changing the value of the constant on the left side of this equality does not change any of the subsequent theoretical predictions. In this respect, the monopolistic-competition model with one firm per region is isomorphic to the perfect-competition model with external economies of scale.

2.10.2 Model with multiple types of patients

This section extends the model to feature multiple types of patients who face different trade costs. There is a finite set of patient types, which are indexed by κ. A patient type is defined by the trade costs $\rho_{ij(k)} = \rho_{ij}^{\kappa} \forall k \in \kappa$. Qualities δ_i, including the outside option δ_0, are the same for all patient types. The demand by patients of type κ residing in location j for procedures performed by providers in location i is now given by

$$Q_{ij}^{\kappa} = \frac{\delta_i N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa}.$$

The aggregate gravity equation is the sum of type-specific gravity equations:

$$Q_{ij} = \sum_{\kappa} Q_{ij}^{\kappa} = \delta_i \sum_{\kappa} \frac{N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa}. \tag{2.15}$$

The free-entry condition (2.4) remains unchanged with the introduction of multiple patient types:

$$R_i = \frac{w_i \delta_i}{A_i Q_i^{\alpha}}.$$

In equilibrium, market clearing requires that

$$Q_i = \left(\frac{w_i \delta_i}{A_i R_i}\right)^{1/\alpha} = \delta_i \sum_j \sum_{\kappa} \frac{N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa} \implies \delta_i = \left(\frac{A_i R_i}{w_i}\right)^{1/(1-\alpha)} \left(\sum_j \sum_{\kappa} \frac{N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa}\right)^{\alpha/(1-\alpha)}.$$

2.10.3 Derivations of results in Section 2.2.5

Abusing notation so that $\mathcal{I}$ is both the set and number of regions, equations (2.2) and (2.3) together constitute $2\mathcal{I}$ equations with $2\mathcal{I}$ unknowns. For the special case of $H(Q_i) = Q_i^{\alpha}$ and $K(\delta_i) = \delta_i$, this reduces to the following $\mathcal{I}$ equations with the unknowns $\{\delta_i\}_{i=1}^{\mathcal{I}}$:

$$\delta_i = \left(\frac{\overline{R} A_i}{w_i}\right)^{\frac{1}{1-\alpha}} \left(\sum_{j \in \mathcal{I}} \frac{\rho_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}} N_j\right)^{\frac{\alpha}{1-\alpha}}$$

Following Costinot et al. (2019), we examine the home-market effect in the neighborhood of a symmetric equilibrium. For brevity, assume $\frac{\overline{R} A_i}{w_i} = 1 \ \forall i$. Note that at the symmetric equilibrium:

$$\bar{\delta}^{\frac{1-\alpha}{\alpha}} = \frac{1}{1 + \bar{\delta} + \sum_{i' \neq i} \bar{\delta}\rho} \bar{N} + \sum_{j \neq i} \frac{\rho}{1 + \bar{\delta} + \sum_{i' \neq j} \bar{\delta}\rho} \bar{N} = \frac{1 + (\mathcal{I} - 1)\rho}{\bar{\Phi}} \bar{N} = \frac{\bar{\Phi} - 1}{\bar{\Phi}} \frac{\bar{N}}{\bar{\delta}}. \tag{2.16}$$

Given $\alpha > 0$, totally differentiating the above system of equations in terms of $\{\mathrm{d}\delta_i, \mathrm{d}N_i\}_{i=1}^{\mathcal{I}}$

and evaluating it at the symmetric equilibrium yields the following expression:

$$\frac{\bar{\Phi}^2}{\bar{N}}\frac{1-\alpha}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}}\mathrm{d}\delta_i = -\left[\mathrm{d}\delta_i + \rho\sum_{i'\neq i}\mathrm{d}\delta_{i'}\right] + \bar{\Phi}\frac{\mathrm{d}N_i}{\bar{N}} + \sum_{j\neq i} -\rho\left[\mathrm{d}\delta_j + \rho\sum_{i'\neq j}\mathrm{d}\delta_{i'}\right] + \sum_{j\neq i}\rho\bar{\Phi}\frac{\mathrm{d}N_j}{\bar{N}}.$$

Given $\mathrm{d}N_1 > 0$ and $\mathrm{d}N_j = 0\ \forall j \neq 1$, we obtain the following expression for $\mathrm{d}\ln\delta_1$:

$$\mathrm{d}\ln\delta_1 = \frac{\frac{\bar{\Phi}}{\bar{\delta}}\mathrm{d}\ln N_1 - (\mathcal{I}-1)(2\rho + ((\mathcal{I}-2)\rho^2))\mathrm{d}\ln\delta_{j\neq 1}}{\frac{\Phi^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} + 1 + (\mathcal{I}-1)\rho^2}. \tag{2.17}$$

Further tedious algebra delivers the following expression for quality changes:

$$\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1} = \frac{(1-\rho)}{\frac{\Phi^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} + (1-\rho)^2}\frac{\bar{\Phi}}{\bar{\delta}}\mathrm{d}\ln N_1. \tag{2.18}$$

Equation (2.16) implies that $\frac{\Phi^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} = \left(\frac{1-\alpha}{\alpha}\right)\frac{\Phi(\Phi-1)}{\bar{\delta}}$ and therefore

$$\begin{aligned}\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1} &= \frac{(1-\rho)}{\left(\frac{1-\alpha}{\alpha}\right)\frac{\Phi(\Phi-1)}{\bar{\delta}} + (1-\rho)^2}\frac{\bar{\Phi}}{\bar{\delta}}\mathrm{d}\ln N_1 \\ &= \left[\frac{1-\alpha}{\alpha}\frac{(\bar{\Phi}-1)}{(1-\rho)\bar{\delta}} + \frac{(1-\rho)\bar{\delta}}{\bar{\Phi}}\right]^{-1}\mathrm{d}\ln N_1 > 0.\end{aligned}$$

The last expression above is reported in Section 2.2.5.

Prior to deriving the weak and strong home-market effects, we obtain an expression for $\frac{\mathrm{d}\ln\delta_j}{\mathrm{d}\ln N_1}$ for $j \neq 1$ around the symmetric equilibrium. Define $\bar{\mathcal{Q}} \equiv \frac{\Phi^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} > 0$. Combining the expressions for $\mathrm{d}\ln\delta_1$ from equation (2.17) and for $\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1}$ from equation (2.18) yields the following:

$$\frac{\mathrm{d}\ln\delta_{j\neq 1}}{\mathrm{d}\ln N_1} = \frac{\bar{\Phi}}{\bar{\delta}}\frac{\bar{\mathcal{Q}}\rho + \rho^3(\mathcal{I}-1) - \rho^2(\mathcal{I}-2) - \rho}{(\bar{\mathcal{Q}} + (1-\rho)^2)(\bar{\mathcal{Q}} + 1 + \rho^2 + 2\rho(\mathcal{I}-1) + \mathcal{I}\rho^2(\mathcal{I}-2))}$$

The weak home-market effect is derived as follows:

$$
\begin{aligned}
\ln Q_{1,j\neq 1} &= \alpha \ln Q_1 + \ln \rho - \ln \Phi_j + \ln N_j \\
\frac{\mathrm{d}\ln Q_{1,j\neq 1}}{\mathrm{d}\ln N_1} &= \alpha \frac{\mathrm{d}\ln Q_1}{\mathrm{d}\ln N_1} - \frac{\alpha}{\Phi_j}\left(\rho Q_1^{\alpha-1}\frac{\mathrm{d}Q_1}{\mathrm{d}\ln N_1} + Q_j^{\alpha-1}\frac{\mathrm{d}Q_j}{\mathrm{d}\ln N_1} + \rho \sum_{i'\neq 1,j} Q_{i'}^{\alpha-1}\frac{\mathrm{d}Q_{i'}}{\mathrm{d}\ln N_1}\right) \\
&= \frac{\mathrm{d}\ln \delta_1}{\mathrm{d}\ln N_1} - \frac{1}{\Phi_j}\left(\rho\delta_1 \frac{\mathrm{d}\ln \delta_1}{\mathrm{d}\ln N_1} + \delta_j \frac{\mathrm{d}\ln \delta_j}{\mathrm{d}\ln N_1} + \rho \sum_{i'\neq 1,j} \delta_{i'}\frac{\mathrm{d}\ln \delta_{i'}}{\mathrm{d}\ln N_1}\right) \\
&= \left(\frac{\bar{N} - Q_{1j}}{\bar{N}}\right)\frac{\mathrm{d}\ln \delta_1}{\mathrm{d}\ln N_1} - \left(\frac{\bar{N} - Q_{0j} - Q_{1j}}{\bar{N}}\right)\frac{\mathrm{d}\ln \delta_j}{\mathrm{d}\ln N_1} \\
&= \left(\frac{\bar{N} - Q_{1j}}{\bar{N}}\right)\left[\frac{\mathrm{d}\ln \delta_1}{\mathrm{d}\ln N_1} - \frac{\mathrm{d}\ln \delta_j}{\mathrm{d}\ln N_1}\right] + \frac{Q_{0j}}{\bar{N}}\frac{\mathrm{d}\ln \delta_j}{\mathrm{d}\ln N_1} \\
&= \frac{\Phi}{\bar{\delta}\bar{N}}\frac{1}{\bar{\mathcal{Q}} + (1-\rho)^2}\Bigg[(Q_{jj} + (\mathcal{I}-2)Q_{1j})(1-\rho) \\
&+ \frac{Q_{0j}}{\bar{\mathcal{Q}} + 1 + \rho^2 + 2\rho(\mathcal{I}-1) + \mathcal{I}\rho^2(\mathcal{I}-2)} \\
&\quad \times \left\{\bar{\mathcal{Q}} + (\rho-1)^2 + 2(\mathcal{I}-1)(\rho-\rho^2) + (\mathcal{I}-1)(\mathcal{I}-2)[\rho^2-\rho^3]\right\}\Bigg] \\
&> 0.
\end{aligned}
$$

The condition for the strong home-market effect is derived as follows:

$$
\begin{aligned}
Q_{1,j\neq 1} - Q_{j\neq 1,1} &= \frac{Q_1^\alpha \rho}{1 + Q_1^\alpha\rho + Q_j^\alpha + \sum_{i\neq 1,j} Q_i^\alpha \rho}N_j \\
&- \frac{Q_j^\alpha \rho}{1 + Q_1^\alpha + Q_j^\alpha\rho + \sum_{i\neq 1,j} Q_i^\alpha \rho}N_1 \\
\mathrm{d}\ln Q_{1,j\neq 1} - \mathrm{d}\ln Q_{j\neq 1,1} &= \mathrm{d}\ln N_j - \mathrm{d}\ln N_1 + \alpha\left[1 + (1-\rho)\frac{\bar{Q}^\alpha}{\bar{\Phi}}\right]\left(\mathrm{d}\ln Q_1 - \mathrm{d}\ln Q_j\right) \\
&= -\mathrm{d}\ln N_1 + \left[1 + (1-\rho)\frac{\bar{\delta}}{\bar{\Phi}}\right]\left(\mathrm{d}\ln \delta_1 - \mathrm{d}\ln \delta_j\right) \\
&= \left[\frac{1 - \frac{1-\alpha}{\alpha}\frac{1+(\mathcal{I}-1)\rho}{1-\rho}}{\frac{1-\alpha}{\alpha}\frac{(1+(\mathcal{I}-1)\rho)}{(1-\rho)} + \frac{(1-\rho)\bar{\delta}}{1+(1+(\mathcal{I}-1)\rho)\bar{\delta}}}\right]\mathrm{d}\ln N_1.
\end{aligned}
$$

There is a strong home-market effect in the neighborhood of the symmetric equilibrium if and only if $\mathrm{d}\ln Q_{1,j\neq 1} - \mathrm{d}\ln Q_{j\neq 1,1} > 0$.

$$\left[\frac{1 - \frac{1-\alpha}{\alpha}\frac{1+(\mathcal{I}-1)\rho}{1-\rho}}{\frac{1-\alpha}{\alpha}\frac{(1+(\mathcal{I}-1)\rho)}{(1-\rho)} + \frac{(1-\rho)\bar{\bar{\delta}}}{1+(1+(\mathcal{I}-1)\rho)\bar{\bar{\delta}}}}\right]\mathrm{d}\ln N_1 > 0 \iff \frac{\alpha}{1-\alpha} > \frac{1+(\mathcal{I}-1)\rho}{1-\rho}$$

This is true if α is large enough and ρ is small enough.

Our difference-in-differences prediction concerns how the effect of market size on net exports varies with the number of potential patients $\bar{N}$. Given the scale elasticity α and (inverse) trade costs ρ, the denominator of the right side of equation (2.6) is increasing in the symmetric-equilibrium quality $\bar{\bar{\delta}}$. For two procedures that both exhibit a strong home-market effect because they have the same scale elasticity and trade costs, the effect of population size on net exports will be larger for the procedure with lower service quality. The symmetric-equilibrium service quality is increasing in the number of potential patients $\bar{N}$ because there are increasing returns (see equation (2.16)). Thus, in the neighborhood of the symmetric equilibrium, the strength of a strong home-market effect is decreasing in the number of potential patients.

2.11 Data appendix

2.11.1 Procedure frequency in main sample compared with aggregate and private data

Medicare provides two public-use files based on 100 percent claims. The first one contains the complete count of procedures billed by HCPCS code but does not have information about providers. We use it to confirm that procedure counts based on the confidential data do not suffer substantial sampling bias. In Figure 2.21, we split procedure codes into deciles based on their national frequencies, separately in the confidential and public datasets. This generates a 100-cell matrix by decile pair. We plot the share of procedures in each cell in this matrix to determine how well the two datasets align. The vast majority of the codes are on the diagonal, with almost all of the remainder adjacent to the diagonal. This suggests that sampling error is not causing us to mischaracterize procedure frequency.

Medicare provides a second public file at the level of physician-by-procedure (HCPCS code). This summary does not contain any patient-level information so cannot be used to study trade flows, but we can use it to replicate analyses based on the location of production and physician experience. This file is censored such that physician-by-procedure pairs with 10 or fewer observations per year are suppressed, which makes for a more complicated bias than simple 20 percent random sampling. Nevertheless, all of the results that can be tested on this sample confirm those found in the 20 percent sample.

Since our procedure frequency measures rely on Medicare data, we would mismeasure frequency if the Medicare population uses a substantially different composition of care from the broader population. For example, childbirth is less common among Medicare beneficiaries. So our frequency measures may not capture the true national frequency of a procedure.

We address this by comparing procedure frequencies between the Medicare public data and private data from the Health Care Cost Institute (HCCI). The HCCI data contain claims for about 55 million privately insured patients (about 35% of individuals with employer-based insurance). We only consider HCPCS codes performed on at least eleven patients in the HCCI data. Note that frequencies are computed for all providers here, not only MDs and DOs. The authors acknowledge the assistance of the Health Care Cost Institute (HCCI) and its data contributors, Aetna, Humana, and Blue Health Intelligence, in providing the claims data analyzed in this section.

We examine whether procedures classified as above median frequency in one dataset are above median frequency in the other dataset. Table 2.18 shows that 88% of the services above median frequency in Medicare are also as above median frequency in the HCCI data. Similarly, 82% of the services below median frequency in Medicare are also below median frequency in the HCCI data.

We next compare classifications of procedures' frequency deciles in Figure 2.22. Analogous to Figure 2.21, this plot visualizes the share of procedures which fall into each of pair of frequency decile bins in HCCI and Medicare data. The two classifications appear to coincide relatively well, with slightly stronger agreement for very frequent procedures compared to rarer procedures in the Medicare public-use data. Overall, the frequency classifications of procedures coincide well between Medicare public-use data and HCCI data.

2.11.2 Additional details on data sources

Physician earnings. The Gottlieb et al. (2020) earnings data depicted in Appendix Figure 2.14 are only available for 111 commuting zones. The American Community Survey (ACS) covers far more CBSAs, but this source top-codes income for a substantial share of doctors.

U.S. News and World Report. The publication produces an overall ranking and rankings for 12 particular specialties. We count the number of times each HRR's hospitals appear on any of these 13 lists.[54] Thus, higher ranking on the horizontal axis indicates a region has some combination of more top-ranked hospitals, or each of its hospitals performs well in many specialty areas.

2.11.3 Geographic price adjustments

Professional fees. To adjust for geographic price variation in the professional fees, we compute a national average price per Healthcare Common Procedure Coding System (HCPCS) code as the sum of the line allowed amount, which includes the line item's Medicare-paid and beneficiary-paid amounts (i.e., deductible, copayment, and coinsurance), divided by the

54. Results are similar when we use other methods to aggregate the rankings information, including when we account for the ordered nature of the lists.

sum of the line service count per HCPCS code nationally. We then apply this average price to all billing for the HCPCS code when computing total spending across services.

Hospital inpatient fees. We use the field "final standard payment amount" in the MedPAR file, which is computed as described in Finkelstein, Gentzkow, and Williams (2016) and Gottlieb et al. (2010). This represents "a standard Medicare payment amount, without the geographical payment adjustments and some of the other add-on payments that go to the hospitals" according to the data documentation.

Hospital outpatient fees. To adjust for geographic price variation in hospital outpatient fees, we compute a national average price per Healthcare Common Procedure Coding System (HCPCS) code, Ambulatory Payment Classifications (APC) code, and revenue center code. HCPCS codes reflect the procedure performed and APC codes reflect a prospective payment system applicable to outpatient analogous to Diagnosis Related Groups (DRGs) for inpatient claims. Revenue center contains information on the place of service, e.g. rehabilitation or acute care, so we consider two procedures performed in different revenue centers as different procedures for price adjustment purposes.

The total amount per claim line is calculated as the sum of the claim (Medicare) payment amount, the primary payer amount, the Part B beneficiary co-insurance amount, the beneficiary Part B deductible amount, and the beneficiary blood deductible amount. These amounts are summed nationally for each {HCPCS code, APC code, revenue center code} triplet, and divided by the frequency of that triplet to obtain a national average price. We then apply this average price to all instances of that {HCPCS code, APC code, revenue center code} combination when computing total spending across services.

2.11.4 Residential measurement error

This appendix uses two methods to investigate potential measurement error in patients' residential location. The first source of potential error is "snowbird" patients, who have multiple residences and therefore may appear to travel farther than they actually do. They may need medical care while spending months in a warmer HRR that is not the one listed as their main residence (or vice versa). Our results are robust to two methods of removing potential snowbirds: excluding Arizona, California, and Florida, following Finkelstein, Gentzkow, and Williams (2016), and excluding the 10% of HRRs with the highest share of second homes in American Community Survey data. These results are in Tables 2.10 and 2.11. The results are little changed by these sample restrictions.

We test for more general location measurement error by examining how far patients appear to travel for dialysis. Since Medicare patients requiring dialysis must generally visit a dialysis center thrice weekly, they are unlikely to go substantial distances for this service. Table 2.12 compares travel distances for dialysis with other care. Dialysis patients appear to travel less than one-quarter as often as other patients—and even less when excluding snowbird states—suggesting that our residential location assignment is largely accurate.

2.11.5 Scale elasticity estimation with unobserved market segments

Our data only contain procedure-level production and consumption in Traditional Medicare (TM), not for Medicare Advantage (MA) or non-Medicare (NM) patients. We quantify how this biases our estimate of the scale elasticity, α, based on geographic variation. Suppose the production function is

$$\ln \delta_i = \alpha \ln Q_i + u_i,$$

where $Q_i = Q_i^{\text{TM}} + Q_i^{\text{MA}} + Q_i^{\text{NM}}$ is the total quantity produced in region i, of which we only observe Q_i^{TM}. When we estimate the scale elasticity α using Q_i^{TM} as a proxy for Q_i, our regression coefficient may be biased:

$$\frac{\text{Cov}(\ln \delta_i, \ln Q_i^{\text{TM}})}{\text{Var}(\ln Q_i^{\text{TM}})} = \frac{\text{Cov}(\alpha \ln Q_i, \ln Q_i^{\text{TM}})}{\text{Var}(\ln Q_i^{\text{TM}})} + \frac{\text{Cov}(u_i, \ln Q_i^{\text{TM}})}{\text{Var}(\ln Q_i^{\text{TM}})} = \alpha\zeta,$$

where ζ, which governs the bias, is the regression coefficient from $\ln Q_i = \zeta \ln Q_i^{\text{TM}} + \mathfrak{u}_i$.

To compute ζ we differentiate the identity $Q_i = Q_i^{\text{TM}} \left(1 + \frac{Q_i^{\text{MA}}}{Q_i^{\text{TM}}} + \frac{Q_i^{\text{NM}}}{Q_i^{\text{TM}}}\right)$ with respect to Q_i^{TM}, which we observe:

$$\frac{\mathrm{d} \ln Q_i}{\mathrm{d} \ln Q_i^{\text{TM}}} = 1 + s_i^{\text{MA}} \varrho_i^{\text{MA}} + s_i^{\text{NM}} \varrho_i^{\text{NM}},$$

where $s_i^{\text{MA}} \equiv \frac{Q_i^{\text{MA}}}{Q_i^{\text{TM}} + Q_i^{\text{MA}} + Q_i^{\text{NM}}}$ is the MA share of production in region i, $\varrho_i^{\text{MA}} \equiv \frac{\mathrm{d} \ln \frac{Q_i^{\text{MA}}}{Q_i^{\text{TM}}}}{\mathrm{d} \ln Q_i^{\text{TM}}}$ is the TM production elasticity of relative production, and s_i^{NM} and ϱ_i^{NM} are similarly defined for non-Medicare (NM) insurance. To make it feasible to estimate these elasticities, we assume that they are constant across regions. If relative quantities produced are uncorrelated with the Traditional Medicare quantity produced ($\varrho^{\text{MA}} = \varrho^{\text{NM}} = 0$), then $\zeta = 1$ and $\alpha\zeta$ is an unbiased estimate of the scale elasticity α.[55] Otherwise, we need estimates of the average production shares $\bar{s}^{\text{MA}}$ and $\bar{s}^{\text{NM}}$ and the regression coefficients ϱ^{MA} and ϱ^{NM} to compute ζ.

We compute the production shares using data on aggregate expenditures and price deflators from prior research. Medicare, including both TM and MA, paid for \$153 billion of the \$525 billion spent nationally on physician services in 2017 (Centers for Medicare and Medicaid Services, 2022). Per capita spending and prices are similar between the two parts of Medicare (Berenson et al., 2015; Gupta, Navathe, and Schwartz, 2022). Given this similarity, we apportion Medicare's production between TM and MA based on relative enrollment and obtain $\bar{s}^{\text{MA}} = 0.111$. Next we consider Non-Medicare (NM) production. Private insurance

55. A special case would be if the quantity of care produced outside of TM is perfectly correlated with volume inside TM, so the shares s_i^{MA} and s_i^{NM} are constant.

spent \$226 billion, which we deflate by a factor of 1.43 to account for the higher prices private insurance pays to make quantities comparable to Medicare (Lopez and Jacobson, 2020). Medicaid spent roughly \$41 billion, which we deflate by its relative price of 0.72 (Zuckerman, Skopec, and Aarons, 2021). We incorporate other residual categories of production without price adjustments.[56] Combining these, we obtain an average $\bar{s}^{\mathrm{NM}} = 0.676$.

To estimate ϱ^{MA} and ϱ^{NM}, we assume that relative production is proportionate to relative resident beneficiaries. We obtain the number of TM beneficiaries and number of MA beneficiaries by HRR from Medicare enrollment data and compute the number of NM patients as total population minus Medicare enrollees.[57] Regressing the respective beneficiary ratios on log TM production yields $\hat{\varrho}^{\mathrm{MA}} = 0.073$ and $\hat{\varrho}^{\mathrm{NM}} = 0.069$. Putting these together means $\hat{\zeta} = 1.055$, so our estimated $\alpha\zeta = 0.66$ from Table 2.5 implies a scale elasticity of $\alpha = \frac{0.66}{1.055} = 0.63$.

2.12 Details of counterfactual calculations

Section 2.12.1 describes how we compute counterfactual equilibrium outcomes relative to baseline equilibrium outcomes in the model. Section 2.12.2 describes the assumptions we make to infer the number of potential patients N_j and hence import shares m_{ij}, which are inputs into these calculations. Section 2.12.3 describes how to compute counterfactual outcomes in the model when there are multiple (observed) types of patients who differ in their trade costs. Section 2.12.4 describes how we infer the number of potential patients of

56. These other categories in the National Health Expenditure data are labeled Other Health Insurance Programs and Other Third Party Payers, along with out-of-pocket spending. Our simplifying approach here amounts to assuming Medicare prices for these residual categories.

57. Ideally we would like to use the quantity of production in NM and MA markets, but we do not have this available at the HRR level. Beneficiaries might seem like a problematic proxy because the composition of NM beneficiaries varies widely across space, with some regions having a high Medicaid share and others a high private share. In aggregate, these two markets turn out to have similar per capita quantities of physician service spending: while private spending is \$1,118 per capita and Medicaid spending is \$550 per capita, the price adjustments mentioned above the quantities are relatively similar at \$782 and \$764, respectively, when valued at Medicare prices.

each type.

2.12.1 Computing equilibrium outcomes in counterfactual scenarios

We compute counterfactual equilibrium outcomes relative to baseline equilibrium outcomes by rewriting the equilibrium system of equations in terms of the initial allocation, constant elasticities, relative exogenous parameters, and relative endogenous equilibrium outcomes, a technique known as "exact hat algebra" in the trade literature.

If $K(\delta) = \delta$ and $H(Q) = Q^{\alpha}$, an equilibrium is a set of quantities and qualities $\{Q_i, \delta_i\}_{i \in \mathcal{I}}$ that simultaneously satisfy equations (2.4) and (2.1) and $Q_i = \sum_j Q_{ij}$. Consider two equilibria: the baseline equilibrium and the counterfactual equilibrium. Define export shares $x_{ij} \equiv \frac{Q_{ij}}{\sum_{j'} Q_{ij'}}$ and import shares $m_{ij} \equiv \frac{Q_{ij}}{N_j}$ in the baseline equilibrium. Denote the counterfactual parameters and equilibrium outcomes by primes. Plugging $Q_i = \sum_j Q_{ij}$ into equation (2.4), we can write the system of equations for each equilibrium as

$$\begin{aligned} \delta'_i &= \left(\frac{R'_i A'_i}{w'_i}\right)\left(\sum_j Q'_{ij}\right)^{\alpha} & \qquad Q'_{ij} &= \delta'_i \frac{\rho'_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta'_{i'} \rho'_{i'j}} N'_j \\ \delta_i &= \left(\frac{R_i A_i}{w_i}\right)\left(\sum_j Q_{ij}\right)^{\alpha} & \qquad Q_{ij} &= \delta_i \frac{\rho_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}} N_j \end{aligned} .$$

Define $\hat{y} \equiv \frac{y'}{y}$ for every variable y. For example, $\hat{\delta}_i \equiv \frac{\delta'_i}{\delta_i}$.

We now rewrite the counterfactual equilibrium equations in terms of baseline equilibrium shares x_{ij}, m_{ij}, the scale elasticity α, (relative) counterfactual exogenous parameters $\hat{A}, \hat{R}, \hat{w}, \hat{\rho}, \hat{N}$, and (relative) counterfactual endogenous qualities $\hat{\delta}$.

First, divide the counterfactual free-entry condition by the baseline free-entry condition to obtain an expression for relative quality:

$$\frac{\delta'_i}{\delta_i} = \frac{\hat{R}_i \hat{A}_i}{\hat{w}_i}\left(\frac{\sum_{j \in \mathcal{I}} Q'_{ij}}{\sum_{j \in \mathcal{I}} Q_{ij}}\right)^{\alpha} = \frac{\hat{R}_i \hat{A}_i}{\hat{w}_i}\left(\sum_{j \in \mathcal{I}} \frac{Q_{ij}}{\sum_{j \in \mathcal{I}} Q_{ij}} \frac{Q'_{ij}}{Q_{ij}}\right)^{\alpha} = \frac{\hat{R}_i \hat{A}_i}{\hat{w}_i}\left(\sum_{j \in \mathcal{I}} x_{ij} \frac{Q'_{ij}}{Q_{ij}}\right)^{\alpha} \tag{2.19}$$

Second, divide the counterfactual gravity equation by the baseline gravity equation to obtain an expression for relative bilateral flows:

$$\begin{aligned}\frac{Q'_{ij}}{Q_{ij}} &= \frac{\delta'_i}{\delta_i}\left(\frac{\frac{\rho'_{ij}}{\sum_{i'\in 0\cup\mathcal{I}}\delta'_{i'}\rho'_{i'j}}N'_j}{\frac{\rho_{ij}}{\sum_{i'\in 0\cup\mathcal{I}}\delta_{i'}\rho_{i'j}}N_j}\right) = \frac{\frac{\delta'_i}{\delta_i}\frac{\rho'_{ij}}{\rho_{ij}}\frac{N'_j}{N_j}}{\sum_{i'\in 0\cup\mathcal{I}}\frac{\delta_{i'}\rho_{i'j}}{\sum_{i'\in 0\cup\mathcal{I}}\delta_{i'}\rho_{i'j}}\frac{\delta'_{i'}\rho'_{i'j}}{\delta_{i'}\rho_{i'j}}} \\ &= \frac{\hat{\delta}_i\hat{\rho}_{ij}\hat{N}_j}{\sum_{i'\in 0\cup\mathcal{I}}\frac{Q_{i'j}}{N_j}\hat{\delta}_{i'}\hat{\rho}_{i'j}} = \frac{\hat{\delta}_i\hat{\rho}_{ij}\hat{N}_j}{m_{0j}+\sum_{i'\in\mathcal{I}}m_{i'j}\hat{\delta}_{i'}\hat{\rho}_{i'j}}\end{aligned}$$

Plug this expression for relative bilateral flows into equation (2.19) and rearrange terms to obtain the following system of $\mathcal{I}$ equations with unknowns $\{\hat{\delta}_i\}_{i=1}^{\mathcal{I}}$:

$$\hat{\delta}_i = \left(\hat{R}_i\hat{A}_i/\hat{w}_i\right)^{\frac{1}{1-\alpha}}\left(\sum_{j\in\mathcal{I}}\frac{x_{ij}\hat{\rho}_{ij}\hat{N}_j}{m_{0j}+\sum_{i'\in\mathcal{I}}m_{i'j}\hat{\delta}_{i'}\hat{\rho}_{i'j}}\right)^{\frac{\alpha}{1-\alpha}}. \tag{2.20}$$

2.12.2 Inferring the number of potential patients

A baseline calibration of our model requires α, x_{ij}, and m_{ij} in order to use equation (2.20) to compute relative counterfactual outcomes. We have estimated α. The export shares $x_{ij} \equiv \frac{Q_{ij}}{\sum_j Q_{ij}}$ are easily computed using the observed trade matrix.[58] The challenge is computing import shares $m_{ij} \equiv \frac{Q_{ij}}{N_j}$ because we do not observe N_j; while we observe the number of Medicare beneficiaries in region j, not all beneficiaries are in the market for all services. This section describes the assumptions we make in order to infer the values of the relevant market size N_j $\forall j \in \mathcal{I}$. Specifically, we assume per capita demand is uniform, outside-option quality is constant across regions, and the average outside-option share is 10%, as described below.

We have estimated $\theta_j = N_j/\Phi_j$ in equation (2.12). We observe the number of beneficiaries

58. Dingel and Tintelnot (2021) document overfitting problems when calibrating gravity models using noisy observed shares. We obtain similar counterfactual outcomes when calibrating our model using gravity-predicted shares.

enrolled in Traditional Medicare in region j, which we denote S_j^{TM}. By definition, $m_{0j} = \frac{\delta_{0j}}{\Phi_j}$. We assume $\delta_{0j} = \delta_0 \; \forall j$ and $N_j \propto S_j^{\text{TM}}$. This implies

$$m_{0j} = \frac{\delta_{0j}}{\Phi_j} = \frac{\delta_0 \theta_j}{N_j} = \frac{\delta_0 \theta_j}{\mathfrak{s} S_j^{\text{TM}}},$$

where $\mathfrak{s}$ is a constant of proportionality. We set $\frac{\delta_0}{\mathfrak{s}}$ such that the average outside-option share is 10%, $\frac{1}{\mathcal{I}} \sum_j m_{0j} = 0.1$. This requires $\frac{\delta_0}{\mathfrak{s}} = \frac{0.1 \times \mathcal{I}}{\sum_j \theta_j / S_j^{\text{TM}}}$. With m_{0j} in hand, we can infer N_j:

$$m_{0j} = 1 - \sum_{i \in \mathcal{I}} m_{ij} = 1 - \frac{1}{N_j} \sum_{i \in \mathcal{I}} Q_{ij} \implies N_j = \frac{1}{1 - m_{0j}} \sum_{i \in \mathcal{I}} Q_{ij}.$$

With N_j in hand, we can compute all import shares, $m_{ij} = \frac{Q_{ij}}{N_j} \; \forall i \in 0 \cup \mathcal{I}, \forall j \in \mathcal{I}$.

We exclude the Anchorage, Alaska HRR from our counterfactual computations. The entire state of Alaska is one (geographically isolated and very large) HRR. The average within-Alaska-HRR procedure incurs more than 60 kilometers of travel. In the gravity regression, Alaska has the smallest exporter fixed effect: very few patients travel to Alaska for care. Alaska's importer fixed effect is quite large because Alaskans import about 15% of their services and the average import traverses $3,616$ kilometers. As a result, the implied outside-option share would exceed one when we set the nationwide average to 10%. We therefore exclude the Alaska HRR from the economy when computing counterfactual outcomes. Given its considerable geographic isolation, Alaska would have little influence on outcomes in other regions.

The qualitative and spatial patterns of counterfactual outcomes are the same if we assume the average outside-option share is 20% rather than 10%.

2.12.3 Counterfactual outcomes with multiple patient types

This section describes how to compute counterfactual equilibrium outcomes relative to baseline equilibrium outcomes when there are multiple patient types who face heterogeneous trade costs. The derivation is very similar to that of Section 2.12.1. Define import shares $m_{ij\kappa} \equiv \frac{Q_{ij\kappa}}{N_{j\kappa}}$ in the baseline equilibrium. Define patient-type shares $n_{j\kappa} \equiv \frac{N_{j\kappa}}{N_j}$. We rewrite the system of baseline and counterfactual gravity equations (2.15) and free-entry condition (2.4) as follows:

$$\begin{array}{ll} \delta'_i = \left(\frac{R'_i A'_i}{w'_i}\right)\left(\sum_j Q'_{ij}\right)^{\alpha} & Q'_{ij} = \delta'_i \sum_{\kappa} \frac{\rho'_{ij\kappa}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta'_{i'} \rho'_{i'j\kappa}} N'_{j\kappa} \\ \delta_i = \left(\frac{R_i A_i}{w_i}\right)\left(\sum_j Q_{ij}\right)^{\alpha} & Q_{ij} = \delta_i \sum_{\kappa} \frac{\rho_{ij\kappa}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j\kappa}} N_{j\kappa} \end{array}.$$

As above, dividing the counterfactual free-entry condition by the baseline free-entry condition yields the expression for relative quality in equation (2.19). Second, divide the counterfactual gravity equation by the baseline gravity equation to obtain an expression for relative bilateral flows:

$$\frac{Q'_{ij}}{Q_{ij}} = \frac{\delta'_i}{\delta_i} \left(\frac{\sum_{\kappa} \frac{\rho'_{ij\kappa}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta'_{i'} \rho'_{i'j\kappa}} N'_{j\kappa}}{\sum_{\kappa} \frac{\rho_{ij\kappa}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j\kappa}} N_{j\kappa}} \right) = \frac{\delta'_i}{\delta_i} \left(\frac{\sum_{\kappa} \frac{\rho'_{ij\kappa}}{\Phi'_{j\kappa}} N'_{j\kappa}}{\sum_{\kappa} \frac{\rho_{ij\kappa}}{\Phi_{j\kappa}} N_{j\kappa}} \right) = \hat{\delta}_i \sum_{\kappa} \frac{n_{j\kappa}}{\frac{m_{ij}}{m_{ij\kappa}}} \frac{\hat{\rho}_{ij\kappa}}{\hat{\Phi}_{j\kappa}} \hat{N}_{j\kappa}$$

Plugging this expression for relative bilateral flows into equation (2.19) and then rearranging terms yields the following system of $\mathcal{I}$ equations with unknowns $\{\hat{\delta}_i\}_{i=1}^{\mathcal{I}}$:

$$\hat{\delta}_i = \left(\hat{R}_i \hat{A}_i / \hat{w}_i\right)^{\frac{1}{1-\alpha}} \left(\sum_{j \in \mathcal{I}} x_{ij} \left(\sum_{\kappa} \frac{n_{j\kappa}}{m_{ij}} \frac{m_{ij\kappa} \hat{\rho}_{ij\kappa}}{m_{0j\kappa} + \sum_{i' \in \mathcal{I}} m_{i'j\kappa} \hat{\delta}_{i'} \hat{\rho}_{i'j\kappa}} \hat{N}_{j\kappa} \right) \right)^{\frac{\alpha}{1-\alpha}}. \quad (2.21)$$

2.12.4 Inferring the number of potential patients of each type

Because we do not observe patients who select the outside option, we make assumptions that allow us to infer $N_{j\kappa}$ and thus $n_{j\kappa}$ and $m_{ij\kappa}$, which are needed to compute counterfactual outcomes using equation (2.21). We start from a type-specific variant of the gravity equation (2.7) with fixed effects, as in the single-type equation (2.12). The estimating equation is

$$\ln \mathbb{E}\left(\overline{R}Q_{ij\kappa}\right) = \ln \delta_i + \ln\left(\frac{N_{j\kappa}}{\Phi_{j\kappa}}\right) + \gamma^{\kappa} X_{ij} = \ln \delta_i + \ln \theta_{j\kappa} + \gamma^{\kappa} X_{ij}.$$

This yields an estimate of $\theta_{j\kappa} = N_{j\kappa}/\Phi_{j\kappa}$.

As in the single-type case above, we assume per capita demand is uniform and outside-option quality is constant across regions. We observe the number of beneficiaries of type κ enrolled in Traditional Medicare in region j, which we denote $S_{j\kappa}^{\mathrm{TM}}$. We assume $\delta_{0j} = \delta_0 \ \forall j$ and $N_{j\kappa} = \mathfrak{s} S_{j\kappa}^{\mathrm{TM}}$, where $\mathfrak{s}$ is a constant of proportionality that is common across types. This implies

$$m_{0j\kappa} = \frac{\delta_0}{\Phi_{j\kappa}} = \frac{\delta_0 \theta_{j\kappa}}{N_{j\kappa}} = \frac{\delta_0 \theta_{j\kappa}}{\mathfrak{s} S_{j\kappa}^{\mathrm{TM}}}.$$

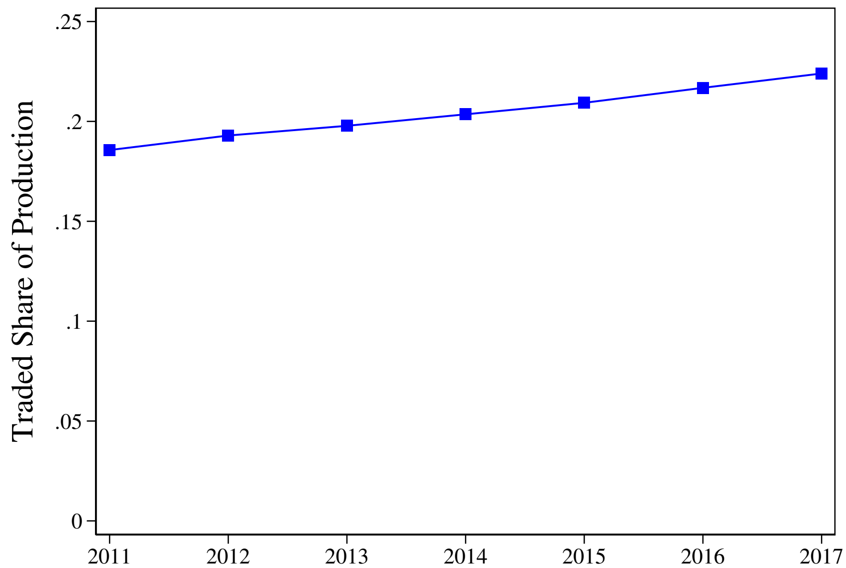
Let $\mathcal{K} = \sum_{\kappa} 1$ denote the number of patient types. We set $\frac{\delta_0}{\mathfrak{s}}$ such that the average outside-option share, across all types, is 10%, $\frac{1}{\mathcal{I}\mathcal{K}} \sum_{j\kappa} m_{0j\kappa} = 0.1$. This implies

$$m_{0j\kappa} = 0.1 \times \frac{\theta_{j\kappa}/S_{j\kappa}^{\mathrm{TM}}}{\frac{1}{\mathcal{I}\mathcal{K}} \sum_{j'\kappa'} \theta_{j'\kappa'}/S_{j'\kappa'}^{\mathrm{TM}}}.$$

Using the resulting $N_{j\kappa} = \frac{1}{1-m_{0j\kappa}} \sum_{i \in \mathcal{I}} Q_{ij\kappa}$ allows us to compute all import shares.

2.13 Additional exhibits

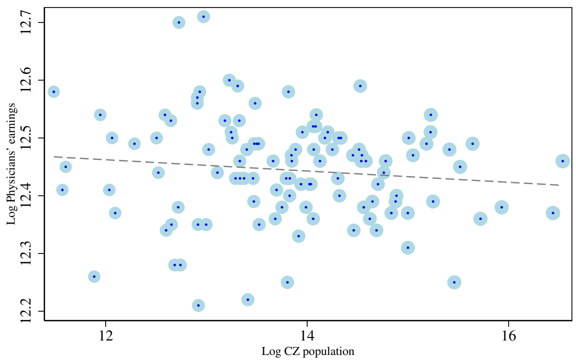
Figure 2.13: Trade in medical services has increased over time



Note: This figure shows the annual exported share of production from 2011 to 2017. Production and trade are computed using the Medicare 20% carrier Research Identifiable Files for the relevant years. Production is exported when the patient's address and the service location are in different HRRs. HRR definitions are from the Dartmouth Atlas Project.

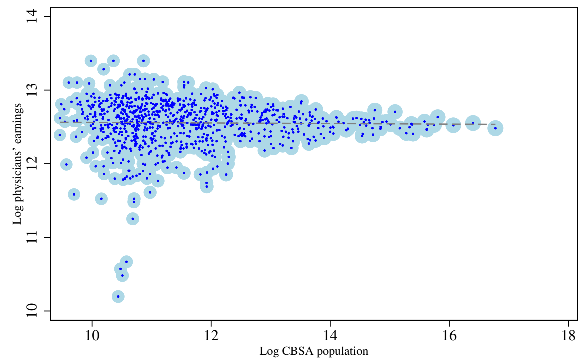
Figure 2.14: Population elasticities of input costs

(a) Physicians' earnings (commuting zones)



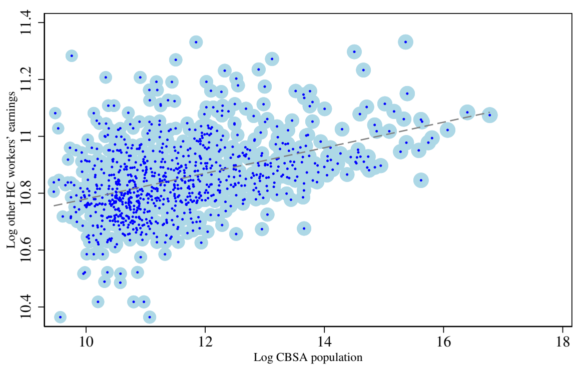
Y = 12.5794 − 0.0098 [0.0082] Log CZ population + ε
N = 111; R^2 = 0.0127

(b) Physicians' earnings (CBSAs)



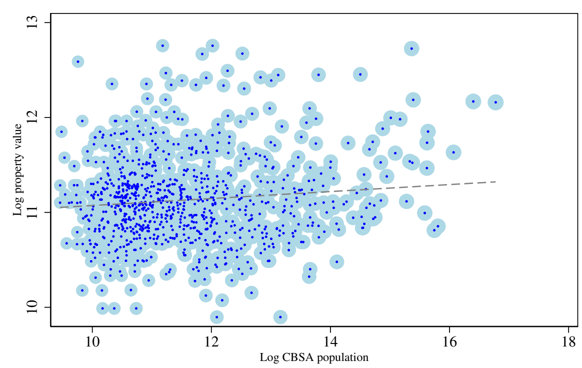
Y = 12.5994 − 0.0038 [0.0082] Log CBSA population + ε
N = 893; R^2 = 0.0002

(c) Other healthcare workers' earnings



Y = 10.3307 + 0.0449 [0.0033] Log CBSA population + ε
N = 896; R^2 = 0.1748

(d) Median house value

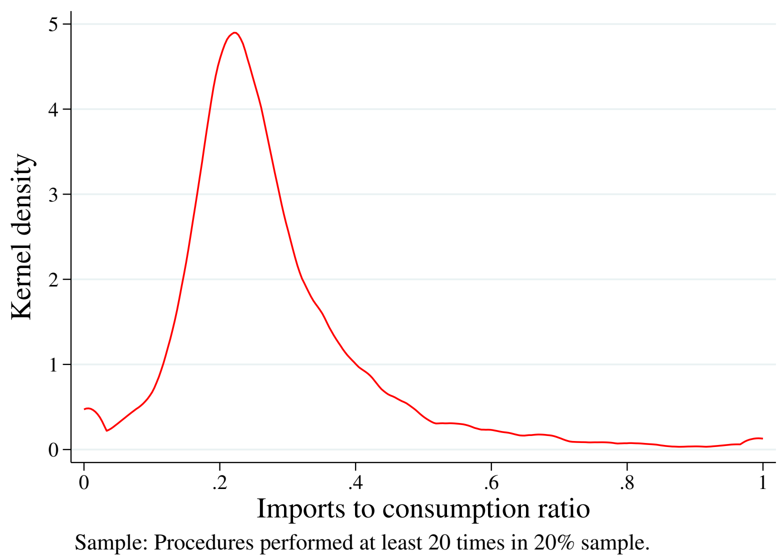


Y = 10.7048 + 0.0368 [0.0116] Log CBSA population + ε
N = 896; R^2 = 0.0110

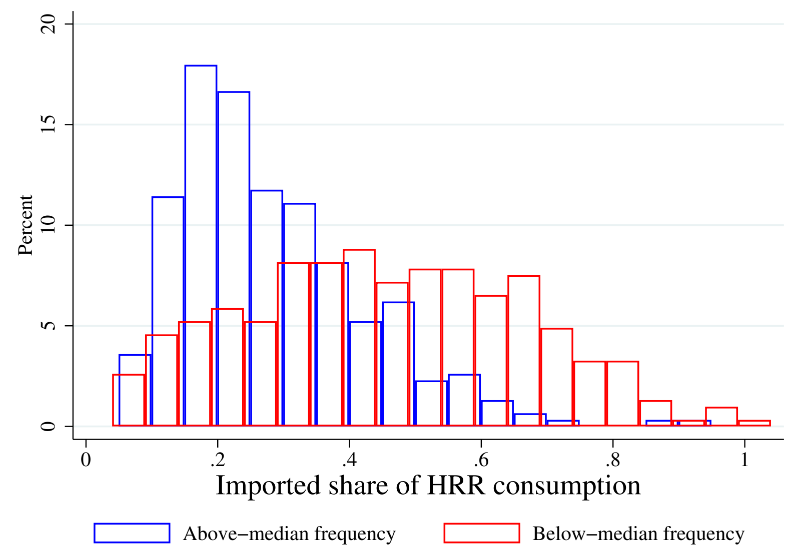
Notes: This figure depicts relationships between input costs and population sizes. Panel (a) shows physicians' earnings across 111 commuting zones using data from Gottlieb et al. (2020). Panels (b), (c), and (d) show variation across CBSAs in physicians' earnings, other healthcare workers' earnings, and median house values (a proxy for real estate and other locally priced inputs) using data from the 2015–2019 American Community Survey.

Figure 2.15: Variation in trade shares across procedures and regions

(a) Distribution of import share by procedure

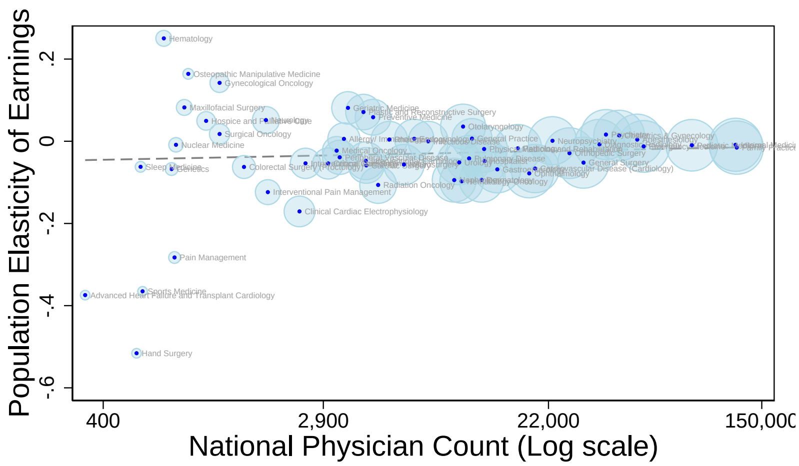


(b) Distributions of import shares for common and rare services



Notes: Panel (a) shows the distribution of the imported consumption share across procedures for procedures performed at least 20 times (in our 20% sample of Medicare claims). Imports are defined as care provided to a patient who lives in one HRR at a service location in a different HRR. Panel (b) splits all services into two groups based on how often they are performed nationally. Those performed less often than the median are shown in red, and those performed more often than the median service are shown in blue. Import shares are substantially higher for the rarer services.

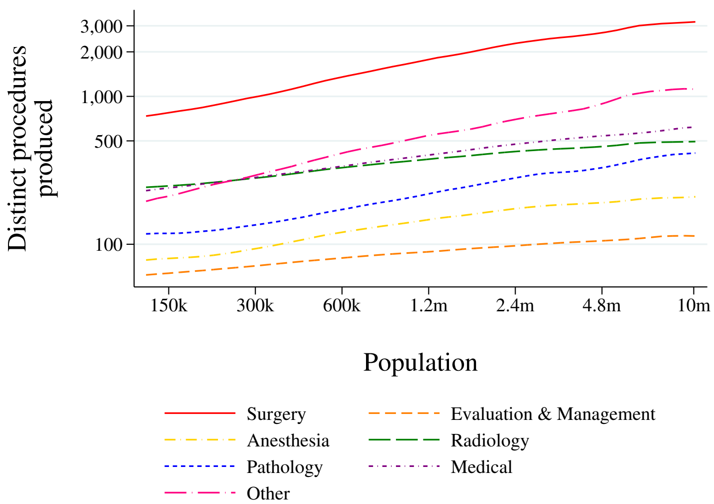
Figure 2.16: Specialists' income patterns do not explain the output-population gradient



Both the regression and scatter markers are weighted by
number of underlying observations for specialty's elasticity estimate
Pearson correlation: 0.113.
Fitted line: y = 0.005 (0.006) * x + -0.077 (0.061)
N = 56
Physician income and count data from Gottlieb et al. (2020)

Notes: This figure shows the population elasticity of income for different medical specialties against the total number of physicians in those specialties. For each specialty, we estimate the elasticity of income with respect to population across commuting zones, using data from Gottlieb et al. (2020). The graph shows that these elasticities are unrelated to the total national count of physicians in those specialties.

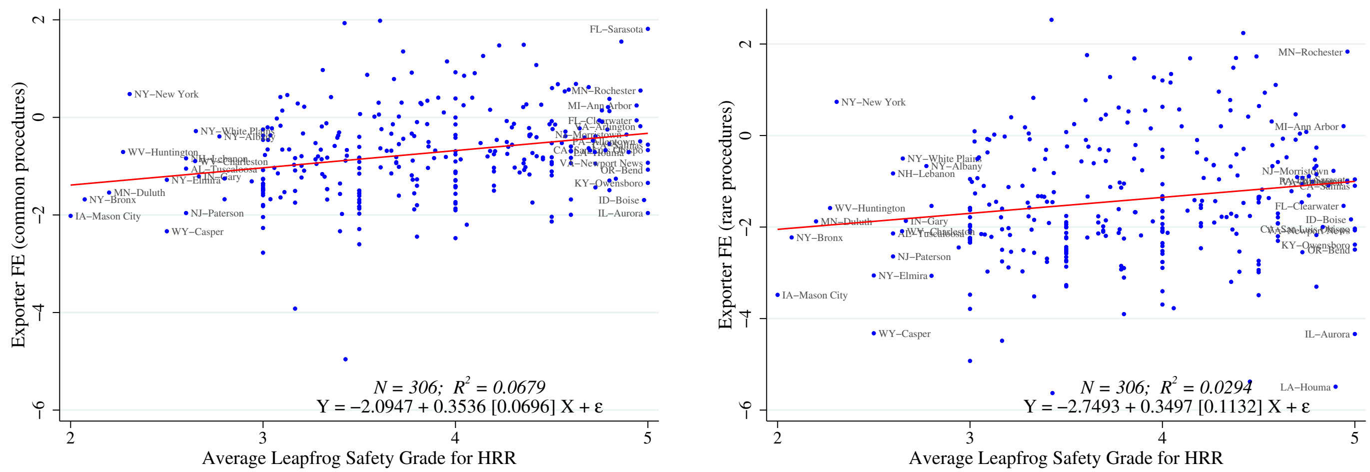
Figure 2.17: Larger markets produce a greater variety of procedures



Notes: This figure shows the local relationship between the number of distinct services performed in the Medicare data in a given HRR and that HRR's population. More populous HRRs perform more unique services; Table 2.20 reports the population elasticities. We use procedure classifications from the American Academy of Professional Coders, which groups codes into surgeries, anesthesia, radiology, pathology, medical, and evaluation & management services (AAPC, 2021). We combine Category II codes, Category III codes and Multianalyte Assays into "other."

Figure 2.18: Leapfrog Safety Grade vs. estimated quality: common and rare

(a) Leapfrog Safety Grade vs. quality for common services (b) Leapfrog Safety Grade vs. quality for rare services

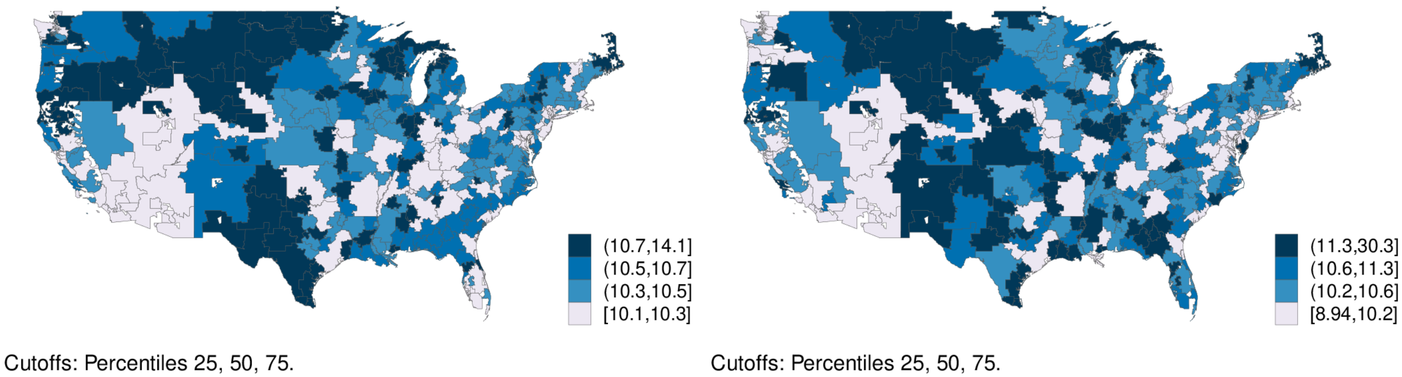


Notes: This figure shows the relationship between exporter fixed effects, estimated separately for common and rare services, and the Leapfrog Safety Grade. The vertical axis shows the exporter fixed effects for each HRR estimated from equation (2.12), in Panel (a) using trade in common services, and in Panel (b) using trade in rare services. The horizontal axis in both panels is the average safety grade for hospitals in an HRR, determined by the Leapfrog Group. The Leapfrog Safety Grades range from A to F, which we scale as integers from 1 (for F) to 5 (for A). We then compute the mean score for all hospitals in the HRR. The Safety Grades are positively associated with the exporter fixed effects for both rare and common procedures.

Figure 2.19: Counterfactual change in quality δ for rare vs. common services when increasing reimbursement by 10% everywhere

(a) Change (%) in quality δ for common services

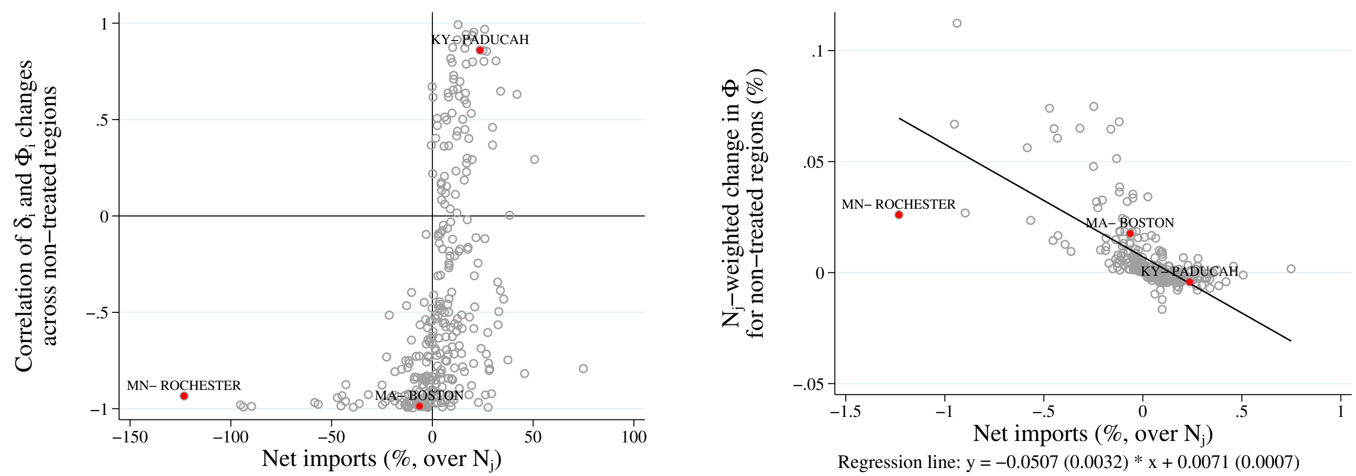
(b) Change (%) in quality δ for rare services



Notes: Both panels show the impacts of increasing reimbursements by 10% everywhere ($\hat{R}_i = 1.1$ for all i) on the quality of production in each region, δ_i. Panel (a) illustrates the change for common services, and Panel (b) for rare services. Each panel is based on the baseline trade matrix for the respective set of services. Panel (a) uses an agglomeration elasticity of $\alpha = 0.6$ and Panel (b) uses $\alpha = 0.9$. The common-services scenario excludes the Alaska HRR and the rare-services scenario excludes four HRRs (Alaska, Hawaii, Houma, La., and Minot, N.D.). The pattern of outcomes is qualitatively similar but the magnitudes vary more for rare services.

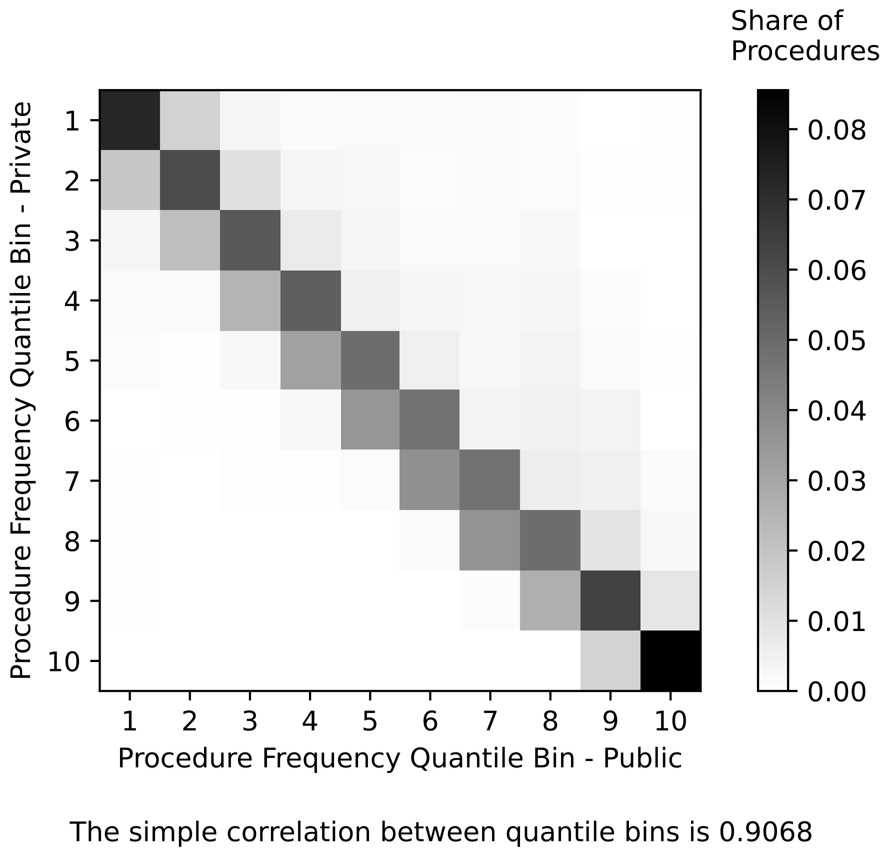
Figure 2.20: Spillovers from higher reimbursements in one region depend on that region's net imports

(a) Correlation of $\hat{\delta}_i$ and $\hat{\Phi}_i$ across non-treated regions (b) Change in non-treated regions' aggregate market access



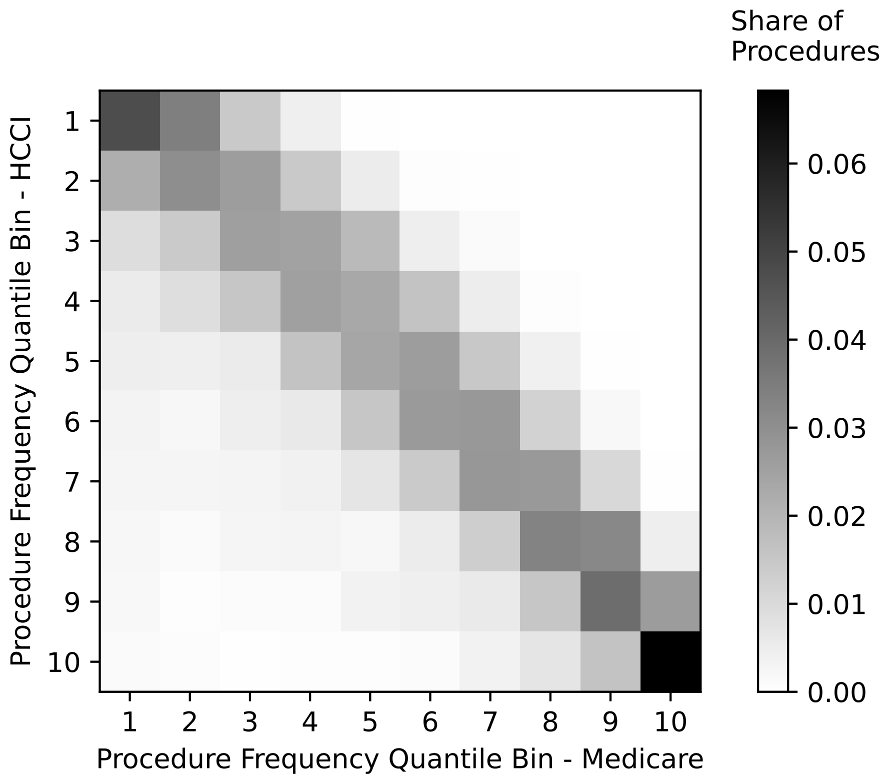
Notes: This figure characterizes counterfactual outcomes when raising reimbursements by 30 percent in one HRR. We conduct this exercise for each region, one at a time, and each observation in each panel represents one such counterfactual scenario. Panel (a) illustrates the contrast in spillovers as a function of net imports of the treated region. The vertical-axis value for each observation reports the correlation—across *all regions other than the treated one* for the exercise in question—between the counterfactual changes $\hat{\delta}_i$ and $\hat{\Phi}_i$. The scatterplot relates these correlations to the *treated* region's net import share, which is plotted on the horizontal axis. When the treated region is a net exporter, changes in quality δ_i and in market access Φ_i for non-treated regions move in opposite directions: a region whose output quality declines experiences an increase in market access through imports from the treated region. However, increasing reimbursements in a net-importing region often has the opposite effect: neighboring regions with quality reductions also experience lower market access, (changes in δ_i and Φ_i are positively correlated). For each counterfactual, the vertical-axis value in Panel (b) shows the aggregate impact on patient market access *excluding the treated region*. The panel relates this impact to the *treated* region's net imports, shown on the horizontal axis. When the treated region is a net importer, the aggregate impact on market access for non-treated regions tends to be smaller or even negative.

Figure 2.21: Deciles of Procedure Frequency in Confidential and Public Medicare Data



Notes: This figure shows the share of procedures in each frequency decile in the Medicare public data compared to the Medicare confidential data. The classification of procedures by frequency deciles appears largely consistent between the two data sources for Medicare patients.

Figure 2.22: Deciles of Procedure Frequency in Medicare and Private Insurance Data



The simple correlation between quantile bins is 0.8287

Notes: This figure shows the share of procedures in each frequency decile in the Medicare versus privately insured data. The classification of procedures by frequency deciles appears largely consistent when comparing public Medicare data with data on privately insured patients from the Health Care Cost Institute (HCCI).

Table 2.7: Higher-income patients are less sensitive to distance: Procedure-level estimates

	(1) 25min visit	(2) cataract removal	(3) knee joint repair	(4) heart artery bypass	(5) gallblader removal
Distance (log)	-2.075	-2.281	-2.255	-2.246	-2.135
	(0.0790)	(0.0829)	(0.0947)	(0.0876)	(0.0855)
Distance (log) × income tercile 2	0.0946	0.143	0.171	0.0987	0.205
	(0.0610)	(0.0819)	(0.0754)	(0.0823)	(0.0685)
Distance (log) × income tercile 3	0.206	0.287	0.227	0.402	0.314
	(0.0777)	(0.0914)	(0.0937)	(0.0927)	(0.0907)
Observations	271,728	268,400	262,352	240,352	250,800
Patient market-income FE & Provider market FE	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the coefficient on log distance for each income tercile from gravity regressions estimated separately for five procedures varying in frequency: 25 min office visit (HCPCS 99214), cataract removal (66984), knee joint repair (27447), heart artery bypass (33533), and gallblader removal (47562). The dependent variable in all regressions is the number of procedures traded. Each regression includes log distance interacted with an income tercile indicator, an indicator for same-HRR observations ($i = j$), an exporting HRR fixed effect, and an income-tercile-importing-HRR fixed effect. The coefficients for higher income terciles are positive, indicating that patients residing in higher-income ZIP codes are less sensitive to distance. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.8: Estimates of a strong home-market effect by CBSA

	(1)	(2)	(3)	(4)	(5)	(6)
Estimation method:	PPML	PPML	PPML	IV	PPML	IV
Instrument:				1940 pop		Bedrock
Provider-market population (log)	0.734	0.739	0.703	0.716	0.739	1.161
	(0.0232)	(0.0234)	(0.0205)	(0.0249)	(0.0259)	(0.307)
Patient-market population (log)	0.395	0.393	0.417	0.396	0.394	0.178
	(0.0290)	(0.0292)	(0.0264)	(0.0261)	(0.0311)	(0.373)
Distance (log)	-2.311	-3.464		-3.403	-3.400	-4.677
	(0.0493)	(0.324)		(0.295)	(0.347)	(1.056)
Distance (log, squared)		0.110		0.104	0.105	0.210
		(0.0323)		(0.0288)	(0.0346)	(0.0850)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.000	0.000	0.000	0.000	0.000	0.063
Observations	857,476	857,476	857,476	857,476	781,456	781,456
Sample:	All CBSAs	All CBSAs	All CBSAs	All CBSAs	Bedrock data	Bedrock data
Distance elasticity at mean		-1.90		-1.92	-1.90	-1.68
Distance deciles			Yes			

Notes: This table reports estimates of equation (2.8), which estimates the presence of weak or strong home-market effects. The dependent variable in all regressions is the value of trade computed by assigning each procedure its national average price. The independent variables are patient- and provider-market log population, log distance between CBSAs, and an indicator for same-CBSA observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Column 5 reports the PPML estimate on the subsample of regions for which we have data on depth to bedrock available ($N = 884^2$). Column 6 uses depth to bedrock in the importing and exporting regions as instruments for current log population in those regions, respectively. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.9: Estimates of a strong home-market effect including facility spending

	(1)	(2)	(3)	(4)
Estimation method:	PPML	PPML	PPML	IV
Provider-market population (log)	0.687	0.700	0.689	0.829
	(0.0576)	(0.0525)	(0.0382)	(0.0586)
Patient-market population (log)	0.226	0.217	0.255	0.266
	(0.0571)	(0.0507)	(0.0345)	(0.0470)
Distance (log)	-1.635	0.877		0.932
	(0.0495)	(0.316)		(0.253)
Distance (log, squared)		-0.254		-0.258
		(0.0319)		(0.0250)
Same hrr	0.434	1.791	4.685	
	(0.174)	(0.238)	(0.0637)	
Observations	93,636	93,636	93,636	93,636
Distance elasticity at mean		-2.76		-2.76
Distance deciles			Yes	

Notes: This table reports estimates of equation (2.8), which estimates the presence of weak or strong home-market effects, when including professional and facility fees. The sample is all HRR pairs ($N = 306^2$). The dependent variable in all regressions is the value of trade when including professional and facility (inpatient and outpatient) fees at national average prices. The independent variables are patient- and provider-market log population, log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Trade flows are computed from the Medicare 20% carrier, MedPAR, and outpatient claims Research Identifiable Files, excluding emergency-room care and skilled nursing facilities. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.10: Estimates of a strong home-market effect excluding AZ, FL, CA

	(1)	(2)	(3)	(4)
Estimation method:	PPML	PPML	PPML	IV
Provider-market population (log)	0.647	0.647	0.649	0.663
	(0.0811)	(0.0701)	(0.0425)	(0.0626)
Patient-market population (log)	0.375	0.383	0.414	0.400
	(0.0809)	(0.0693)	(0.0421)	(0.0570)
Distance (log)	-1.748	1.690		1.707
	(0.0608)	(0.428)		(0.397)
Distance (log, squared)		-0.360		-0.361
		(0.0429)		(0.0396)
Observations	67,600	67,600	67,600	67,600
Distance elasticity at mean		-3.35		-3.35
Distance deciles			Yes	

Notes: This table reports estimates of equation (2.8), which estimates the presence of weak or strong home-market effects, excluding snowbird states. The sample is all HRR pairs, excluding those in Arizona, Florida, or California. The dependent variable in all regressions is the value of trade computed by assigning each procedure its national average price. The independent variables are patient- and provider-market log population, log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.11: Estimates of a strong home-market effect excluding HRRs with high second-home share

	(1)	(2)	(3)	(4)
Estimation method:	PPML	PPML	PPML	IV
Provider-market population (log)	0.654	0.663	0.661	0.679
	(0.0652)	(0.0641)	(0.0453)	(0.0571)
Patient-market population (log)	0.369	0.362	0.392	0.382
	(0.0639)	(0.0619)	(0.0424)	(0.0564)
Distance (log)	-1.675	0.364		0.372
	(0.0509)	(0.307)		(0.279)
Distance (log, squared)		-0.210		-0.211
		(0.0300)		(0.0273)
Observations	76,176	76,176	76,176	76,176
Distance elasticity at mean		-2.64		-2.64
Distance deciles			Yes	

Notes: This table reports estimates of equation (2.8), which estimates the presence of weak or strong home-market effects, excluding HRRs with a high second-home share. The sample is all HRR pairs excluding those in the top 10% based on the share of housing units that are vacant for seasonal/recreational purposes in the 2013–2017 American Community Survey. See Table 2.10 notes on the variables, instruments, geographic units, and standard errors.

Table 2.12: Travel for dialysis

Distance (km)	Share of output				
	All (Professional)	All (Facility)	All (Dialysis)	No snowbird states (Dialysis)	Snowbird states (Dialysis)
$[0, 50)$	0.77	0.77	0.94	0.94	0.93
$[50, 100)$	0.12	0.12	0.03	0.03	0.03
$[100, .)$	0.11	0.11	0.03	0.02	0.04

Notes: For the care described in each column and the distance intervals in each row, the entries in this table report the share of patients traveling that distance from their residential ZIP code to the service location's ZIP code. The first column shows professional claims (from Medicare's "carrier" file), the second column shows facility (hospital) claims, and the third column shows dialysis claims. The remaining columns split dialysis claims between "snowbird" states (AZ, CA, and FL, following Finkelstein, Gentzkow, and Williams 2016) and other states. In non-snowbird states, the table shows that 94% of patients travel less than 50 km from their home for dialysis, and only 2% more than 100 km. This is less than one-fifth as much as for other facility or professional care, suggesting that residential location is recorded correctly for almost all patients.

Table 2.13: Contrasting geographies of colonoscopies and LVAD insertions

	Colonoscopy	LVAD Insertion
Code	G0121	33979
N	58,798	333
Physicians	13,475	177
$\hat{\beta}_p^{\text{production}}$	0.00	0.71
$\hat{\beta}_p^{\text{consumption}}$	-0.01	0.03
Share traded (HRR)	0.15	0.50
Share traded (CBSA)	0.15	0.48
Median distance traveled (km)	18.44	65.50
Share > 100km	0.06	0.37

Notes: This table reports statistics for two HCPCS codes: screening colonoscopy (G0121) and LVAD insertion (33979). We report the number of times the procedure is performed in 2017 in our 20% sample of Medicare patients and the number of distinct physicians performing it. The population elasticities of production and consumption are estimated using the Poisson models in equations (2.9) and (2.10) based on production HRR and patients' residential HRR, respectively. We also report the shares of procedures in which the patient and service location are in different HRRs or CBSAs, the median distance traveled for all care, and the share in which the patient and service location are more than 100 kilometers apart.

Table 2.14: Estimates of a stronger home-market effect for rare diagnoses including facility spending

	(1)	(2)	(3)	(4)	(5)	(6)
λ_X Provider-market population (log)	0.665	0.659	0.649		0.662	
	(0.0573)	(0.0560)	(0.0557)		(0.0516)	
λ_M Patient-market population (log)	0.240	0.241	0.246		0.239	
	(0.0567)	(0.0551)	(0.0548)		(0.0498)	
μ_X Provider-market population (log) $\times$ rare			0.223	0.209	0.231	0.207
			(0.0243)	(0.0206)	(0.0236)	(0.0204)
μ_M Patient-market population (log) $\times$ rare			-0.0793	-0.0753	-0.0841	-0.0700
			(0.0216)	(0.0158)	(0.0210)	(0.0161)
Observations	187,272	147,814	147,814	147,814	147,814	147,814
Distance controls	Yes	Yes	Yes	Yes		
Distance [quadratic] controls					Yes	Yes
Patient-provider-market-pair FEs				Yes		Yes

Notes: This table reports estimates of equation (2.11), which introduces interactions with an indicator for whether a diagnosis is "rare" (provided to less patients than the median diagnosis, when adding up all diagnoses nationally). The dependent variable in all regressions is the value of trade when including professional and facility (inpatient and outpatient) fees at national average prices. The interactions with patient- and provider-market population reveal whether the home-market effect is larger for rare diagnoses. The unit of observation is {rare indicator, exporting HRR, importing HRR} so the number of observations is 2×306^2 in column 1. All diagnoses are included. Columns 2 onwards drop HRR pairs with zero trade in both diagnosis groups, which leads to a larger sample than in Table 2.4 because trade in facility fees is included in addition to professional fees for all diagnoses. Column 2 shows that this restriction has a negligible impact on the estimated log population coefficients. Columns 1–4 control for distance using the log of distance between HRRs. Columns 5 and 6 add a control for the square of log distance. Columns 4 and 6 introduce a fixed effect for each ij pair of patient market and provider market, so these omit all covariates that are not interacted with the rare indicator. The positive coefficient on provider-market population $\times$ rare across all columns indicates that the home-market effect is stronger for rare than for common services. The negative coefficient on patient-market population $\times$ rare across all columns indicates that the *strong* home-market effect has a larger magnitude for rare services. Trade flows are computed from the Medicare 20% carrier, MedPAR, and outpatient claims Research Identifiable Files, excluding emergency-room care and skilled nursing facilities. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.15: Home-market effect is stronger for rare services controlling for patient engagement

	(1)	(2)
Provider-market population (log) × common × high engagement	-0.0355	-0.0354
	(0.0349)	(0.0347)
Provider-market population (log) × rare × low engagement	0.231	0.244
	(0.0482)	(0.0370)
Provider-market population (log) × rare × high engagement	0.481	0.360
	(0.0808)	(0.141)
Patient-market population (log) × common × high engagement	0.0440	0.0450
	(0.0257)	(0.0255)
Patient-market population (log) × rare × low engagement	-0.146	-0.125
	(0.0374)	(0.0243)
Patient-market population (log) × rare × high engagement	-0.477	-0.575
	(0.0923)	(0.271)
Distance (log) × common × high engagement	-0.0548	0.146
	(0.0209)	(0.118)
Distance (log) × rare × low engagement	0.0488	0.716
	(0.0375)	(0.171)
Distance (log) × rare × high engagement	-0.127	2.458
	(0.0814)	(2.764)
Distance (log, squared) × common × high engagement		-0.0193
		(0.0100)
Distance (log, squared) × rare × low engagement		-0.0615
		(0.0152)
Distance (log, squared) × rare × high engagement		-0.277
		(0.324)
Observations	226,936	226,936
Distance controls	Linear	Quadratic
Patient-provider-market-pair FEs	Yes	Yes
Additional distance elasticity at mean for high engagement: common procedures	-0.05	-0.13
Additional distance elasticity at mean for high engagement: rare procedures	-0.18	-1.34

Notes: This table reports estimates of a variant of equation (2.11), which adds interactions with indicators for whether a procedure is "rare" (provided less often than the median procedure) and for whether a procedure is "high engagement" (median number of distinct claims per patient for the procedure in a given year is above one) or low engagement. The unit of observation is {rare indicator, high-engagement indicator, exporting HRR, importing HRR}, and the dependent variable is the value of trade. Each column includes fixed effects for each ij pair of patient market and provider market, rare versus common procedures, and high- versus low-engagement procedures, plus indicators for three categories (common × high-engagement, rare × low-engagement, and rare × high-engagement) interacted with patient- and provider-market populations and distance covariates. Covariates for common × low-engagement procedures are omitted, since they would lead to collinearity with the ij fixed effects. Column 2 adds a control for the square of log distance and its interactions. The negative coefficient on provider-market population and the positive coefficient on patient-market population for common and high-engagement procedures indicate that the home-market effect is slightly less *strong* compared to common and low-engagement procedures, even though these effects are not all statistically different from zero. The positive coefficient on provider-market population × rare and the negative coefficient on patient-market population × rare for both high- and low-engagement procedures indicates that the *strong* home-market effect is stronger for rare services, whether they are high- or low-engagement. The distance elasticity is more negative for high-engagement procedures (both rare and common). Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 2.16: Gravity regression by procedure: individual procedures exhibit a strong home-market effect

	(1)	(2)	(3)	(4)	(5)	(6)
Procedure:	Colonoscopy	Cataract surgery	Brain tumor	Brain radiosurgery	LVAD	Colon removal
HCPCS code:	G0121	66982	61510	61798	33979	44155
Provider-market population (log)	0.515	0.466	0.928	1.148	1.251	0.992
	(0.0690)	(0.0729)	(0.0884)	(0.119)	(0.168)	(0.165)
Patient-market population (log)	0.351	0.436	0.191	0.165	0.181	-0.143
	(0.0692)	(0.0690)	(0.0726)	(0.0817)	(0.141)	(0.147)
Distance (log)	0.446	0.965	1.018	1.484	2.176	3.097
	(0.395)	(0.495)	(0.534)	(0.686)	(0.910)	(1.630)
Distance (log, squared)	-0.217	-0.270	-0.268	-0.304	-0.366	-0.500
	(0.0394)	(0.0491)	(0.0564)	(0.0697)	(0.0922)	(0.171)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.100	0.407	0.000	0.000	0.000	0.000
Observations	93,636	93,636	93,636	93,636	93,636	93,636
Distance elasticity at mean	-2.66	-2.90	-2.82	-2.87	-3.07	-4.07
Total count	58,798	43,604	1,922	752	333	112

Notes: This table reports estimates of equation (2.8) for procedure-level trade for six selected HCPCS codes, which vary in how common they are. For all procedures, the sample is all HRR pairs ($N = 306^2$). The dependent variable in all regressions is the value of trade in the procedure (computed using each procedure's national average price). The independent variables are patient- and provider-market log population, log distance and square of log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market. The bottom row reports the total national count of the procedure in our sample. Common procedures include screening colonoscopy (column 1) and cataract surgery (column 2). In a screening colonoscopy, the physician visualizes the large bowel with a camera to look for cancer. In a cataract surgery, the surgeon removes a cloudy lens from the eye to improve vision. Relatively rare procedures include brain radiosurgery (column 3), brain tumor removal (column 4), left ventricular assist device (LVAD) implantation (column 5) and colon removal (column 6). In brain radiosurgery, an area of the brain is irradiated, often to kill a tumor. In an LVAD implantation, a pump is implanted in the chest to assist a failing heart in pumping blood. Brain tumor and colon removals involve surgical removal of the respective structures.

Table 2.17: Scale elasticity estimates for CBSAs

Panel A: All services	Baseline	No Diagonal	Controls
OLS	0.888	1.052	0.907
	(0.009)	(0.017)	(0.010)
2SLS: population (log)	0.845	1.023	0.852
	(0.010)	(0.016)	(0.013)
2SLS: population (1940, log)	0.848	0.928	0.851
	(0.014)	(0.025)	(0.017)
2SLS: bedrock depth	0.810	0.762	0.812
	(0.038)	(0.099)	(0.044)
Panel B: Rare services			
OLS	0.941	1.108	0.945
	(0.010)	(0.028)	(0.011)
2SLS: population (log)	0.914	1.106	0.909
	(0.013)	(0.026)	(0.016)
2SLS: population (1940, log)	0.942	1.019	0.941
	(0.017)	(0.044)	(0.022)
2SLS: bedrock depth	0.814	0.095	0.807
	(0.063)	(0.393)	(0.078)

Notes: This table reports estimates of α from ordinary least squares (OLS) or two-stage least squares (2SLS) regressions of the form $\widehat{\ln \delta_i} = \alpha \ln Q_i + \ln R_i + \ln w_i + u_i$ using core-based statistical areas (CBSAs) as the geographic units. The dependent variable $\widehat{\ln \delta_i}$ is estimated in equation (2.12), Q_i is region i's total production for Medicare beneficiaries, R_i is Medicare's Geographic Adjustment Factor, the w_i covariate includes mean two-bedroom property value and mean annual earnings for non-healthcare workers, and u_i is an error term. In the rows labeled "2SLS" we instrument for $\ln Q_i$ using the specified instruments. The $\ln R_i$ and $\ln w_i$ controls are omitted in the columns labeled "no controls". In the columns labeled "no diag", Q_{ii} observations were omitted when estimating $\widehat{\ln \delta_i}$ in equation (2.12). Standard errors (in parentheses) are robust to heteroskedasticity.

Table 2.18: Classification of rare and common procedures in Medicare vs. private insurance data

Above median HCCI Above median CMS	0	1	total
0	82	18	100
1	12	88	100

Notes: This table compares the percentage of procedures classified as rare (above median frequency equals one) or common (above median frequency equals zero) in the public Medicare data versus the private insurance data from the Health Care Cost Institute (HCCI). Classifying procedures as rare versus common is consistent when using Medicare or privately insured data.

Table 2.19: Specialization earnings and frequency

Dependent variable: Per capita population elasticity	(1)	(2)	(3)
Number of physicians in specialization (log, national)	-0.0716		-0.0677
	(0.0139)		(0.0137)
Mean earnings (log)		-0.245	-0.174
		(0.0697)	(0.0543)
Observations	209	209	209
R-squared	0.199	0.050	0.223

Notes: This table reports estimates of a regression of per capita population elasticity of physician count on the national count of physicians and mean earnings. Each observation is an NPPES taxonomy code. Earnings (wage and business income) data from Gottlieb et al. (2020) are reported by Medicare specialty groups. We use a crosswalk to map Medicare specialty groups to NPPES taxonomy codes. The estimation sample excludes 11 taxonomy codes that are not mapped to any Medicare specialty. Standard errors (in parentheses) are robust to heteroskedasticity.

Table 2.20: Larger markets produce a greater variety of procedures

	(1) All	(2) Anesthesia	(3) E&M	(4) Medical	(5) Other	(6) Pathology	(7) Radiology	(8) Surgery
Population (log)	0.357	0.292	0.169	0.294	0.428	0.358	0.201	0.400
	(0.00736)	(0.0132)	(0.00528)	(0.00663)	(0.0115)	(0.0204)	(0.00610)	(0.00959)
Observations	306	306	306	306	306	306	306	306

Notes: This table reports the population elasticity of the number of distinct service codes produced in a region, estimated using Poisson pseudo-maximum likelihood (PPML). Column 1 shows the coefficient including all service types. The remaining columns show the coefficients for specific categories of service types. We use procedure classifications from the American Academy of Professional Coders, which groups codes into surgeries, anesthesia, radiology, pathology, medical, and evaluation & management ("E&M") services (AAPC, 2021). We combine Category II codes, Category III codes and Multianalyte Assays into "other."

CHAPTER 3

FIRMS, MARKETS, AND THE DIVISION OF LABOR: THE CASE OF PHYSICIANS

Abstract: Why and how do physicians co-locate to provide care? I establish novel facts regarding this question. First, the number of healthcare organizations grows with an elasticity near one with market size, so that a doubling of population results in twice as many healthcare organizations. Notably, the average size of healthcare organizations does not increase measurably with market size. I also show that the composition of organizations varies substantially with market size, even though they remain the same size. As market size grows, physicians co-locate more with same-specialty colleagues, individually produce a narrower set of services, and collectively produce a larger set and volume of services. These results suggest that coordination costs substantially constrain organization size. In addition, they imply that same-specialty colleagues become more valuable as the market size grows due to an increasingly fine division of labor, allowing for production efficiencies.

3.1 Introduction

The impact of team formation within firms and firms' organizational decisions on aggregate output, productivity, and labor market outcomes has received renewed attention recently (Boerma, Tsyvinski, and Zimin, 2021; Bonhomme, 2021; Adenbaum, 2022; Freund, 2022). Yet, empirically examining the role of the internal organizations of firms on aggregate market outcomes remains challenging due to the limited availability of detailed micro-level data at the worker-task level within firms. These questions are particularly relevant in healthcare, with the recent rise of health systems, and given its potential implications for individual provider quality and aggregate patient outcomes. In addition, studying such questions in the healthcare setting is particularly promising due to the availability of uniquely detailed micro-level data on individual provider billing information.

In this chapter, I investigate why and how physicians co-locate in firms to provide care. Using detailed Medicare data at the physician level linked to information on firms at which they practice, I document novel facts about the internal organization of healthcare firms. I show that the elasticity of the number of healthcare organizations to market size is about one, suggesting that an increase in population by 10% increases the number of healthcare organizations by the same percentage. The average size of healthcare organizations does not vary with market size. However, the boundaries of individual healthcare organizations appear to contract with market size, with a decrease in the number of physician specialties represented by firms as market size grows. These results are reminiscent to results from Garicano and Hubbard (2003, 2009) in the context of U.S. law firms. This suggests that healthcare organizations are on average more specialized in more populous regions, but also that same specialty colleagues become more valuable to doctors as market size grows.

Firms' labor composition varies with market size while firms' size remain constant on average. Previous work by Baumgardner (1988*b*) shows evidence that specialists tend to be located in larger cities. I re-produce these results using the universe of specialties and procedures billed to Medicare for each physician: physicians are more specialized in larger cities. Chapter 2 shows that there are large economies of scale in the production of healthcare services, and that one mechanism for it appears to be division of labor. Larger cities allow physicians to specialize more because of a larger patient base.

Taken together, this evidence suggests that firms and their composition play a role in allowing physicians to specialize *within* their specialty. I evaluate it directly in the data, by leveraging within-city cross-sectional variation in the firm size at which physicians practice. I find evidence of increasing returns, where the elasticity of a firm's specialty revenue to the specialty size at that firm is greater than one. I also find that physicians tend to be more specialized in specific tasks when organizations they practice at are larger in their specialty, suggesting that division of labor may be a driving force of healthcare firms' boundaries and their internal organization.

These findings are tightly related to the work related to knowledge-hierarchies within firms allowing for division of labor and worker specialization and the crucial role of referrals (Garicano, 2000; Garicano and Santos, 2004). Referrals are likely to play a crucial role in healthcare, and should be key determinants of a physician's ability to specialize on a specific set of procedures or diseases. Gathering under the same firm may allow physicians to credibly share such referrals and specialize, ultimately producing higher quality care. There is substantial evidence in the medical literature linking physicians' specialization to health outcomes, as outlined in Chapter 1 of this dissertation.

Understanding the role of healthcare firms in determining physician specialization has direct policy implications. First, the rise of health-systems, or multi-organizations firms in healthcare, raises concerns for competition policy. However, if referrals within- and between-organizations may allow physicians to specialize further, leading to better matches between a patient's diagnosis and their doctor's skills, health systems could ultimately improve health outcomes. There is an inherent trade-off between reduction in competition and efficiency of production, and this work aims as highlighting the second element of this trade-off in the context of healthcare.

Second, and related, the adoption of new technologies such as electronic medical records (EMRs) may make referrals less costly for patients and physicians. The literature has found a mixed impact of EMR adoption in healthcare. Agha (2014) finds little impact of health information technology adoption on quality of care using patient outcomes for hospitals. McCullough, Parente, and Town (2016) and Miller and Tucker (2011) show a positive impact of EMR adoption on patient outcomes for complex conditions requiring careful monitoring and care coordination. At the firm level, Lee, McCullough, and Town (2013) find a modest impact of expanded IT adoption on hospital productivity. Yet, McCullough and Snir (2010) examine the relationship between IT and physician-hospital integration where IT allows for monitoring, and find that physician-hospital integration and monitoring are complementary. They argue that this relationship may hinge on contracting costs. Such technologies can have an impact on the internal organization of firms and reduce the costs of inter-organizations referrals, allowing for more specialized organizations and incentivizing multi-organizations firms. This is likely to be driven by "superstar firms" (Hsieh and Rossi-Hansberg, 2023). Understanding how these new technologies led to changes in the internal organization of healthcare firms, and how different firms responded to these changes, may provide insightful avenues to better understand these different results.

Ongoing work on this Chapter includes both leveraging the richness of Medicare Research Identifiable Files and building a theoretical model endogenizing firms' internal organizational choices to market size where physician specialization is increasing in market size. Medicare claims data will allow to precisely assign each physician's activity to each specific organization, which the public use data does not allow to do. The model will allow for multi-organization firms to evaluate the health-systems and will include the costs of referrals to examine the role of new technologies such as EMRs.

3.2 The internal organization of healthcare firms

3.2.1 Data

I use data from Medicare, the health insurer for Americans aged 65 years old and older and the disabled. In addition to covering about 60 million beneficiaries every year, Medicare is largely accepted by physicians throughout the U.S., which allows to cover close to the universe of physicians in the country.

I match 2017 data from the Medicare provider utilization and payment data - physicians and other suppliers, which describes each procedure, identified by a HCPCS code, billed to Medicare in a given year by a physician identified by a unique identifier, to the doctors and clinicians CMS public use dataset, which includes physician-level information including healthcare organizations physicians work with in a given year.[1] Each dataset contains about 1 million unique individual identifiers in 2017. I complement this data with the National Plan and Provider Enumeration System (NPPES) dataset to obtain primary practice locations for physicians and healthcare entities as well as physician primary specializations.

The analysis sample is described in Table 3.1: it contains close to half a million physicians linked to about 52,000 unique healthcare organizations in 2017. A healthcare organization can include multi-establishment organizations, but it is not a proxy for ownership since multiple healthcare organizations may be owned by the same entity. I perform the analysis at the organization level since organizations likely constitute the appropriate scope for referrals between physicians, and I will call such organizations "firms".

There are two main limitations in using these public use datasets that will be overcome by using the Medicare Research Identifiable Files and additional external datasets. First, healthcare organizations are observed at the physician level; I cannot allocate a physician's activity to each specific healthcare organization. Consequently, the average healthcare organization size at the physician level is a simple average across their linked organizations, hereafter denoted as "firms", irrespective of the share of their activity at each of them. Using the Medicare Research Identifiable Files will allow us to precisely link a physician's activity to a healthcare organization, and even get at a more precise definition of a healthcare establishment, to ultimately overcome this limitation. Second, healthcare organizations in the doctors and clinicians CMS public use dataset likely include different types of organizations such as hospitals, doctors offices, and outpatient facilities. I am exploring external datasets to match the healthcare organization identifiers to external datasets with information on the

1. The information in the Medicare provider utilization and payment data - physicians and other suppliers comes from Medicare claims data, aggregated at the physician and HCPCS code level when a minimum of eleven unique patients were treated. The doctors and clinicians CMS public use dataset is sourced from the provider enrollment chain and ownership system (PECOS).

organization type.

3.2.2 Facts

I first investigate the elasticity of the number of healthcare organizations to market size. Healthcare organizations include different locations; I infer the primary location of an organization as the location at which they make the most medicare revenue or the location with the most physicians.[2]

Figure 3.1 shows the relationship between the number of healthcare establishments and market size for two alternative definitions of a market: hospital referral regions (HRRs) and core-based statistical areas (CBSAs). For both market definitions, the elasticity of the number of healthcare organizations to market size is about one: an increase in population by 10% increases the number of healthcare organizations by 10%.

Second, I examine how the average organization size varies with market size in Figure 3.2. The average healthcare organization size appears to be constant with market size. Figure 3.2 shows that markets above the 75th percentile in market sizes exhibit more large and very small organizations than markets below the 25th percentile in population. Overall, the distributions of healthcare organization sizes remains very similar across market sizes. This finding is consistent with transaction costs at the organization level restricting the size of organizations.

Third, I evaluate how organization boundaries respond to market size. I find that the average number of distinct physician specialties represented in an organization decreases with market size in Figure 3.3: organizations appear to be more specialized as market size grows, going from about 4 distinct specialties on average in smallest markets to about 2.5 distinct specialties on average in largest markets. Since average organization size remains

2. Note that more than 90% of organizations included in our analysis have all establishments located in the same HRR. The 10% that cross HRR boundaries are likely to be "superstar firms" in the spirit of Hsieh and Rossi-Hansberg (2023) as organizations spreading across space; they will be explored in future analyses.

constant with market size, this suggests that the value of same specialty colleagues increases with market size. I examine this directly in Figure 3.3, where the share of the organization in the same specialty at the physician level increases from about 35% in smallest HRRs to about 45% in largest HRRs.

Taken together, these results indicate that while the average size of healthcare organizations remains constant with market size, the composition of labor within these organizations, i.e. their internal organization, contracts to market size. Healthcare organizations are more specialized in larger markets, with fewer distinct specialties per organization and more same specialty colleague for physicians in these organizations.

3.2.3 Division of labor within the firm

Specialists tend to be located in larger markets, suggesting that larger markets allow for division of labor (Baumgardner, 1988*a*). I leverage the richness of the data to re-demonstrate these facts in our data including all specialties and a wider set of procedures. Figure 3.4 shows that the share of physicians that are specialists, i.e. not in Family Medicine, Internal medicine, or General Practictioners, increase significantly with market size, whether markets are defined as HRRs or CBSAs. Using physician-level measure of specialization, including the number of distinct procedures billed, the share in a physician's activity represented by their top five procedures, and Herfindahl-Hirschman index (HHI) measures using billed HCPCS codes, I show that physicians in larger markets are statistically significantly more specialized across all alternative measures of specializations in Tables 3.2 and 3.3.

Consistent with physicians wanting to specialize more in larger markets, healthcare organizations also appear to become more specialized as market size grows. The value of same specialty colleagues appears to be greater in larger markets, which may be driven by the need to divide labor further among physicians *within* specialties. Said differently, intra-specialty referrals appear to become more valuable as market size grows.

To investigate this hypothesis, I leverage within-city cross-sectional variation in organization sizes to evaluate the relationship between organization size and the specialization of an entire specialty in the organization. I run the following regression

$$\ln Y_{jsh} = \alpha + \beta_1 \ln \text{same spe size}_{js} + \beta_2 \ln \text{other spe size}_j + \lambda_s + \lambda_h + \epsilon_{jsh} \tag{3.1}$$

where Y_{jsh} is the output of firm j's specialty s in HRR h, $\text{same spe size}_{js}$ the number of physicians in that specialty at firm j, other spe size_j the number of physicians in other specialties at firm j, and λ_s and λ_h are specialty and market fixed effects respectively.

Results are reported in Tables 3.4 and 3.5. Increasing the size of a specialty within an organization significantly increases the number of distinct procedures produced by that specialty and increases the organization's specialization represented by the share represented by the top five procedures for this specialty at this firm. These results indicate that organizations with more physicians in a specialty are able to procedure a wider variety of procedures while also producing more of their most common procedures. Furthermore, they appear to achieve economies of scale, since the elasticity of the total Medicare revenue for that specialty in that firm to the size of the organization is statistically larger than one.

To tie this evidence to individual physician specialization, I run a similar regression at the physician-level, such that

$$\ln Y_{jsh} = \alpha + \mu_1 \ln \text{same spe size}_{js} + \mu_2 \ln \text{other spe size}_j + \delta X_j + \lambda_s + \lambda_h + \epsilon_{jsh} \tag{3.2}$$

where Y_{jsh} is the output of physician j in HRR h, $\text{same spe size}_{js}$ the number of physicians in the same specialty at firm j, other spe size_j the number of physicians in other specialties at firm j, δX_j are physician-level covariates, and λ_s and λ_h are specialty and market fixed effects respectively. Physician-level covariates include the total Medicare revenue billed by the physician, the number of organizations they are associated with, their number of hospital affiliations, and their medical school graduation year.

Tables 3.6 and 3.7 report results for this regression. Physicians working in organizations with more same specialty colleagues tend to specialize more: they produce a narrower set of procedures and produce more of their top five procedures. There is substantial heterogeneity in these results, as illustrated in Table 3.8 reporting results from the same regression but for specific specialties separately. While internal medicine, surgeons, and hospitalists tend to produce a narrower set of procedures when they have more same specialty colleagues, orthopedic surgeons and cardiac surgeons on the contrary tend to produce a wider set of procedures.

This heterogeneity result is interesting because there may be some heterogeneity across specialties depending in inter-specialties relationships. For example, internal medicine physicians practicing in a specialized firm may be able to focus better on their specialty, and avoid performing procedures that can be performed by other specialties, which may not be the case for a solo internal medicine physican in a doctor office. For orthopedic surgeons, who are particularly specialized, it is unlikely that they performe non-orthopedic procedures in small firms. However, having more colleagues in their specialty may allow them to share "routine" orthopedic care while freeing time for rarer orthopedic procedures, hence expanding the scope of their practice within their specialty, in line with results obtained at the firm's specialty level. More work is required to further examine these potential relationships across specialties, since I observe which specialties tend to co-locate in firms, and how this evolves as market size grows. These illustrate two potential gains from specialization for physicians that may have a different impact on patient outcomes.

Overall, specialties within firms appear to benefit from size: they produce a wider set of procedures, yet are more focused on their "specialty" procedures, and achieve economies of scale since the elasticity of size to total revenue is more than one. I find suggestive evidence that these gains are achieved through division of labor, since physicians working at larger firms in their specialty tend to be more specialized.

3.3 Conclusion

In this Chapter, I highlight new facts about the internal organization of healthcare firms. The elasticity of the number of healthcare organization to market size is about one and the average size of these healthcare organizations remains constant with market size. However, organizations buondaries contract with market size: the number of distinct physician specialties per healthcare organization decreases with market size. In other words, same specialty colleagues appear to become more valuable as market size grows.

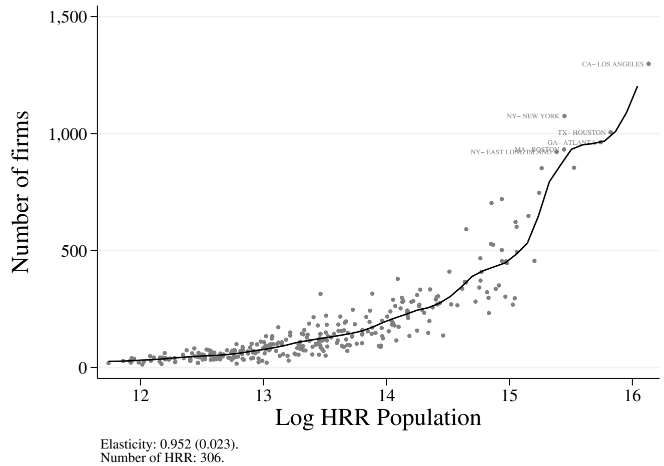
I find suggestive evidence that healthcare organizations facilitate division of labor, so that physicians prefer same specialty colleague in larger markets in which they want to specialize more *within* their specialty. Organizations with more physicians in a specialty produce a wider variety of procedures, produce more of their top procedures, and achieve economies of scale so that the elasticity of revenue to specialty size is larger than one. In parrallel, physicians with more colleague in their specialty tend to be more specialized Taken together, this evidence suggests that organizations facilitate division of labor, possibly through referrals (Garicano and Santos, 2004), which allows physicians in an organization to produce more specialized care.

There are several limitations to the current analysis that will be overcome with ongoing work using Medicare Research Identifiable Files. Medicare claims data will allow to assign each physicians' activity to each specific organization, but also will permit to examine the consequences of specialization and referrals on health outcomes. Additional work is needed to better understand the heterogeneity across different types of organizations, i.e. whether they include physician practices, outpatient centers, or hospitals. Finally, examining the role of new technologies is a promising avenue to isolate the role of referrals between physicians in the same organization.

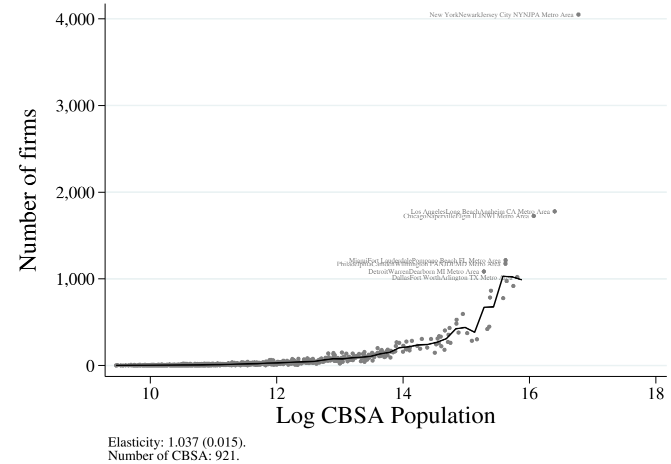
3.4 Figures and Tables

Figure 3.1: Number of healthcare organizations to market size

(a) HRRs



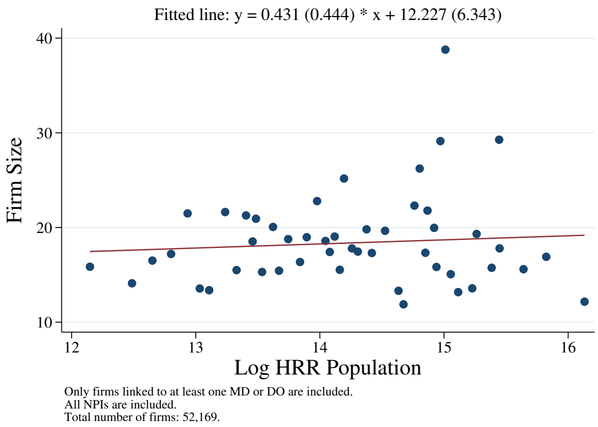
(b) CBSAs



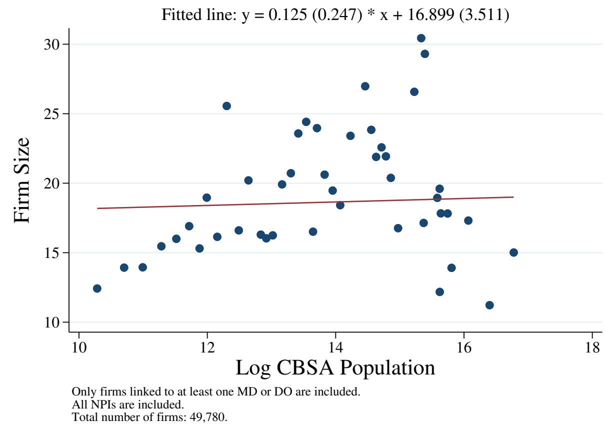
Notes: These graphs show how the number of healthcare organizations varies with market size, when markets are defined as HRRs or CBSAs. Market size corresponds to the population in the relevant market definition. The ZIP code of the organization corresponds to the ZIP code with associated physicians with the largest total Medicare revenue. HRR definitions are from the Dartmouth Atlas Project. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Figure 3.2: Average organization size constant with market size

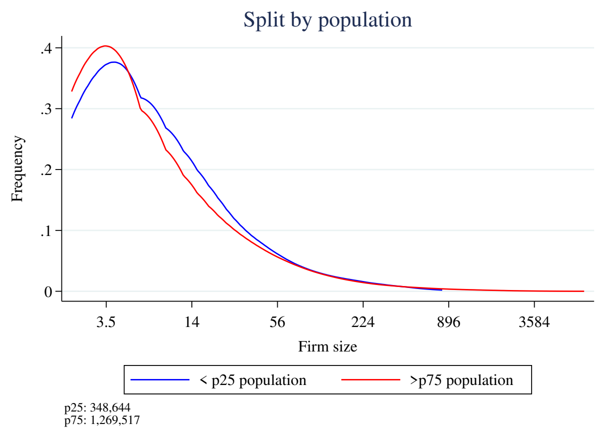
(a) HRRs



(b) CBSAs



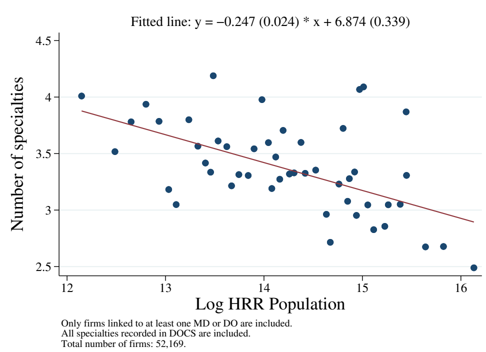
(c) Distribution by population (HRRs)



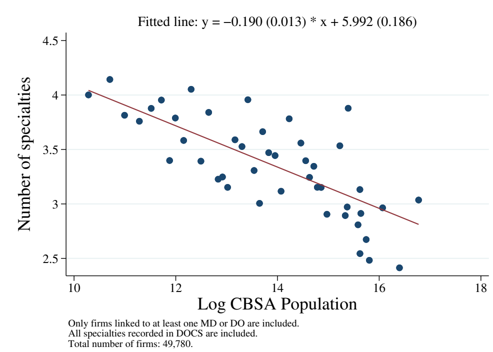
Notes: The two top graphs show that average organization size is constant with market size, whether markets are defined as HRRs or CBSAs. The graphs summarize this relationship using a binned scatter plot with fifty equally sized bins. The bottom graph compares the distribution of firm size for markets below the 25th percentile and above the 75th percentile in population using a kernel-weighted polynomial regression. Market size corresponds to the population in the relevant market definition. The ZIP code of the organization corresponds to the ZIP code with associated physicians with the largest total Medicare revenue. HRR definitions are from the Dartmouth Atlas Project. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Figure 3.3: Number of healthcare organizations to market size

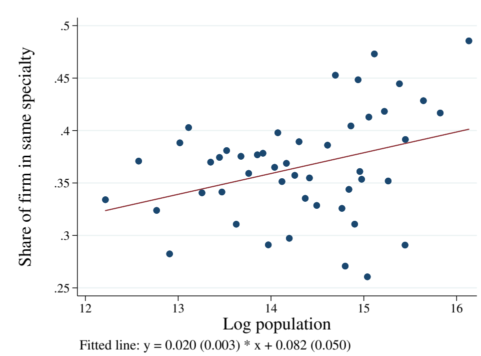
(a) Distinct specialties per firm, HRRs



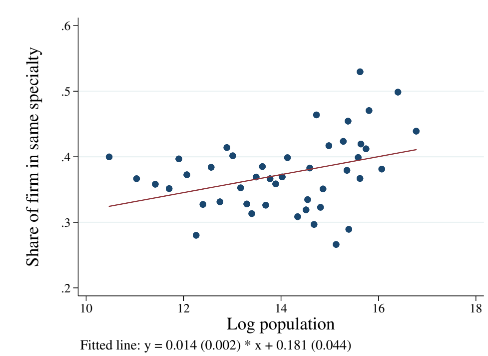
(b) Distinct specialties per firm, CBSAs



(c) Share of firm in same specialty, HRRs



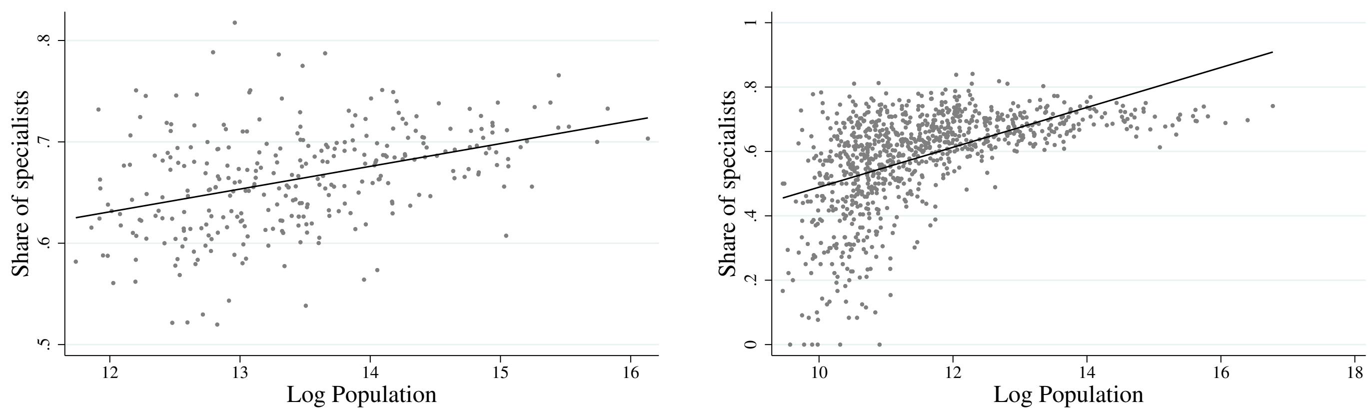
(d) Share of firm in same specialty, CBSAs



Notes: Top graphs show that the average number of distinct specialties represented by organization decreases with market size, for both HRRs and CBSAs. Bottom graphs show that the share of the firm in the same specialty as the physician increases wit market size, for both HRRs and CBSAs. All graphs summarize these relationships using a binned scatter plot with fifty equally sized bins. Market size corresponds to the population in the relevant market definition. The ZIP code of the organization corresponds to the ZIP code with associated physicians with the largest total Medicare revenue. The ZIP code of the physician corresponds to their primary practice ZIP code. HRR definitions are from the Dartmouth Atlas Project. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Figure 3.4: Share of specialists increases with market size

(a) HRRs (b) CBSAs



Pearson correlation: 0.396.
Fitted line: y = 0.022 (0.003) * x + 0.362 (0.040)
N = 306.

Pearson correlation: 0.522.
Fitted line: y = 0.062 (0.003) * x + −0.131 (0.039)
N = 926.

Notes: These graphs show that the share of specialists increases with market size, when markets are defined as HRRs or CBSAs. Specialists are defined as physicians that are not in internal medicine, general practicioners, or family medicine. Market size corresponds to the population in the relevant market definition. HRR definitions are from the Dartmouth Atlas Project. Physician locations correspond to the primary practice ZIP codes. Physician locations and specialties come from the 2017 National Plan and Provider Enumeration System (NPPES) data.

Table 3.1: Summary statistics analysis sample

	Physician level	Firm level
Distinct procedures	13.2	39.3
	(13.0)	(64.7)
Medicare revenue (x1,000)	105.2	1,153.7
	(227.8)	(4,304.1)
Share top 5 procedures	0.9	0.7
	(0.2)	(0.2)
Number of firms linked	1.1	-
	(0.7)	-
Firm size	394.6	18.4
	(772.9)	(95.0)
Firm size same specialty	29.4	-
	(58.9)	-
Distinct MD or DO specialties	24.7	2.8
	(31.0)	(5.9)
N	495,880.0	52,169.0

Notes: This table displays basic summary statistics for the analysis sample. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. The entirety of a physician's activity is assigned to each organization they are associated with. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Table 3.2: Physicians are more specialized in larger markets (HRRs)

	(1) Log distinct procedures	(2) Share top 5 procedures	(3) HHI revenue	(4) HHI quantity
Log population	-0.069	0.010	131.231	120.518
	(0.008)	(0.002)	(13.802)	(11.851)
Log income per capita	-0.015	0.002	8.471	13.525
	(0.004)	(0.001)	(8.564)	(13.060)
Log median schooling category	-0.863	0.106	-453.205	-663.419
	(0.269)	(0.035)	(644.525)	(735.348)
Percentage completed college	0.001	-0.000	10.597	11.533
	(0.002)	(0.000)	(5.935)	(6.655)
Percentage above 65 yo	0.014	-0.002	-22.101	-13.433
	(0.001)	(0.000)	(4.372)	(4.579)
Log land area	0.021	-0.004	-29.752	-30.322
	(0.003)	(0.001)	(6.936)	(5.796)
Constant	4.077	0.610	2,285.188	2,368.252
	(0.328)	(0.039)	(732.913)	(896.277)
Observations	579,435	579,429	579,435	579,435
R-squared	0.298	0.361	0.194	0.130
Specialty FE	Yes	Yes	Yes	Yes

Notes: This table shows that physicians are more specalized in larger markets. Physician locations correspond to the primary practice ZIP codes, and are mapped to total population, income per capita, median schooling category, percentage that completed college, percentage of population above 65 years old, and land area using the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. HRR definitions are from the Dartmouth Atlas Project. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets.

Table 3.3: Physicians are more specialized in larger markets (CBSAs)

	(1) Log distinct procedures	(2) Share top 5 procedures	(3) HHI revenue	(4) HHI quantity
Log population	-0.024	0.005	48.830	44.240
	(0.005)	(0.001)	(10.974)	(10.794)
Log income per capita	-0.325	0.045	564.833	683.022
	(0.032)	(0.007)	(102.681)	(84.136)
Log median schooling category	-0.488	0.043	-749.722	-889.329
	(0.224)	(0.042)	(552.361)	(611.941)
Percentage completed college	0.002	-0.000	5.824	3.274
	(0.002)	(0.000)	(4.431)	(5.077)
Percentage above 65 yo	0.013	-0.002	-24.480	-17.561
	(0.001)	(0.000)	(3.865)	(3.763)
Log land area	-0.005	-0.001	-0.616	-2.259
	(0.006)	(0.001)	(16.121)	(13.990)
Constant	6.348	0.304	-1,973.520	-3,100.581
	(0.385)	(0.059)	(945.286)	(946.492)
Observations	567,806	567,801	567,806	567,806
R-squared	0.304	0.366	0.197	0.132
Specialty FE	Yes	Yes	Yes	Yes

Notes: This table shows that physicians are more specalized in larger markets. Physician locations correspond to the primary practice ZIP codes, and are mapped to total population, income per capita, median schooling category, percentage that completed college, percentage of population above 65 years old, and land area using the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets.

Table 3.4: Benefits from size of a specialty within organizations (HRRs)

	(1) Log distinct procedures	(2) Log share top 5 procedures	(3) Log Medicare revenue
Log size same specialty	0.548	-0.094	1.145
	(0.003)	(0.001)	(0.004)
Log size other specialties	-0.043	0.005	-0.129
	(0.002)	(0.000)	(0.004)
Constant	2.397	-0.207	11.200
	(0.005)	(0.001)	(0.009)
Observations	128,516	128,516	128,516
R-squared	0.523	0.469	0.567
Specialty FE	Yes	Yes	Yes
HRR FE	Yes	Yes	Yes
Mean level outcome	22.6	.79	425825
Mean same spe size	4.1	4.1	4.1
R2 within	0.294	0.164	0.387

Notes: This table shows that organizations with more physicians in a specialty produce more distinct procedures, are more specialized, and achieve an elasticity of revenue to size larger than one. HRR definitions are from the Dartmouth Atlas Project. The ZIP code of the organization corresponds to the ZIP code with associated physicians with the largest total Medicare revenue. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Table 3.5: Benefits from size of a specialty within organizations (CBSAs)

	(1) Log distinct procedures	(2) Log share top 5 procedures	(3) Log Medicare revenue
Log size same specialty	0.546	-0.094	1.137
	(0.003)	(0.001)	(0.004)
Log size other specialties	-0.044	0.005	-0.134
	(0.002)	(0.000)	(0.004)
Constant	2.407	-0.208	11.248
	(0.005)	(0.001)	(0.010)
Observations	122,522	122,522	122,522
R-squared	0.536	0.479	0.578
Specialty FE	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes
Mean level outcome	22.8	.79	437592
Mean same spe size	4.2	4.2	4.2
R2 within	0.298	0.163	0.390

Notes: This table shows that organizations with more physicians in a specialty produce more distinct procedures, are more specialized, and achieve an elasticity of revenue to size larger than one. The ZIP code of the organization corresponds to the ZIP code with associated physicians with the largest total Medicare revenue. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Table 3.6: Physicians with more colleagues in their specialty are more specialized (HRRs)

	(1) Log distinct procedures	(2) Log share top 5 procedures
Log size same specialty	-0.013	0.003
	(0.001)	(0.000)
Log size other specialties	-0.005	0.003
	(0.000)	(0.000)
Constant	-2.721	1.300
	(0.143)	(0.046)
Observations	521,491	521,491
R-squared	0.779	0.547
Physician covariates	X	X
Specialty FE	Yes	Yes
HRR FE	Yes	Yes
Mean level outcome	14.2	0.8
Mean same spe size	33	33
R2 within	0.652	0.241

Notes: This table shows that physicians with more colleagues in their specialty are more specialized. HRR definitions are from the Dartmouth Atlas Project. Physician-level covariates include the total Medicare revenue billed by the physician, the number of organizations they are associated with, their number of hospital affiliations, and their medical school graduation year. Physician locations correspond to the primary practice ZIP codes. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Table 3.7: Physicians with more colleagues in their specialty are more specialized (CBSAs)

	(1) Log distinct procedures	(2) Log share top 5 procedures
Log size same specialty	-0.012	0.003
	(0.001)	(0.000)
Log size other specialties	-0.004	0.003
	(0.000)	(0.000)
Constant	-2.881	1.342
	(0.145)	(0.047)
Observations	502,314	502,314
R-squared	0.780	0.548
Physician covariates	X	X
Specialty FE	Yes	Yes
CBSA FE	Yes	Yes
Mean level outcome	14.2	0.8
Mean same spe size	34	34
R2 within	0.650	0.238

Notes: This table shows that physicians with more colleagues in their specialty are more specialized. Physician-level covariates include the total Medicare revenue billed by the physician, the number of organizations they are associated with, their number of hospital affiliations, and their medical school graduation year. Physician locations correspond to the primary practice ZIP codes. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

Table 3.8: The impact of firm size varies across specialties (HRRs)

Log distinct procedures	(1) IM	(2) Surgery	(3) Hospitalist	(4) Ortho Surgery	(5) Cardio
Log size same specialty	-0.101	-0.029	-0.067	0.050	0.010
	(0.002)	(0.004)	(0.006)	(0.004)	(0.003)
Constant	6.774	-4.889	-4.245	-4.329	-13.658
	(0.359)	(0.638)	(1.301)	(0.599)	(0.555)
Observations	93,674	17,050	8,105	15,021	22,718
R-squared	0.668	0.764	0.643	0.776	0.814
Physician covariates	X	X	X	X	X
Specialty FE	Yes	Yes	Yes	Yes	Yes
HRR FE	Yes	Yes	Yes	Yes	Yes
Mean level outcome	12.4	8.8	8.1	14.6	22.4
Mean same spe size	77	14	17	14	21
R2 within	0.636	0.733	0.573	0.737	0.784

Notes: This table shows that firm size has a different impact on physician output across specialties. Physician locations correspond to the primary practice ZIP codes. Physician-level information come from the 2017 Medicare provider utilization and payment data - physicians and other suppliers and National Plan and Provider Enumeration System (NPPES) datasets. Linkages of physicians to organizations and organization information comes from the 2017 doctors and clinicians CMS public use dataset.

REFERENCES

AAPC. 2021. "What Is CPT?" Available online at `https://www.aapc.com/resources/medical-coding/cpt.aspx`.

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality effects and choice across private health insurance plans." *The Quarterly Journal of Economics*, 136(3): 1557–1610.

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms." *Econometrica*, 67(2): 251–333.

Abowd, John M, Francis Kramarz, Paul Lengermann, and Sébastien Pérez-Duarte. 2004. "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States." *Unpublished Manuscript*, 5.

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *The Quarterly Journal of Economics*, 119(3): 1049–1090.

Adenbaum, Jacob. 2022. "Endogenous firm structure and worker specialization." Working Paper.

Adhvaryu, Achyuta, Vittorio Bassi, Anant Nyshadham, and Jorge A Tamayo. 2020. "No line left behind: Assortative matching inside the firm." National Bureau of Economic Research.

Agarwal, Sumit, J. Bradford Jensen, and Ferdinando Monte. 2020. "Consumer Mobility and the Local Structure of Consumption Industries." NBER Working Paper 23616.

Agha, Leila. 2014. "The effects of health information technology on the costs and quality of medical care." *Journal of Health Economics*, 34: 19–30.

Allen, Treb, Simon Fuchs, Sharat Ganapati, Alberto Graziano, Rocio Madera, and Judit Montoriol-Garriga. 2021. "Urban Welfare: Tourism in Barcelona."

Anderson, James E., Catherine A. Milot, and Yoto V. Yotov. 2014. "How much does geography deflect services trade? Canadian answers." *Int'l Econ. Rev.*, 55(3): 791–818.

Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3): 673–697.

Andrews, Martyn J, Leonard Gill, Thorsten Schank, and Richard Upward. 2012. "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias." *Economics Letters*, 117(3): 824–827.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.

Baicker, Katherine, and Amitabh Chandra. 2010. "Understanding Agglomerations in Health Care." *Agglomeration Economics*, 211–236. University of Chicago Press.

Bartelme, Dominick G., Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Textbook Case for Industrial Policy: Theory Meets Data." NBER NBER Working Paper 26193.

Bartik, Timothy, and George Erickcek. 2007. "Higher Education, the Health Care Industry, and Metropolitan Regional Economic Development: What Can 'Eds & Meds' Do for the Economic Fortunes of a Metro Area's Residents?" Upjohn Inst. Working Paper 08-140.

Battaglia, Emily. forthcoming. "The Effect of Hospital Closures on Maternal and Infant Health." *The Effect of Hospital Maternity Ward Closures on Maternal and Infant Health.*

Baumgardner, James R. 1988*a*. "Physicians' services and the division of labor across local markets." *Journal of Political Economy*, 96(5): 948–982.

Baumgardner, James R. 1988*b*. "Physicians' Services and the Division of Labor across Local Markets." *Journal of Political Economy*, 96(5): 948–982.

Berenson, Robert A., Jonathan H. Sunshine, David Helms, and Emily Lawton. 2015. "Why Medicare Advantage plans pay hospitals traditional Medicare prices." *Health Affairs*, 34(8): 1289–1295.

Berry, Christopher R., and Edward L. Glaeser. 2005. "The divergence of human capital levels across cities." *Papers in Regional Science*, 84(3): 407–444.

Birkmeyer, John D, Jonathan F Finks, Amanda O'reilly, Mary Oerline, Arthur M Carlin, Andre R Nunn, Justin Dimick, Mousumi Banerjee, and Nancy JO Birkmeyer. 2013. "Surgical skill and complication rates after bariatric surgery." *New England Journal of Medicine*, 369(15): 1434–1442.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2020. "Healthy business? Managerial education and management in health care." *Review of Economics and Statistics*, 102(3): 506–517.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin. 2021. "Sorting with Teams." *arXiv preprint arXiv:2109.02730*.

Bonhomme, Stéphane. 2021. "Teams: Heterogeneity, sorting, and complementarity." *arXiv preprint arXiv:2102.01802*.

Bonhomme, Stéphane, and Elena Manresa. 2015. "Grouped patterns of heterogeneity in panel data." *Econometrica*, 83(3): 1147–1184.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2023. "How much should we trust estimates of firm effects and worker sorting?" *Journal of Labor Economics*, 41(2): 291–322.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A distributional framework for matched employer employee data." *Econometrica*, 87(3): 699–739.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2022. "Discretizing unobserved heterogeneity." *Econometrica*, 90(2): 625–643.

Burstein, Ariel, Sarah Lein, and Jonathan Vogel. 2022. "Cross-border shopping: evidence and welfare implications for Switzerland."

Capps, Cory, David Dranove, and Mark Satterthwaite. 2003. "Competition and market power in option demand markets." *The RAND Journal of Economics*, 737–763.

Card, David, Alessandra Fenizia, and David Silver. 2023. "The health impacts of hospital delivery practices." *American Economic Journal: Economic Policy*, 15(2): 42–81.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly Journal of Economics*, 128(3): 967–1015.

Centers for Medicare and Medicaid Services. 2022. "National Health Expenditure Data."

Centers of Disease Control and Prevention. 2023. "Heart Disease Facts." https://www.cdc.gov/heartdisease/facts.htm, Accessed: 2023-05-30.

Chade, Hector, Jan Eeckhout, and Lones Smith. 2017. "Sorting through search and matching models in economics." *Journal of Economic Literature*, 55(2): 493–544.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016*a*. "Health care exceptionalism? Performance and allocation in the US health care sector." *American Economic Review*, 106(8): 2110–44.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016*b*. "Productivity dispersion in medicine and manufacturing." *American Economic Review*, 106(5): 99–103.

Chen, Yiqun. 2021. "Team-specific human capital and team performance: evidence from doctors." *American Economic Review*, 111(12): 3923–3962.

Chipman, John S. 1970. "External Economies of Scale and Competitive Equilibrium." *The Quarterly Journal of Economics*, 84(3): 347–385.

Clapesattle, Helen. 1969. *The Doctors Mayo.* Mayo Foundation for Medical Education.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Treatment Patterns and Patient Health?" *American Economic Review*, 104(4): 1320–1349.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2017. "In the Shadow of a Giant: Medicare's Influence on Private Payment Systems." *Journal of Political Economy*, 125(1): 1–39.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. 2018. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *The Quarterly Journal of Economics*, 134(1): 51–107.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams. 2019. "The More We Die, The More We Sell? A Simple Test of the Home-Market Effect." *The Quarterly Journal of Economics*, 134(2): 843–894.

Currie, Janet, and W Bentley MacLeod. 2017. "Diagnosing expertise: Human capital, decision making, and performance among physicians." *Journal of Labor Economics*, 35(1): 1–43.

Dafny, Leemore, Kate Ho, and Robin S Lee. 2019. "The price effects of cross-market mergers: theory and evidence from the hospital industry." *The RAND Journal of Economics*, 50(2): 286–325.

Davis, Donald R., and David E. Weinstein. 2003. "Market access, economic geography and comparative advantage: an empirical test." *J. of International Economics*, 59(1): 1–23.

Davis, Donald R., and Jonathan I. Dingel. 2020. "The comparative advantage of cities." *Journal of International Economics*, 123(C).

Davis, Donald R., Jonathan I. Dingel, Joan Monras, and Eduardo Morales. 2019. "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127(4): 1684–1738.

Deryugina, Tatyana, and David Molitor. 2021. "The Causal Effects of Place on Health and Longevity." *Journal of Economic Perspectives*, 35(4): 147–70.

Diamond, Rebecca. 2016. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *American Economic Review*, 106(3): 479–524.

Dingel, Jonathan I. 2017. "The Determinants of Quality Specialization." *Review of Economic Studies*, 84(4): 1551–1582.

Dingel, Jonathan I., and Felix Tintelnot. 2021. "Spatial Economics for Granular Settings." NBER Working Paper 27287.

Dingel, Jonathan I, Joshua D Gottlieb, Maya Lozinski, and Pauline Mourot. 2023. "Market size and trade in medical services." National Bureau of Economic Research.

Disdier, Anne-Célia, and Keith Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics*, 90(1): 37–48.

Dorsey, E. Ray, and Eric J. Topol. 2016. "State of Telehealth." *New England Journal of Medicine*, 375(2): 154–161. PMID: 27410924.

Dranove, David, and Christopher Ody. 2016. "Evolving measures of provider market power." *Am. J. of Health Econ.*, 2(2): 145–160.

Dranove, David, and Mark A. Satterthwaite. 2000. "The Industrial Organization of Health Care Markets." *Handbook of Health Economics*, 1: 1093–1139.

Dranove, David, Mark Shanley, and Carol Simon. 1992. "Is Hospital Competition Wasteful?" *The RAND Journal of Economics*, 23(2): 247–262.

Dubin, Jeffrey A, and Daniel L McFadden. 1984. "An econometric analysis of residential electric appliance holdings and consumption." *Econometrica: Journal of the Econometric Society*, 345–362.

Eckert, Fabian, Sharat Ganapati, and Conor Walsh. 2020. "Skilled Scalable Services: The New Urban Bias in Economic Growth." CESifo Working Paper Series 8705.

Eeckhout, Jan. 2018. "Sorting in the labor market." *Annual Review of Economics*, 10: 1–29.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying sortingâĂŤin theory." *The Review of Economic Studies*, 78(3): 872–906.

Einav, Liran, Amy Finkelstein, and Heidi Williams. 2016. "Paying on the margin for medical care: Evidence from breast cancer treatments." *American Economic Journal: Economic Policy*, 8(1): 52–79.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2022. "Producing Health: Measuring Value Added of Nursing Homes."

Ericson, Keith Marzilli, and Amanda Starc. 2015. "Measuring consumer valuation of limited provider networks." *American Economic Review*, 105(5): 115–19.

Falcettoni, Elena. 2021. "The Determinants of Physicians' Location Choice: Understanding the Rural Shortage." Mimeo, Federal Reserve.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." *QJE*, 131(4): 1681–1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review*, 111(8): 2697–2735.

Fischer, Stefanie J., Heather Royer, and Corey D. White. 2022. "Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the US." NBER Working Paper 30141.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. Lee Lucas, and Etoile L. Pinder. 2003*a*. "The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care." *Ann. Intern. Med.*, 138(4): 273–287.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. Lee Lucas, and Etoile L. Pinder. 2003*b*. "The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care." *Annals of Internal Medicine*, 138(4): 288–298.

Flexner, Abraham. 1910. "Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching." Carnegie Fnd. Bulletin 4.

Freund, Lukas. 2022. "Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities." *Available at SSRN 4312245.*

Fujita, Masahisa, and Paul Krugman. 1995. "When is the economy monocentric?: von Thünen and Chamberlin unified." *Regional Science and Urban Econ.*, 25(4): 505–528.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of political economy*, 108(5): 874–904.

Garicano, Luis, and Tano Santos. 2004. "Referrals." *American Economic Review*, 94(3): 499–525.

Garicano, Luis, and Thomas N Hubbard. 2003. "Firms' boundaries and the division of labor: Empirical strategies." *Journal of the European Economic Association*, 1(2-3): 495–502.

Garicano, Luis, and Thomas N Hubbard. 2009. "Specialization, firms, and markets: The division of labor within and between law firms." *The Journal of Law, Economics, & Organization*, 25(2): 339–371.

Garthwaite, Craig, Christopher Ody, and Amanda Starc. 2022. "Endogenous quality investments in the us hospital market." *Journal of Health Economics*, 102636.

Gaynor, Martin, and William B Vogt. 2003. "Competition among hospitals." *The RAND Journal of Economics*, 34(3): 764–785.

Gaynor, Martin S., Samuel A. Kleiner, and William B. Vogt. 2013. "A structural approach to market definition with an application to the hospital industry." *The Journal of Industrial Economics*, 61(2): 243–289.

Ghaferi, Amir A, John D Birkmeyer, and Justin B Dimick. 2009. "Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients." *Annals of Surgery*, 250(6): 1029–1034.

Glaeser, Edward L. 2005. "Reinventing Boston: 1630–2003." *Journal of Economic Geography*, 5(2): 119–153.

Gottlieb, Daniel J, Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S Skinner, and Jason M Sutherland. 2010. "Prices donâĂŹt drive regional Medicare spending variations." *Health Affairs*, 29(3): 537–543.

Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2020. "Who Values Human Capitalists' Human Capital? Healthcare Spending and Physician Earnings." University of Chicago, mimeo.

Gotto, Antonio M., and Jennifer Moon. 2016. *Weill Cornell Medicine: A History of Cornell's Medical School.* Ithaca: Cornell University Press.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2015. "Mergers when prices are negotiated: Evidence from the hospital industry." *American Economic Review*, 105(1): 172–203.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings." *Journal of Political Economy*, 118(3): 536–598.

Gupta, Atul. 2021. "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program." *American Economic Review*, 111(4): 1241–83.

Gupta, Atul, Amol Navathe, and Aaron Schwartz. 2022. "How Does Medicare Advantage Affect Health Care Use? Evidence from Beneficiary Transitions." University of Pennsylvania, mimeo.

Hanson, Gordon H., and Chong Xiang. 2004. "The Home-Market Effect and Bilateral Trade Patterns." *American Economic Review*, 94(4): 1108–1129.

Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade.* MIT Press.

Ho, Kate, and Robin S. Lee. 2017. "Insurer competition in health care markets." *Econometrica*, 85(2): 379–417.

Ho, Kate, and Robin S. Lee. 2019. "Equilibrium provider networks: Bargaining and exclusion in health care markets." *American Economic Review*, 109(2): 473–522.

Ho, Katherine. 2009. "Insurer-Provider Networks in the Medical Care Market." *American Economic Review*, 99(1): 393–430.

Hoxby, Caroline M. 2014. "What Lies Ahead for America's Children and Their Schools." Chapter Covering the Costs. Chester E. Finn & Richard Sousa.

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2021. "The Industrial Revolution in Services." Center for Economic Studies, U.S. Census Bureau Working Papers 21-34.

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2023. "The industrial revolution in services." *Journal of Political Economy Macroeconomics*, 1(1): 3–42.

Huckman, Robert S, and Gary P Pisano. 2006. "The firm specificity of individual performance: Evidence from cardiac surgery." *Management Science*, 52(4): 473–488.

Hull, Peter. 2018. "Estimating hospital quality with quasi-experimental data." *Available at SSRN 3118358.*

Jackson, C Kirabo. 2013. "Match quality, worker productivity, and worker mobility: Direct evidence from teachers." *Review of Economics and Statistics*, 95(4): 1096–1116.

Jensen, J. Bradford, and Lori G. Kletzer. 2005. "Tradable Services: Understanding the Scope and Impact of Services Offshoring." *Brookings Trade Forum*, 75–116.

Jochmans, Koen, and Martin Weidner. 2019. "Fixed-effect regressions on network data." *Econometrica*, 87(5): 1543–1560.

Johnston, Maximilian J, Sonal Arora, Dominic King, George Bouras, Alex M Almoudaris, Rachel Davis, and Ara Darzi. 2015. "A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery." *Surgery*, 157(4): 752–763.

Kessler, Daniel P., and Mark B. McClellan. 2000. "Is hospital competition socially wasteful?" *The Quarterly Journal of Economics*, 115(2): 577–615.

Khoury, Stephanie, Jonathan M. Leganza, and Alex Masucci. forthcoming. "Health Professional Shortage Areas and Physician Location Decisions." *Am. J. of Health Econ.*

Kline, Patrick, and Enrico Moretti. 2013. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics*, 129(1): 275–331.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out estimation of variance components." *Econometrica*, 88(5): 1859–1898.

Kolstad, Jonathan T. 2013. "Information and quality when motivation is intrinsic: Evidence from surgeon report cards." *American Economic Review*, 103(7): 2875–2910.

Kolstad, Jonathan T, and Michael E Chernew. 2009. "Quality and consumer decision making in the market for health insurance and health care services." *Medical Care Research and Rreview*, 66(1_suppl): 28S–52S.

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review*, 70(5): 950–59.

Lee, Jinhyung, Jeffrey S McCullough, and Robert J Town. 2013. "The impact of health information technology on hospital productivity." *The RAND Journal of Economics*, 44(3): 545–568.

Lee, Sanghoon. 2010. "Ability sorting and consumer city." *Journal of Urban Economics*, 68(1): 20–33.

Levy, Antoine, and Jacob Moscona. 2020. "Specializing in Density: Spatial Sorting and the Pattern of Trade." Mimeo, MIT. Available online at `https://economics.mit.edu/files/16986`.

Lewis, Matthew S., and Kevin E. Pflum. 2015. "Diagnosing hospital system bargaining power in managed care networks." *American Economic Journal: Economic Policy*, 7(1): 243–74.

Lewis, Matthew S., and Kevin E. Pflum. 2017. "Hospital systems and bargaining power: Evidence from out-of-market acquisitions." *The RAND Journal of Economics*, 48(3): 579–610.

Lipsey, Robert E. 2009. "Measuring International Trade in Services." In *International Trade in Services and Intangibles in the Era of Globalization.* , ed. Marshall B. Reinsdorf and Matthew J. Slaughter, 27–74. University of Chicago Press.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2016. "Matching, sorting and wages." *Review of Economic Dynamics*, 19: 63–87.

Lopez, Eric, and Gretchen Jacobson. 2020. "How Much More Than Medicare Do Private Insurers Pay? A Review of the Literature."

Marshall, Alfred. 1890. *Principles of Economics.* London: Macmillan and Co.

McCullough, Jeffrey S, and Eli M Snir. 2010. "Monitoring technology and firm boundaries: Physician–hospital integration and technology utilization." *Journal of Health Economics*, 29(3): 457–467.

McCullough, Jeffrey S, S Parente, and Robert J Town. 2016. "Health information technology and patient outcomes: The role of organizational and informational complementarities." *The RAND Journal of Economics*, 47(1): 07–236.

McDermott, Kimberly W, and Lan Liang. 2021. "Overview of operating room procedures during inpatient stays in US Hospitals, 2018: statistical Brief# 281."

Metcalfe, Robert D, Alexandre B Sollaci, and Chad Syverson. 2023. "Managers and Productivity in Retail." National Bureau of Economic Research.

Miller, Amalia R, and Catherine E Tucker. 2011. "Can health care information technology save babies?" *Journal of Political Economy*, 119(2): 289–324.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J. Redding. 2021. "The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data." NBER Working Paper 28497.

Moretti, Enrico. 2004. "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *J. of Econometrics*, 121(1-2): 175–212.

Moretti, Enrico. 2011. "Local labor markets." In *Handbook of Labor Economics*. Vol. 4, 1237–1313. Elsevier.

Morris, Carl N. 1983. "Parametric empirical Bayes inference: theory and applications." *Journal of the American Statistical Association*, 78(381): 47–55.

Muñoz, Mathilde. 2022. "Trading Non-Tradables: The Implications of Europe's Job Posting Policy."

Mukamel, Dana B, David L Weimer, and Alvin I Mushlin. 2006. "Referrals to High-Quality Cardiac Surgeons: Patients' Race and Characteristics of Their Physicians." *Health Services Research*, 41(4p1): 1276–1295.

Munoz, Pablo, and Cristobal Otero. 2023. "Managers abd public hospital performance." *Working Paper*.

Neufeld, Jonathan D., and Charles R. Doarn. 2015. "Telemedicine Spending by Medicare: A Snapshot from 2012." *Telemed J E Health*, 21(8): 686–693.

Newhouse, Joseph P. 1990. "Geographic access to physician services." *Annual Review of Public Health*, 11(1): 207–230.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*a*. "Does the Geographical Distribution of Physicians Reflect Market Failure?" *The Bell Journal of Economics*, 13(2): 493–505.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*b*. "Does the Geographical Distribution of Physicians Reflect Market Failure?" *The Bell Journal of Economics*, 13(2): 493–505.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*c*. "The geographic distribution of physicians: Is the conventional wisdom correct?" Santa Monica: RAND Corp. Publ. No. R-2734.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*d*. "The geographic distribution of physicians: Is the conventional wisdom correct?" Santa Monica: RAND Corp. Publ. No. R-2734.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*e*. "Where have all the doctors gone?" *JAMA*, 247(17): 2392–2396.

Petek, Nathan. 2022. "The marginal benefit of hospitals: Evidence from the effect of entry and exit on utilization and mortality rates." *Journal of Health Economics*, 86: 102688.

Rosenblatt, R. A., and L. G. Hart. 2000. "Physicians and rural America." *The Western Journal of Medicine*, 173(5): 348–351.

Rosenthal, Meredith B., Alan Zaslavsky, and Joseph P. Newhouse. 2005*a*. "The Geographic Distribution of Physicians Revisited." *Health Services Research*, 40(6p1): 1931–1952.

Rosenthal, Meredith B., Alan Zaslavsky, and Joseph P. Newhouse. 2005*b*. "The Geographic Distribution of Physicians Revisited." *Health Services Research*, 40(6p1): 1931–1952.

Rosenthal, Stuart S., and William C. Strange. 2008. "The attenuation of human capital spillovers." *Journal of Urban Economics*, 64(2): 373–389.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. "IPUMS USA: Version 12.0." Minneapolis, MN: IPUMS.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Silber, Jeffrey H, Sankey V Williams, Henry Krakauer, and J Sanford Schwartz. 1992. "Hospital and patient characteristics associated with death after surgery: a study of adverse occurrence and failure to rescue." *Medical Care*, 615–629.

Silver, David, and Jonathan Zhang. 2022. "Impacts of Basic Income on Health and Economic Well-Being: Evidence from the VA's Disability Compensation Program." NBER Working Paper 29877.

Skinner, Lucy, Douglas O. Staiger, David I. Auerbach, and Peter I. Buerhaus. 2019. "Implications of an aging rural physician workforce." *NEJM*, 381(4): 299–301.

Small, Kenneth A, and Harvey S Rosen. 1981. "Applied welfare economics with discrete choice models." *Econometrica*, 105–130.

Town, Robert, and Gregory Vistnes. 2001. "Hospital competition in HMO networks." *Journal of Health Economics*, 20(5): 733–753.

Ward, Sarah T, Justin B Dimick, Wenying Zhang, Darrell Campbell, and Amir A Ghaferi. 2019. "Association between hospital staffing models and failure to rescue." *Annals of Surgery*, 270(1): 91.

Wilson, Chad T, Steven Woloshin, and Lisa M Schwartz. 2007. "Choosing where to have major surgery: who makes the decision?" *Archives of Surgery*, 142(3): 242–246.

Zuckerman, Stephen, Laura Skopec, and Joshua Aarons. 2021. "Medicaid Physician Fees Remained Substantially Below Fees Paid By Medicare In 2019." *Health Affairs*, 40(2): 343–348.

Zwischenberger, Brittany A, Oliver K Jawitz, and Jennifer S Lawton. 2021. "Coronary surgery in women: how can we improve outcomes." *JTCVS techniques*, 10: 122–128.