Should Top Surgeons Practice at Top Hospitals? Sorting and Complementarities in Healthcare*

Please click here for latest version.

Pauline Mourot

Chicago Booth

October 2023

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 effect 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. Surgeon sorting across hospitals has large consequences for aggregate patient outcomes. 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.

*I am highly indebted to my advisors Joshua Gottlieb, Neale Mahoney, Matthew Notowidigdo, and Stéphane Bonhomme for their guidance and support. This paper benefited from conversations with Zarek Brot-Goldberg, Olivia Bordeu-Gazmuri, Jonathan Dingel, Alex Frankel, Arjun Gopinath, Luis Garicano, Anders Humlum, Thibaut Lamadon, Jacob Leshno, Magne Mogstad, Jack Montjoy, Karthik Srinivasan, Chad Syverson, Walter Zhang, and seminar participants at the Booth School of Business, the Economics Department, and the Harris School of Public Policy at the University of Chicago. All errors are my own.
Author's email address: pmourot@chicagobooth.edu.

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 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 effect 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, I can address identification issues for sorting which have been outlined in two-way fixed effect models when using worker earnings (Eeckhout and Kircher, 2011).

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 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 effect model. I

[1] See for example Chade, Eeckhout, and Smith (2017), Eeckhout (2018) and Card et al. (2018) for reviews of these literatures.

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

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, positive assortative matching is large 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. 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 evaluate the robustness of my results to alternative classifications for surgeons and hospitals. I find that my results are robust to alternative numbers of groups. I also examine alternative classification strategies, either including more moments to the k-means algorithm or relying on other classification methods such as simply grouping surgeons and hospitals in quintiles of risk-adjusted survival. Results are similar across these alternative classification approaches.

Using partial equilibrium reallocation exercises, I show that surgeon sorting across hospitals has a large impact on aggregate patient survival. 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 performances are hospital-specific. They show that surgeon performances are 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. Assuming that markets clear, the direction of sorting maps into the existence of complementarities.[3] Relaxing this assumption, 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. This paper has two main contributions to this literature. First, it documents positive assortative matching in a specific yet important labor market on a measure of output directly. Second, it shows the importance of including interactions to both better understand the relative role of workers and firms, but also to quantify the impact of sorting on aggregate output.

[3]In the traditional two-way fixed effect 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 may not be monotonic in firms' productivity in the presence of frictions, hence preventing from identifying sorting from wage data alone.

The rest of the paper is organized as follows. I describe the institutional setting and provide an overview of the data in Section 2. I delineate the empirical strategy in Section 3. Section 4 evaluates the sensibility of estimated parameters. I detail the imperfect substitutability and sorting results and assess their robustness in Section 5. Finally, I quantify the impact of surgeon sorting across hospitals on aggregate patient survival using partial equilibrium reallocation exercises in Section 6. Section 7 concludes.

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.

2.1 Institutional Setting

2.1.1 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.

2.1.2 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 A.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.

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 A.1.

2.2 Data

2.2.1 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 A.2. I am able to match 90% of the CABG surgeries from the 20% sample Carrier file as reported in Table 1.

I restrict the sample in four main dimensions as reported in Table 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.[4] Patient covariates are reported in Table 2. The

[4]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

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.

2.2.2 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 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 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 B.1. 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.

independent.

Multi-homing happens for a susbtantial share of a surgeon's activity. Appendix Table B.2 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.

2.2.3 Motivating facts

The variation in patient survival across surgeons and hospitals separately is quantitatively important for CABG surgery. In Figure 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 A.4. I compute the risk-adjusted survival at the patient level using a logit model including patient observables as delineated in Appendix A.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.[5]

Note that the averages across hospitals include the impact of surgeons and vice versa in Figure 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.

3 Empirical strategy

I leverage the existence of multi-homers and traditional movers in a two-way fixed effect 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

[5] Without measurement error adjustment, the variation across hospitals and surgeons is larger and larger for surgeons, as delineated in Appendix Table B.3. 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.

of complementarities between surgeons and hospitals, using a non-parametric production function of survival in the second step. Assuming selection on observables, the patient-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.

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 $E[\epsilon_{ijht}|\alpha_j, \psi_h, X_{it}] = 0$, the conditional expectation of patient survival is equal to the production function g.

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

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.

$$\frac{\partial^2 g(\alpha_j, \psi_h, \bar{X})}{\partial \alpha_j \partial \psi_h} \tag{1}$$

When equation (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) 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) is equal to zero, the contributions of surgeons and hospitals to the production function are independent.

Figure 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 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 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 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

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

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 $D_{ijht} = 1$:

$$\begin{aligned} E[Y_{ijht}|D_{ijht} = 1, \alpha_j, \psi_h, X_{it}] &= E[Y^*_{ijht}|D_{ijht} = 1, \alpha_j, \psi_h, X_{it}] \\ &= g(\alpha_j, \psi_h, X_{it}) + E[\epsilon_{ijht}|D_{ijht} = 1, \alpha_j, \psi_h, X_{it}] \end{aligned} \tag{2}$$

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.[6] I address this issue by grouping surgeons and hospitals in a first step in the

[6] Noise in the individual fixed effects also magnifies the bias on the sorting parameter, called limited

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 $E[\epsilon_{ijht}|D_{ijht}, \alpha_j, \psi_h, X_{it}] = E[\epsilon_{ijht}|\alpha_j, \psi_h, X_{it}] = 0$ in equation (2). This is the exogenous network assumption, common in two-way fixed effect 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 3.3.

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.[7] 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 effect 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 providers' moments that identify individual provider types. Using provider-level average survival to identify the individual provider types requires a provider's average survival to be increasing in the individual provider type. To see what this assumption implies, note that the average survival

mobility bias, which has been shown to be quantitatively important in the literature (Kline, Saggio, and Sølvsten, 2020; Bonhomme, Lamadon, and Manresa, 2022).

[7]Theoretical properties of the two-step grouped fixed effect 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.

at hospital h in the population can be written as

$$E[Y^*_{ijht}|\psi_h] = \int \underbrace{g(\alpha_j, \psi_h, X_{it})}_{\text{Production function}} \underbrace{f(\alpha_j|\psi_h)}_{\text{Sorting}} d\alpha_j \tag{3}$$

where $g(\alpha_j, \psi_h, X_{it})$ 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

$$E[Y^*_{ijht}|\psi_h] = \psi_h + E[\alpha_j|\psi_h]$$

With positive assortative matching, $E[\alpha_j|\psi_h]$ 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, $E[\alpha_j|\psi_h]$ 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 risk-adjusted survival, which prevents from separating their individual types from their average risk-adjusted survival. Consequently, identification of the individual quality from individual moments requires that these individual moments are increasing in the provider's quality. 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 obatined 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 A.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 A.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.[8]

Grouping surgeons and hospitals also addresses biases in the variance and covariance estimates in two-way fixed effect models (Bonhomme, Lamadon, and Manresa, 2019, 2022). This bias, called limited mobility bias (Abowd et al., 2004; Andrews et al., 2008, 2012), is

[8]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 an arbitrage 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).

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.

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

$$g(\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}) = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} \tag{4}$$

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) 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 retrieve complementarities between surgeon and hospital quality on the average patient.

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 $\alpha_{l(j)}$, $\psi_{k(h)}$, $\kappa_{l(j)k(h)}$:

$$\epsilon_{ijht} \perp D_{ijht} | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, \ \forall i, j, h, t \tag{5}$$

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:

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t] = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t \quad (6)$$

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 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

$$\begin{aligned} D_{ijht} &= 1\{u_{ih} \geq u_{ih'}, \ \forall h'\} \\ \text{with } u_{ih} &= \delta_h - \tau \ln(d_{ih}) + \eta_{ih} \end{aligned} \quad (7)$$

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.[9]

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

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

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

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.

Part $\phi_s(\eta_{is} - \mu_\eta)$ is hospital-specific and captures selection into a specific hospital. 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, $\phi_s > 0$.

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 goes to hospital h. Following Dubin and McFadden (1984), the observed survival conditional on the choice of hospital and all observed covariates $\alpha_{l(j)}, \psi_{k(h)}, X_{it}, \ln d_{i1}, ..., \ln d_{iH}$ is such that

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

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

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

[9] 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.

and $\hat{p}_{is}$ is the predicted probability for patient i to choose hospital s from the demand model in equation (7). Exact derivations are included in Appendix A.7. Note that the control function is positive when $s = h$ but negative otherwise since $\ln \hat{p}_{is} < 0$ with $0 < \hat{p}_{is} < 1$.

I estimate the demand model market by market, using hospital referral regions (HRRs) as market definitions.[10] 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:

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t + \sum_{s \in \mathcal{H}} \phi_s \hat{\theta}_{is}(h) + \varphi \hat{\theta}_{ih}(h) \tag{8}$$

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, or $cov(\epsilon_{ijht}, d_{ih}) = 0$, $\forall i, j, h, t$. 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.[11] 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 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 treatment is probably limited in this setting, I do not address it directly. As detailed in Appendix A.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.

[10] 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.

[11] 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.

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.

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 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 5.3.

There exist a substantial amount of variation in survival across groups. Figure 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 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 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, 30% of surgeons are observed at more than one hospital *group*, as depicted in Appendix Figure B.1. Yet, the number of surgeons observed at multiple hospital groups remains large.

4.2 Estimated parameters are sensible

Risk-adjustment parameters. The coefficients on patient observables for risk adjustment are sensible. Table 3 reports coefficients on patient covariates from estimating equations (6) and (8). 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 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.[12] 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 B.6.

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 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 (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 B.7. 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

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

specifications.

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 B.8, 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 B.6, 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 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 7 and Appendix Table B.9, 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 B.2 and Appendix Table B.9. 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.[13] Overall, patient selection into higher-survival providers appears to be minimal in this setting.

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 mecha-

[13]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.

nism 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.

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 8a, similarly to Figure 3 in Section 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 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 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.[14]

[14] 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.

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-days 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 survival for their patients varies widely with the quality of the hospital they operate at. These facts are illustrated in Figure 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

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \gamma_t] = \alpha_{l(j)} + \psi_{k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t \tag{9}$$

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

$$Var(Y_{ijht} - \sum_p \hat{\beta}_p X_{it,p} - \hat{\gamma}_t - \hat{\epsilon}_{ijht}) = Var(\hat{\alpha}_{l(j)}) + Var(\hat{\psi}_{k(h)}) + 2 \times cov(\hat{\alpha}_{l(j)}, \hat{\psi}_{k(h)}) \tag{10}$$

Table 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 B.10. 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.

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 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 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 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 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 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.

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 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 B.3. 21% of surgeries are performed outside of the patient's HRR.[15] The relationship between the chosen hospital and distance also appears log-linear, supporting the functional form assumption in equation (7).

I examine whether the distance to the hospital is predictive of hospital choice in Panel 13a of Figure 13. This figure depicts the relationship between the predicted probabilities to choose a hospital from the hospital choice model depicted in equation (7) estimated for each HRR separately and the distance between the patient ZIP code of residence and the hospital ZIP code. Appendix Figure B.4 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 B.11, the coefficient on the logarithm of distance is relatively stable with and without covariates, lending support for the exclusion restriction.

Second, I examine the stability of distance parameters in the demand model described

[15]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).

by equation (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 13b of Figure 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.[16]

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 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 B.5. 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 B.6. While there is no systematic 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 14, surgeon groups and hospitals groups are imperfect substitutes in the production function of survival. Estimated slopes in column (2) of Table 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

[16]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.

of a model without interactions between surgeon and hospital groups also leads to similar conclusions. As reported in Table 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[17]

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}] = \frac{\exp\left(\alpha_{l(j)} + \psi_{k(h)} + \sum_s \beta_s X_{it,s}\right)}{1 + \exp\left(\alpha_{l(j)} + \psi_{k(h)} + \sum_p \beta_p X_{it,p}\right)} \tag{11}$$

Since the predicted log odds of survival are linear in the hospital and surgeon group fixed effects and patient covariates, I decompose the variance as

$$Var\left(\ln\left(\frac{\hat{p}_{ijht}}{1 - \hat{p}_{ijht}}\right) - \sum_p \hat{\beta}_p X_{it,p}\right) = Var(\hat{\alpha}_{l(j)}) + Var(\hat{\psi}_{k(h)}) + 2 \times cov(\hat{\psi}_{k(h)}, \hat{\alpha}_{l(j)}) \tag{12}$$

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 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 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 B.12.

Appendix Figure B.7 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.

[17]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.

Alternative classifications. I examine the robustness of results to using alternative classifications in Panel D of Table 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 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 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 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 7 also indicates imperfect substitutability between surgeon and hospital quality.

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.

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 (6) using the new assigned surgeon and hospital groups.

As reported in Table 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 effect 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 (Eeckhout and Kircher, 2011; Bonhomme, Lamadon, and Manresa, 2019; Adhvaryu et al., 2020). 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.

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*,*b*; Baumgardner, 1988; Rosenthal, Zaslavsky, and Newhouse, 2005; Dingel et al., 2023). However, as indicated in Section 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.

Results reported in Table 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.

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 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). 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.

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 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 directly on a 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.

References

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.

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.

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.

Baumgardner, James R. 1988. "Physicians' services and the division of labor across local markets." *Journal of Political Economy*, 96(5): 948–982.

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.

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.

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, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics*, 36(S1): S13–S70.

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 of Disease Control and Prevention. 2023. "Heart Disease Facts." https://www.cdc.gov/heartdisease/facts.htm#:~:text=due%20to%20death.-,Coronary%20Artery%20Disease,killing%20375%2C476%20people%20in%202021.&text=About%201%20in%2020%20adults,have%20CAD%20(about%205%25).&text=In%202021%2C%20about%202%20in,less%20than%2065%20years%20old., 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.

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.

Dingel, Jonathan I, Joshua D Gottlieb, Maya Lozinski, and Pauline Mourot. 2023. "Market size and trade in medical services." National Bureau of Economic Research.

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.

Eeckhout, Jan. 2018. "Sorting in the labor market." *Annual Review of Economics*, 10: 1–29.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying sortingin 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."

Freund, Lukas. 2022. "Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities." *Available at SSRN 4312245*.

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.

Hoxby, Caroline M. 2014. "What Lies Ahead for America's Children and Their Schools." Chapter Covering the Costs. Chester E. Finn & Richard Sousa.

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*.

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.

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.

McDermott, Kimberly W, and Lan Liang. 2021. "Overview of operating room procedures during inpatient stays in US Hospitals, 2018: statistical Brief# 281."

Morris, Carl N. 1983. "Parametric empirical Bayes inference: theory and applications." *Journal of the American Statistical Association*, 78(381): 47–55.

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*.

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*. "The geographic distribution of physicians: Is the conventional wisdom correct?" Santa Monica: RAND Corp. Publ. No. R-2734.

Rosenthal, Meredith B., Alan Zaslavsky, and Joseph P. Newhouse. 2005. "The Geographic Distribution of Physicians Revisited." *Health Services Research*, 40(6p1): 1931–1952.

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.

Small, Kenneth A, and Harvey S Rosen. 1981. "Applied welfare economics with discrete choice models." *Econometrica: Journal of the Econometric Society*, 105–130.

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.

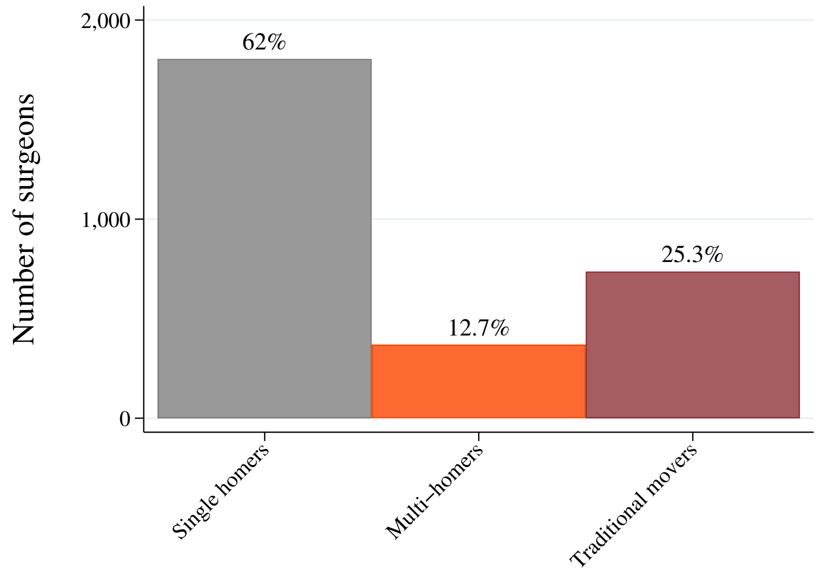
Wilson, Chad T, Steven Woloshin, and Lisa M Schwartz. 2007. "Choosing where to have major surgery: who makes the decision?" *Archives of Surgery*, 142(3): 242–246.

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.

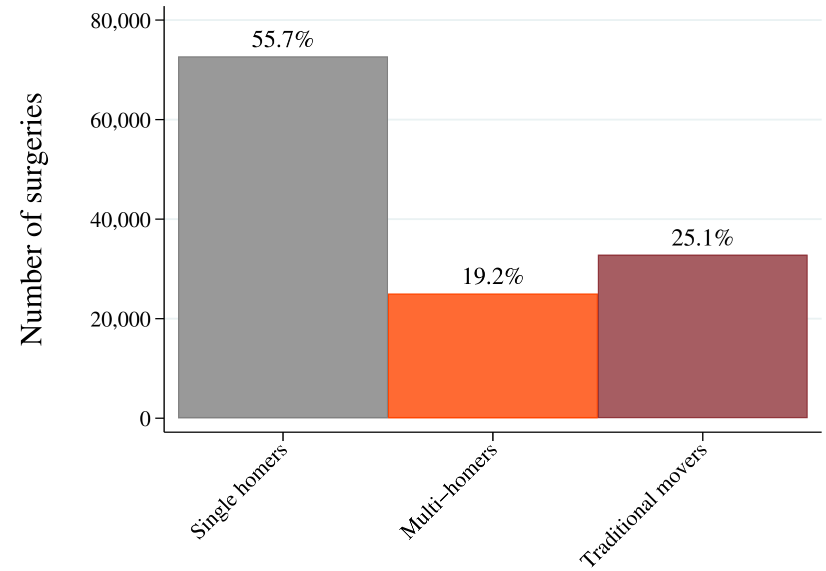
Exhibits

Figure 1: Proportion of “single-homers,” “multi-homers,” and “traditional movers”

(a) Proportion of surgeons

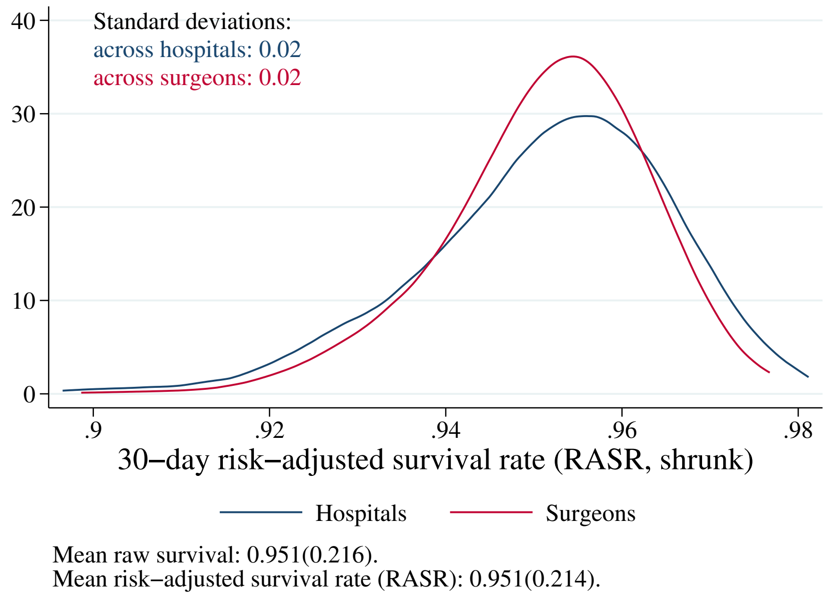


(b) Proportion of surgeries



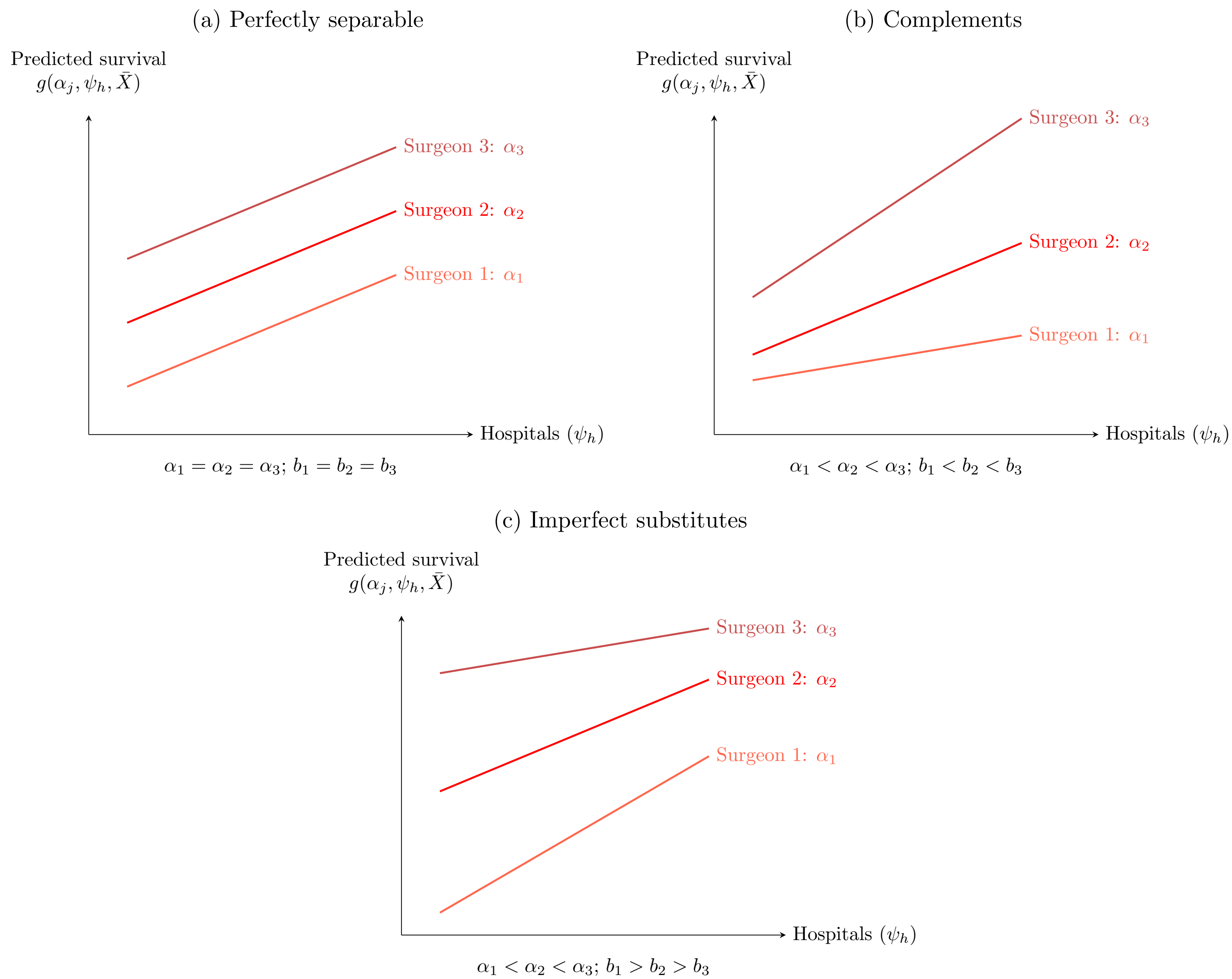
Notes: “Multi-homers” are defined as surgeons who performed CABG surgeries at more than one hospital within a year for four years of more in the sample. They represent 12.7% of all surgeons performing CABG in the sample, and 19.2% of CABG surgeries in the sample. “Traditional movers” are surgeons who 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 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 A.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 A.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 3: Impact of alternative assumptions on interactions between surgeon and hospital quality on predicted survival

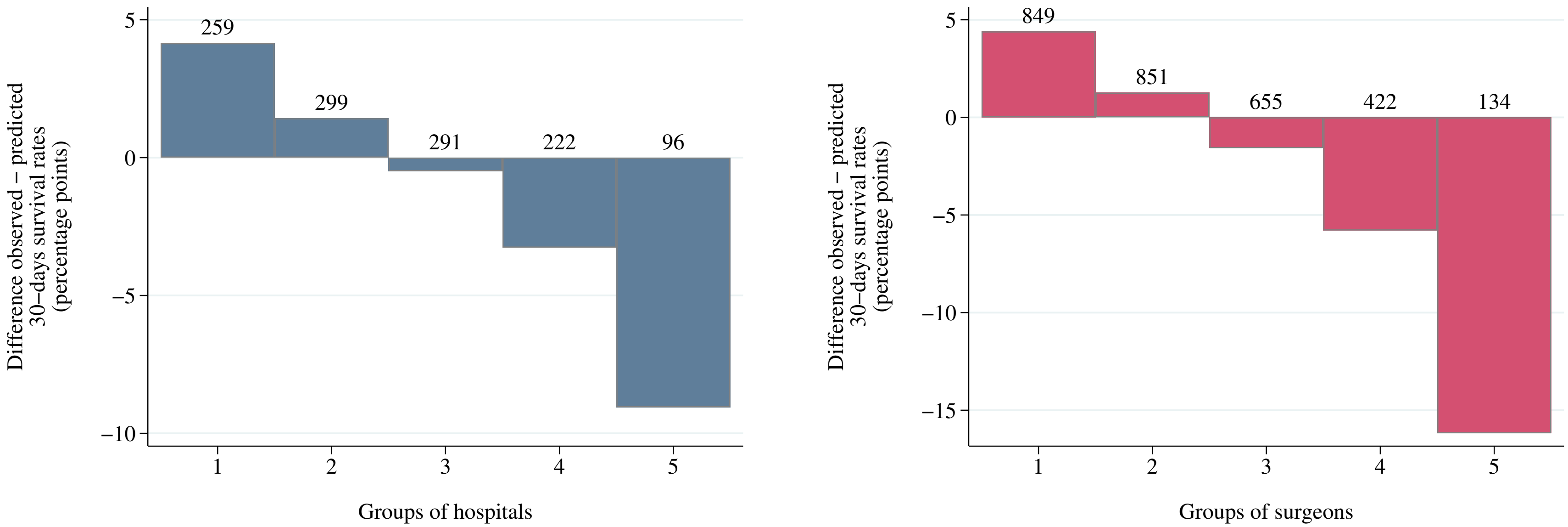


Notes: These figures illustrate the impact of alternative assumptions on the interactions between surgeon and hospital quality on predicted patient survival across providers. Panel 3a describes the case where surgeons and hospitals are perfectly separable. In this case, the slope across hospitals types b_j is equal for all surgeons: the return to allocating surgeons to high-ψ_h hospitals is independent of the surgeon. Panel 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 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 4: Risk-adjusted survival rate variation across k-means groups

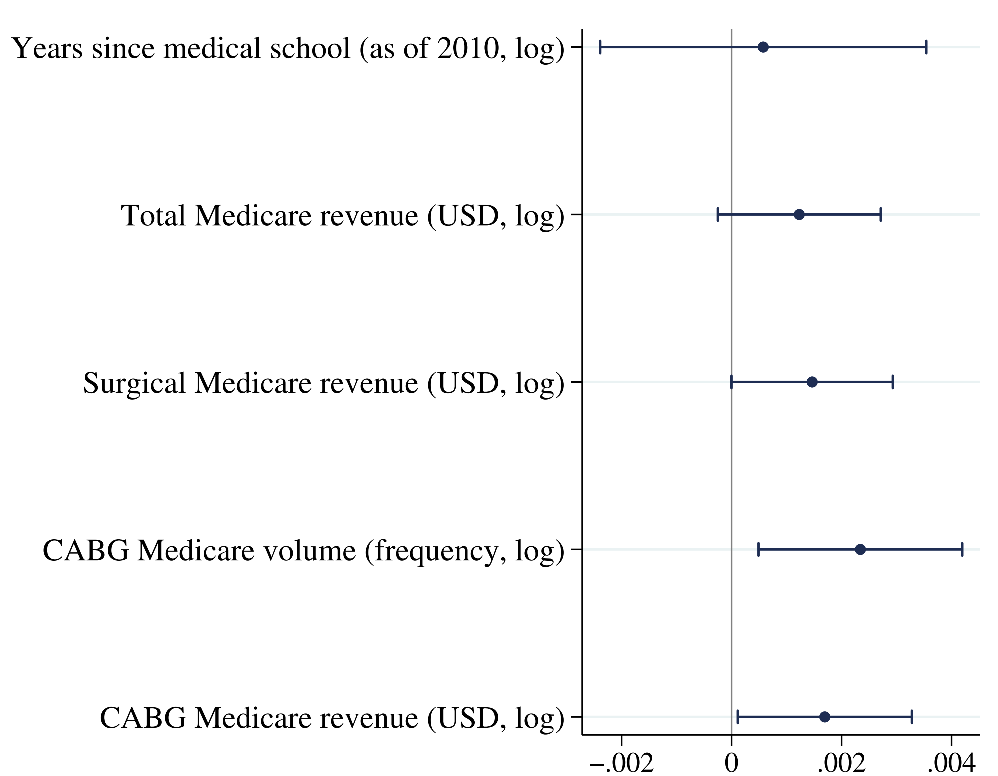
(a) Hospitals

(b) Surgeons



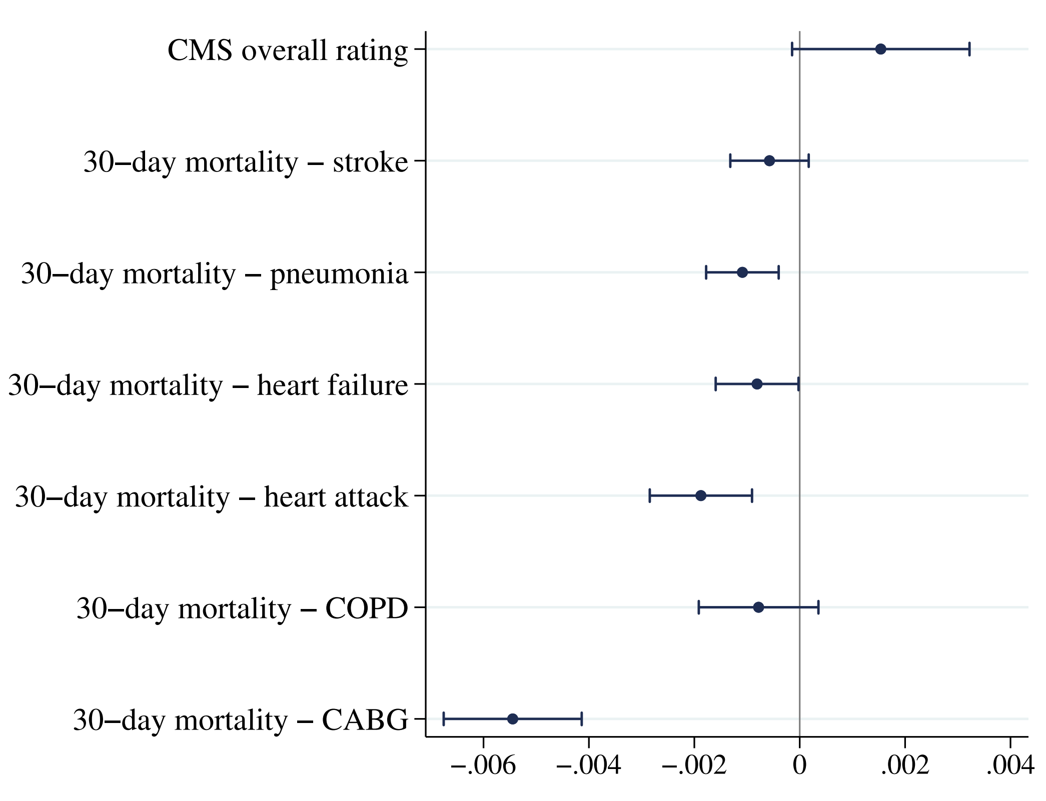
Notes: The variation in risk-adjusted 30-day survival across groups of hospitals and surgeons resulting from k-means clustering is large. Risk-adjusted survival is expressed as the difference between the average observed and average predicted 30-day survival. Numbers on top of bars indicate the number of hospitals or surgeons per group. Predicted survival is computed as described in Section 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 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 5: Correlation of estimated surgeon group effects with external measures of surgeons' skill



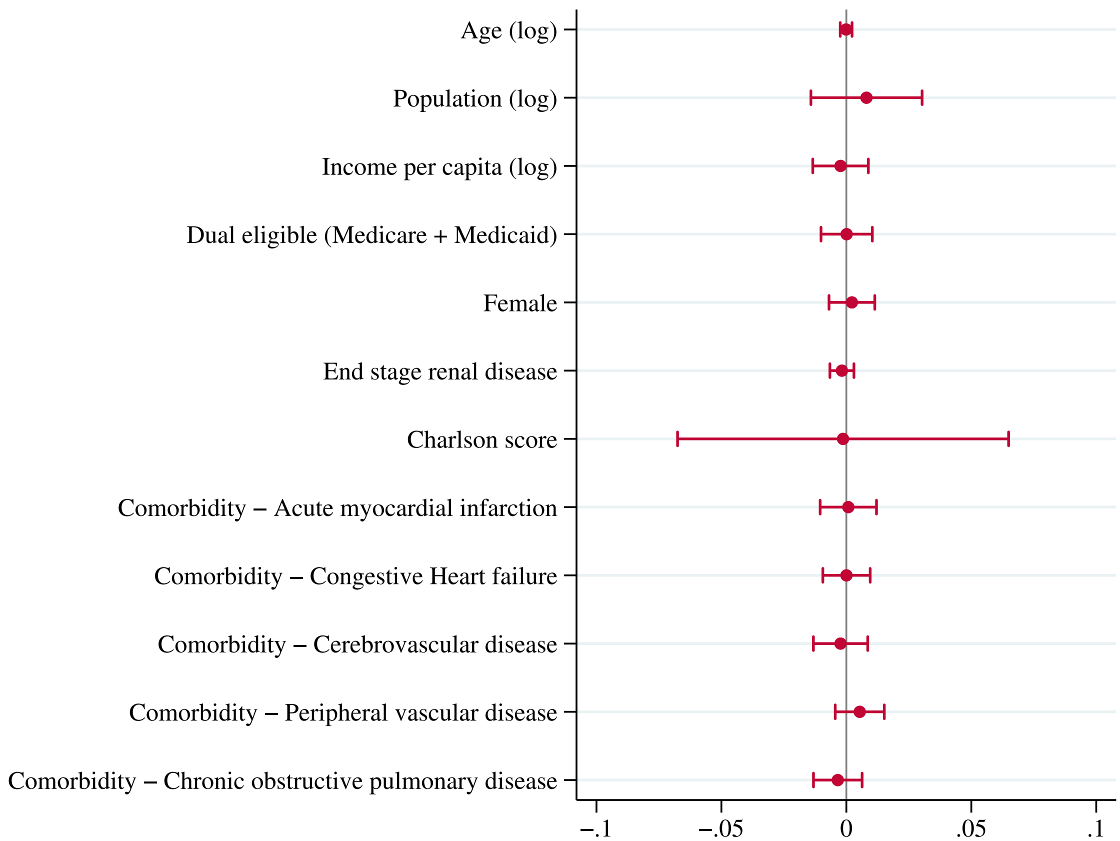
Notes: This graph reports the point estimates and 95% confidence intervals from regression of the surgeon-group estimated effects on surgeon-level covariates. The surgeon-group estimates include the fixed effects with interactions as $\widehat{\alpha_l} + \frac{1}{K}\sum_k \widehat{\kappa_{lk}}$ from equation (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 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 (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.

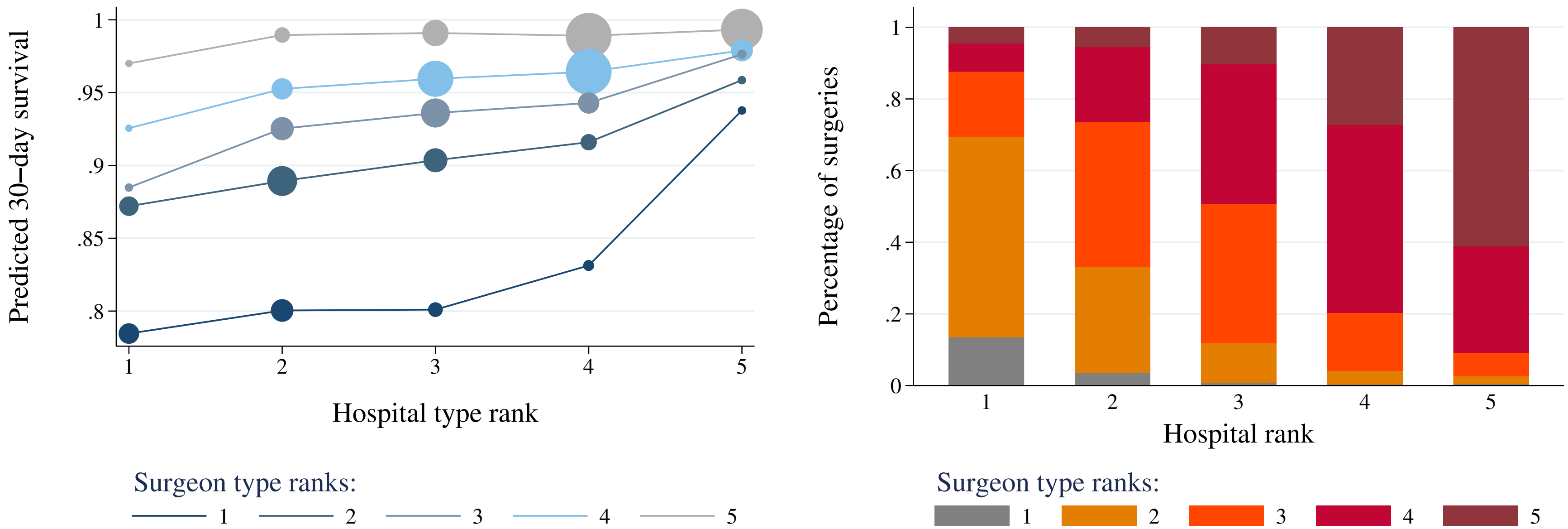
Figure 7: No evidence that surgeons systematically triage sicker patients into higher-survival hospitals



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 (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 3. Confidence intervals displayed are 95% confidence intervals constructed using clustered standard errors at the surgeon level.

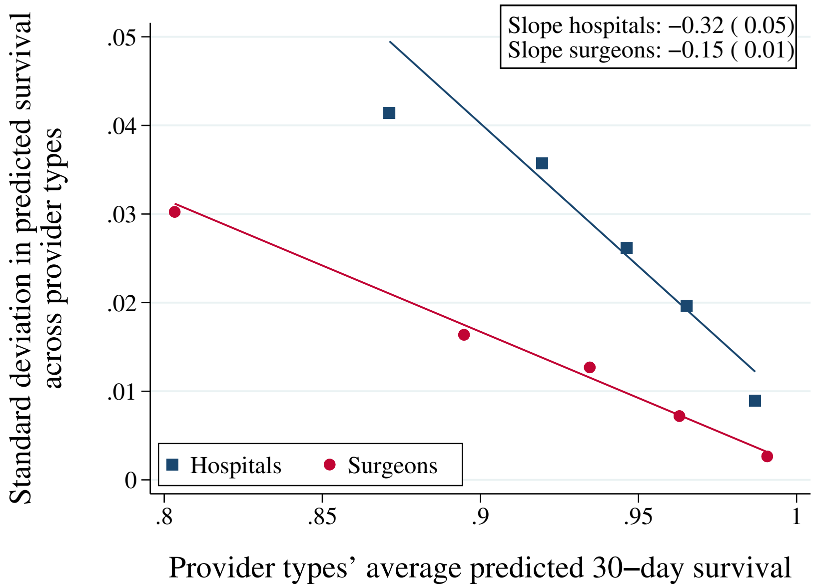
Figure 8: Imperfect substitutability and sorting

(a) Imperfect substitutability (b) Sorting



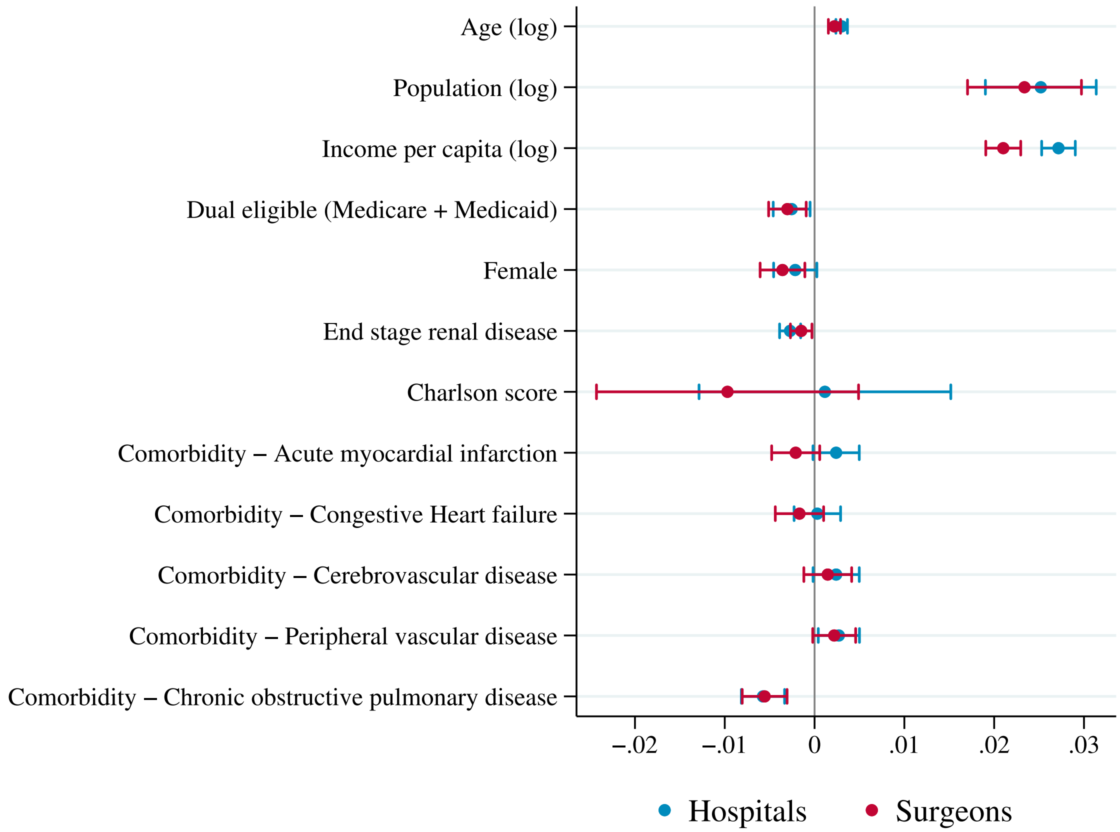
Notes: These graphs show results when assuming selection on observables as delineated in equation (6). Panel 8a displays the predicted 30-day survival for the average patient in the data across hospital and surgeon groups where groups are described by their relative rankings. The production function 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 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 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 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 9: Lower dispersion in survival among high-survival providers



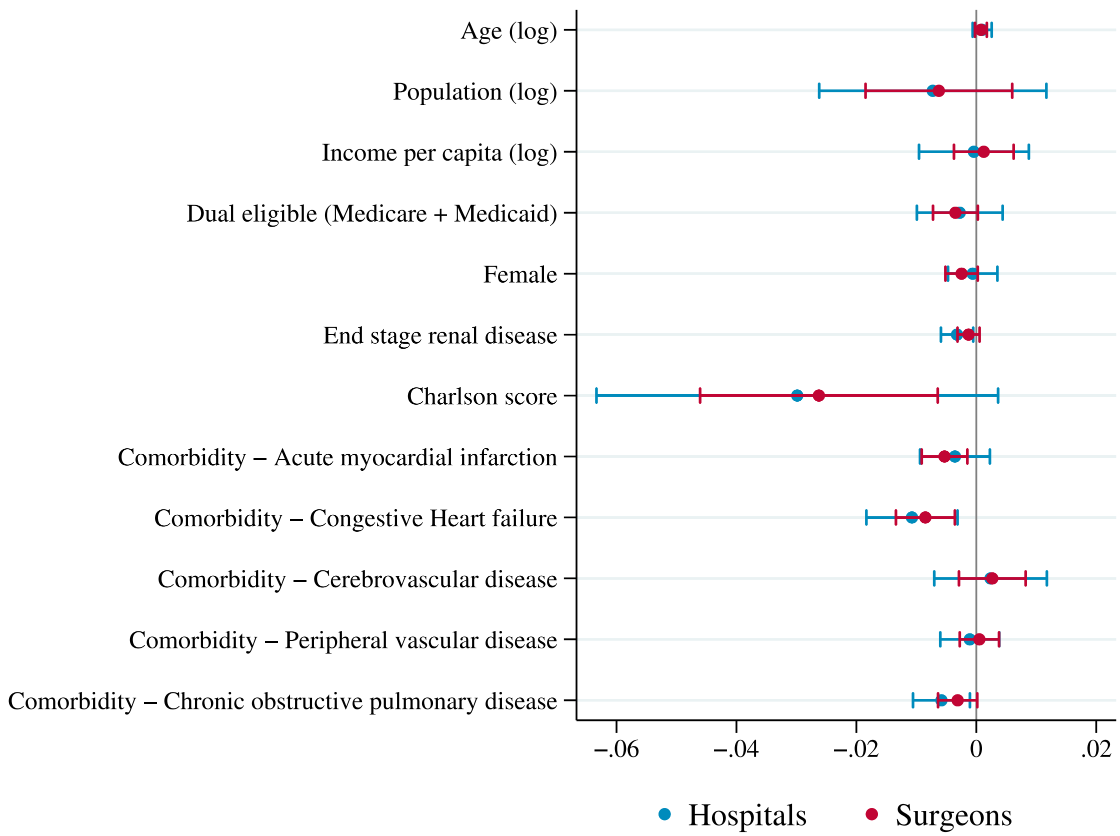
Notes: This graph shows that high-survival hospitals exhibit more similar survival across surgeons. This is depicted by the hospitals serie, which relates the standard deviation in predicted 30-day risk-adjusted survival across surgeon groups for each hospital group to the average predicted 30-day risk-adjusted 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 (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 10: Relationship between patient observables and provider rankings



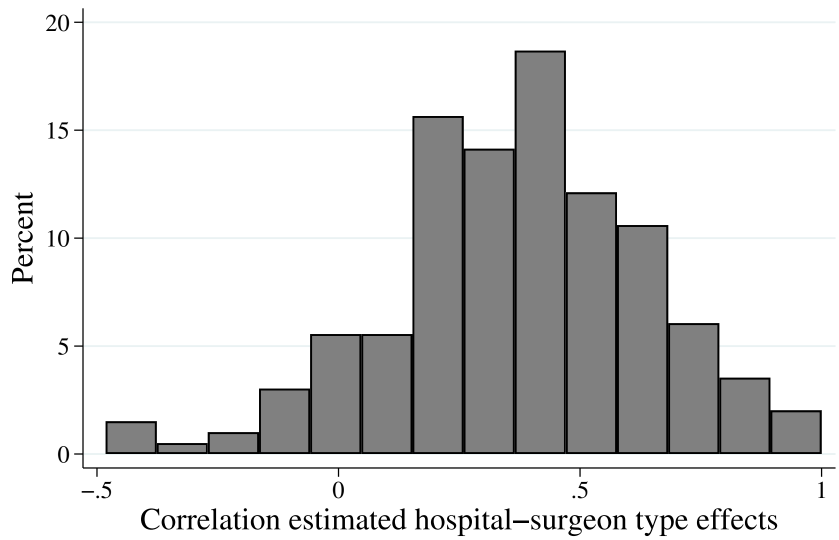
Notes: This graph shows the relationship between patient observables and the rank of the provider they receive surgery from. Highly ranked surgeons and hospitals tend to treat older patients living in highly populated high income ZIP codes, suggesting that high-type providers tend to be located in such 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 (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 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 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 (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 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 12: Sorting within hospital referral regions (HRRs)



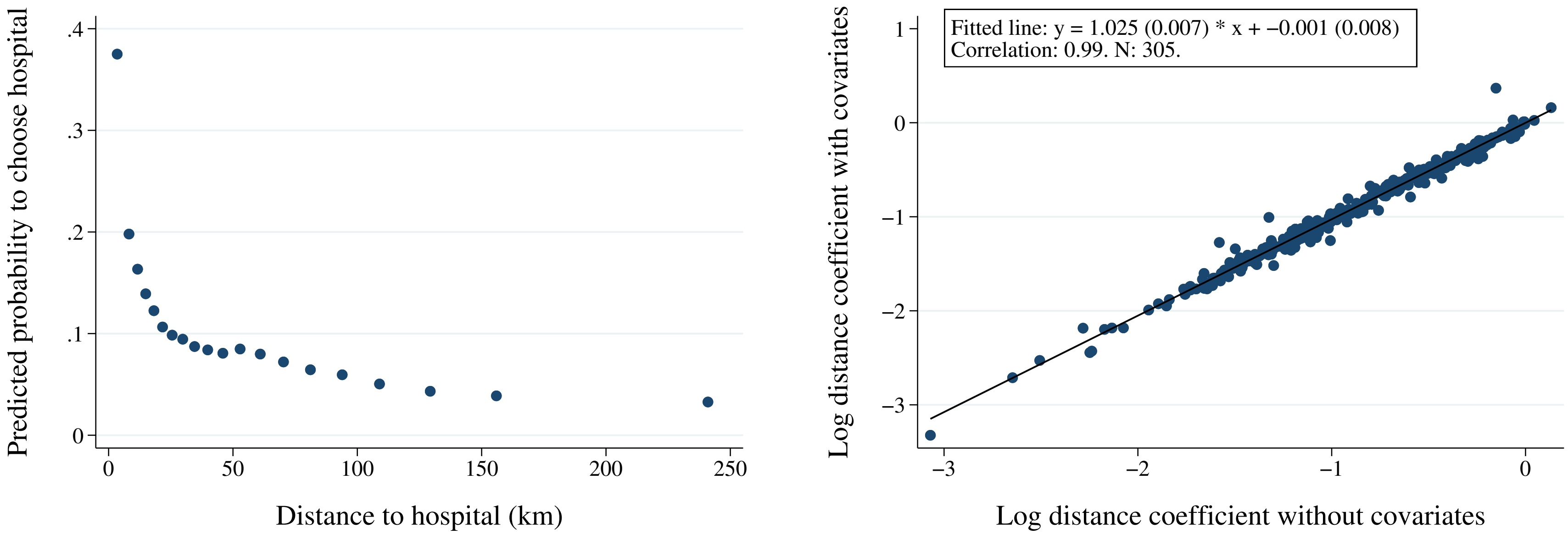
Number of HRRs: 228.
77 HRRs are missing because they only have one hospital or surgeon type.

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 (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 13: Distance to the hospital is a strong predictor of hospital choice within an HRR

(a) Probability to choose a hospital and distance

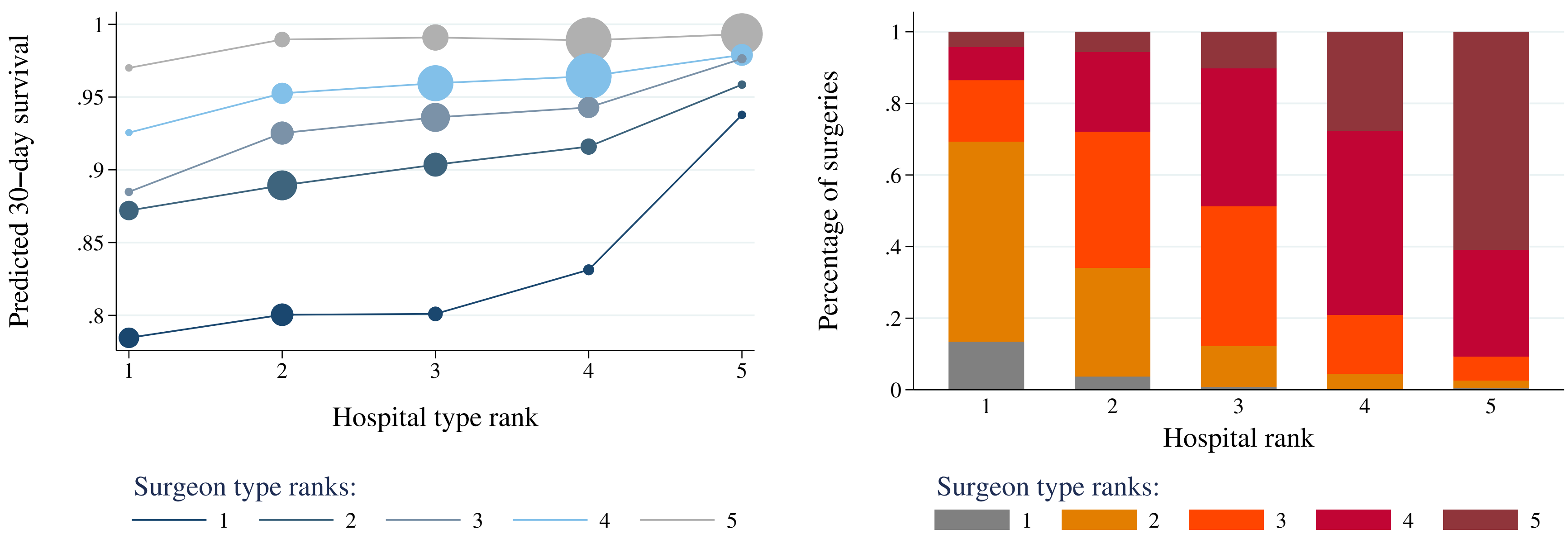
(b) Distance parameters per HRR



Notes: Panel 13a depicts the relationship between the predicted probabilities to choose a hospital using the demand model delineated in equation (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 sumarize this relationship using a binned scatter plot with twenty equally sized bins. Panel 13b depicts the estimated demand parameter for the logarithm of distance τ in the specification without patient observables from equation (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 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 (8). Panel 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 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 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 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: 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 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-days mortality	0.05	0.22
60-days 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 in the twelve months prior to the surgery for years 2012-2017, and in the year of surgery for 2011. 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 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 (6) and (8). 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 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 (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 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 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 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 (10). Fixed effects are estimated following equation (9). 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 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 types			
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	67.38	7.84	24.78	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 (10) 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 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 (6) or (8), 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 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 8: Alternative allocations of surgeons to hospitals: surgeon sorting has large consequences for aggregate 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 (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.

A Appendix

A.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.

A.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.[18]

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

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

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.

A.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

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

where X_{it} include patient covariates included in Table 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.

A.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 A.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 2.

A.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:

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

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

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

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.

A.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.[19] 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 B.4. 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 B.5, a large amount of hospitals and surgeons are misclassified so that the correlation between the 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,

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

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.

A.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.

I impose the following linear structure to the conditional expectation of ϵ_{ijht}

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

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 ψ is choice-specific and identifies selection on gains.

Assume the following production function of survival g:

$$Pr(y^*_{ijht}=1|\alpha_{l(j)},\psi_{k(h)},\kappa_{l(j)k(h)},X_{it},\epsilon_{ijht})=\alpha_{l(j)}+\psi_{k(h)}+\kappa_{l(j)k(h)}+\beta X_{it}$$

where y_{ijht} is a binary variable equal to one when patient i survives, $\alpha_{l(j)}$ and $\psi_{k(h)}$ are the unobserved heterogeneity for surgeon and hospital types respectively, $\kappa_{l(j)k(h)}$ the interaction between surgeon and hospital types, and X_{it} are patient-level observables and ϵ_{ijht} are unobserved health shocks.

Now 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. Note that $D_i=h$ is such that

$$u_{ih}\geq\max_{s\neq h}u_{is}\iff\delta_h-\tau\ln d_{ih}+\lambda_j+\eta_{ih}\geq\max_{s\neq h}\delta_s-\tau\ln d_{is}+\lambda_j+\eta_{is}$$

so it depends on δ_h, τ, and $(\ln d_{i1},...,\ln d_{iH})$. In other words, it is a restriction on the $(\eta_{i1},...,\eta_{iH})$

The observed survival is conditional on the choice of hospital and all covariates such that

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

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

To obtain it, we have the potential survival as potential outcome (not conditional on hospital choice):

$$E(y^*_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}, \eta_{i1}, ..., \eta_{iH}) =$$
$$\alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} + \sum_{s \in \mathcal{H}} \phi_s(\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta)$$

so we get

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

To derive the control functions, we have

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

Using Small and Rosen (1981), we have

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

so that

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

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

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

Use

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

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

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

Therefore, the control function is such that

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

Note that the control function is positive when $s = h$ but negative otherwise since $\ln \hat{p}_{is} < 0$ with $0 < \hat{p}_{is} < 1$.

Overall, we have

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

and the observed survival, conditional on the choice of hospital, is

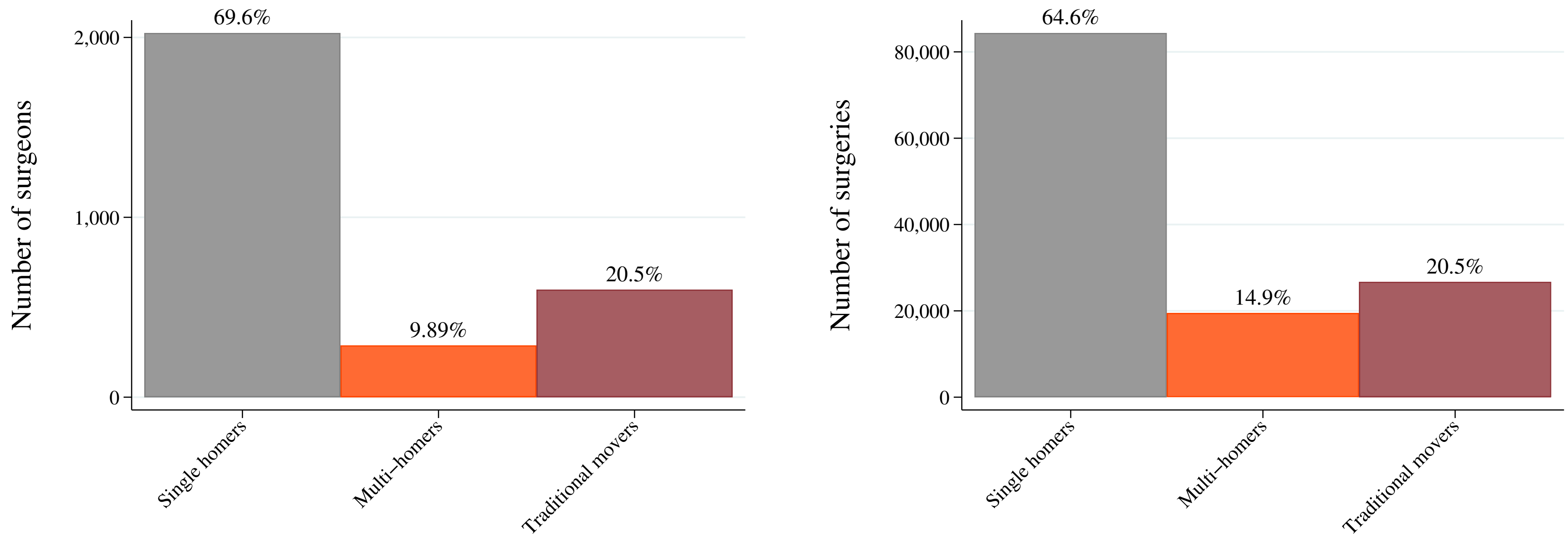
$$E(y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, X_{it}, \ln d_{i1}, ..., \ln d_{iH}, D_i = h) = \\ \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} + \sum_{s \in \mathcal{H}} \phi_s \theta_{is}(h) + \varphi \theta_{ih}(h)$$

with $\theta_{is}(h)$ as defined above. Note that d_{ih} are the instruments, excluded from the second step, with the exclusion restriction $cov(\epsilon_{ijht}, d_{ih}) = 0$.

B Additional exhibits

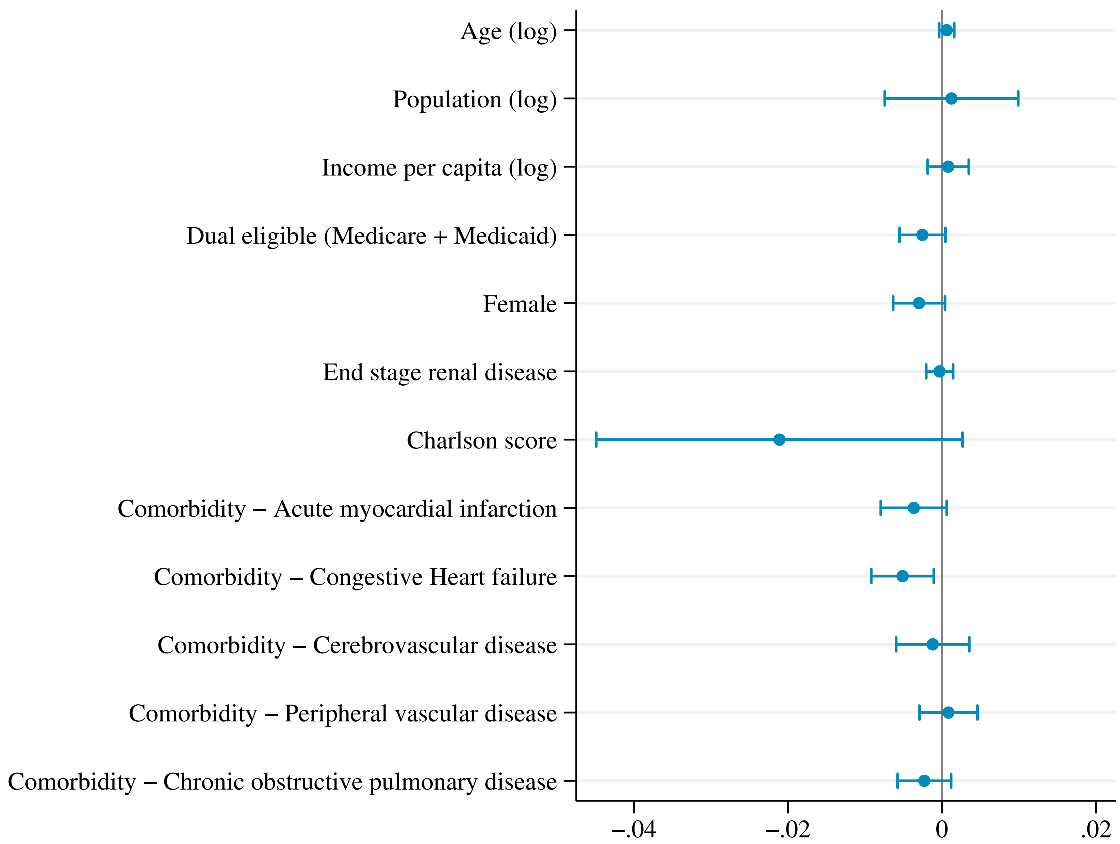
Figure B.1: Proportion of “single-homers,” “multi-homers,” and “traditional movers” using hospital groups

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



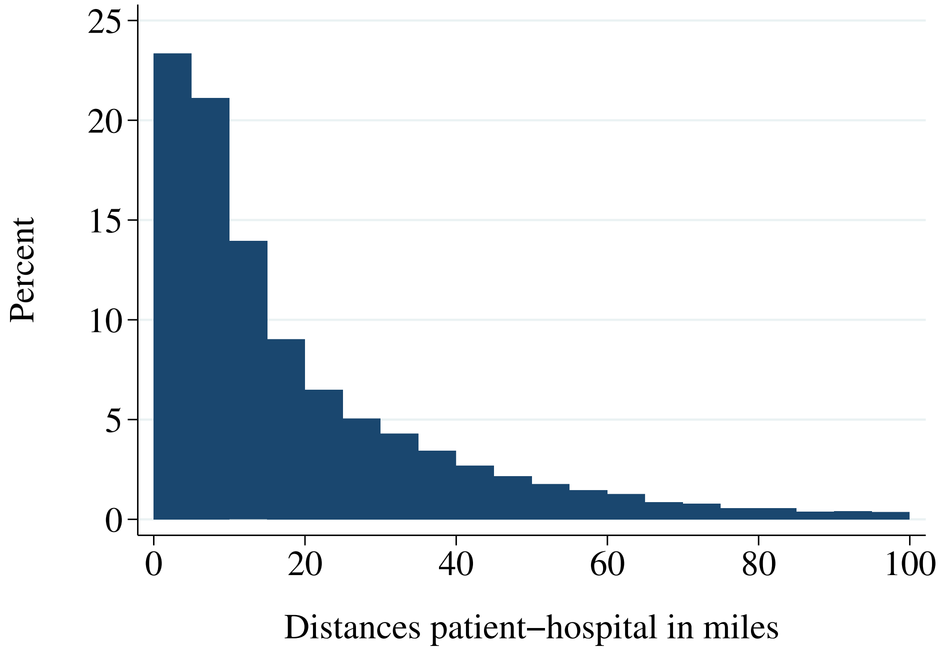
Notes: The fraction of surgeons observed at more than one hospital *group* falls to about 30%, compared to Figure 1 in which the fraction of surgeons observed at more than one hospital is close to 40%. “Multi-homers” are defined as surgeons who performed CABG surgeries at more than one hospital *group* 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 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 B.2: 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 (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 3. Confidence intervals displayed are 95% confidence intervals constructed using clustered standard errors at the hospital level.

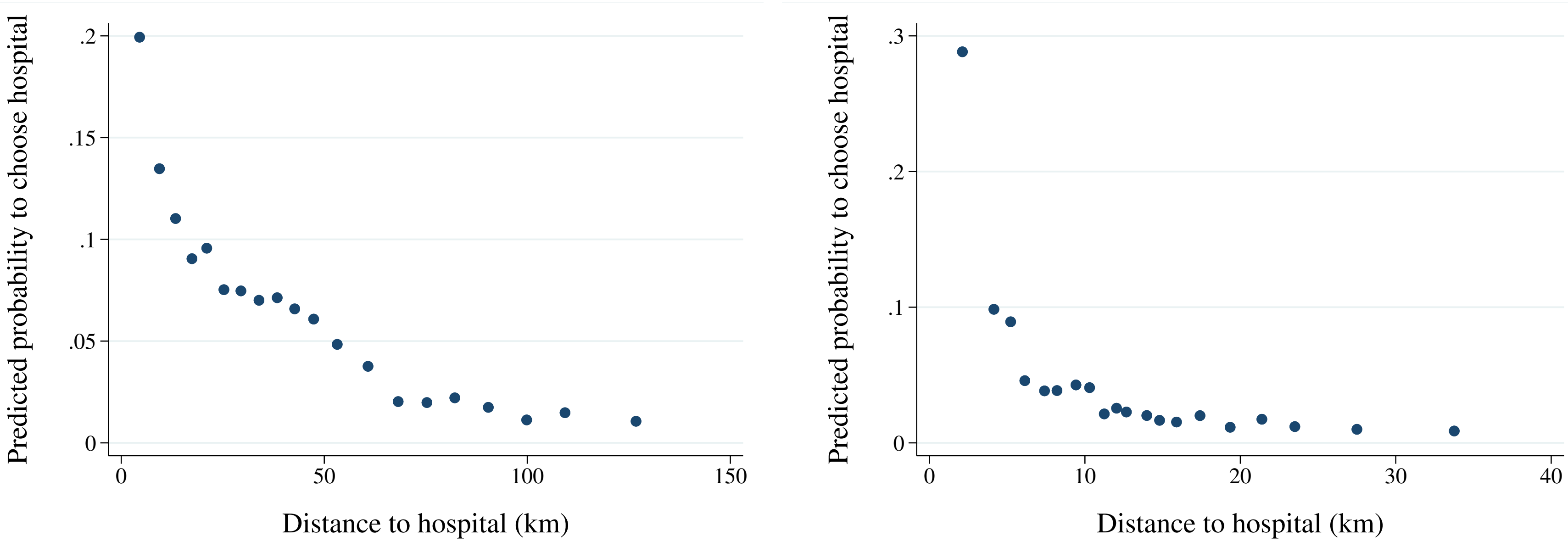
Figure B.3: Distribution of distances between patients and their chosen hospital



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.

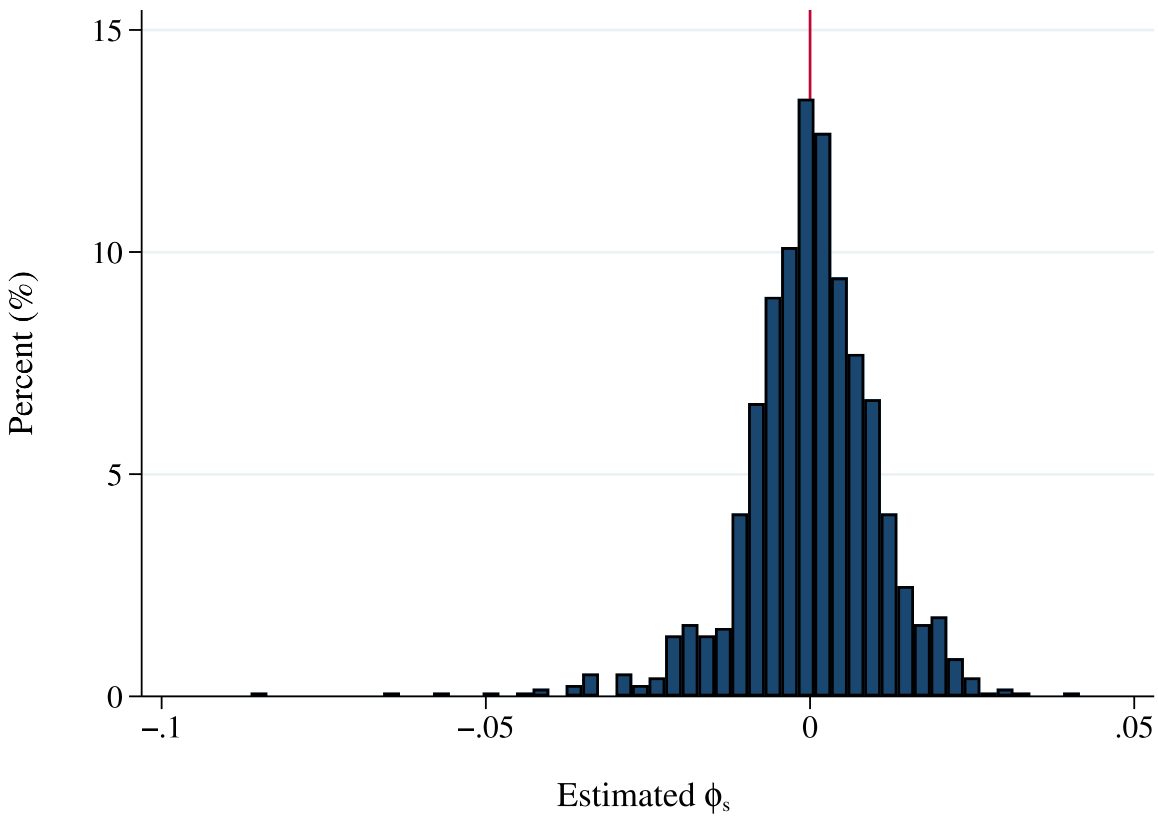
Figure B.4: Distance to the hospital is a strong predictor of hospital choice within HRRs: Boston and Chicago

(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 (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.

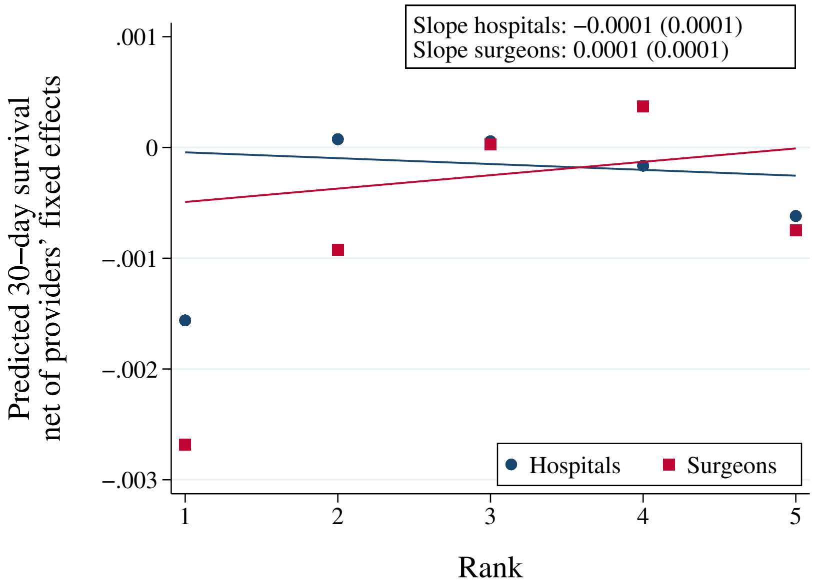
Figure B.5: Distribution of estimated control function parameters $\hat{\phi}_s$ across hospitals



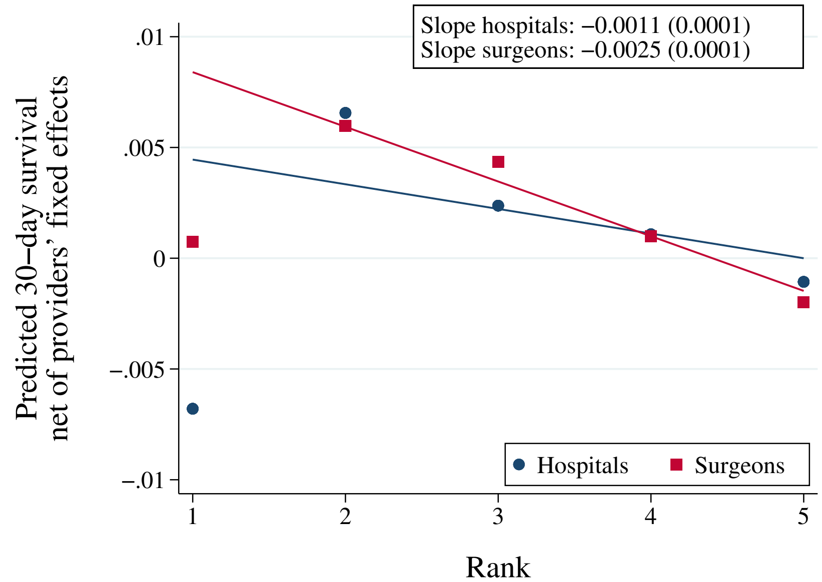
Notes: This graph shows the distribution of the estimated control function parameters $\hat{\phi}_s$ from equation (8), with s denoting a specific hospital. When the estimated coefficient is negative, sickest patients tend to select into that specific hospital. Conversely, when the estimated coefficient 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 B.6: Relationship between predicted survival net of provider fixed effects and provider rankings

(a) Including patient covariates only



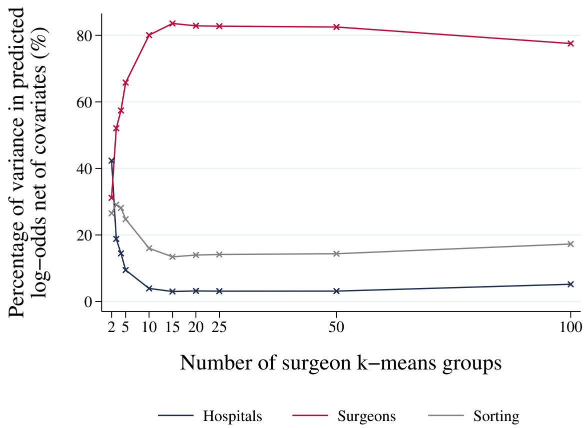
(b) Including control function arguments



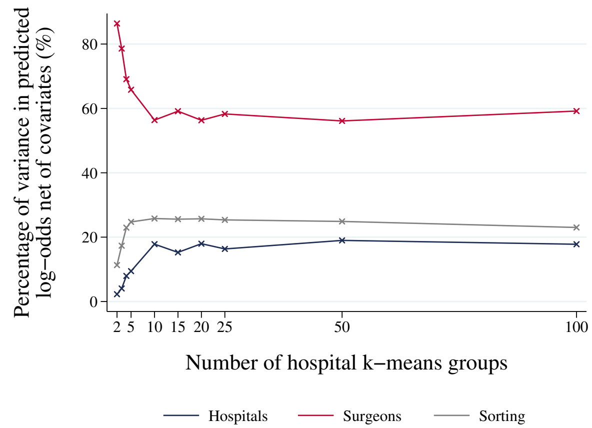
Notes: This figure reports the relationship between predicted 30-day survival net of provider fixed effects and the rank of providers from the model in equations (6) and (8). There is no systematic relationship between predicted survival net of provider fixed effects and the ranking of their provider 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 (6) and (8) respectively. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 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 B.7: Robustness of the variance decomposition to alternative numbers of k-means groups

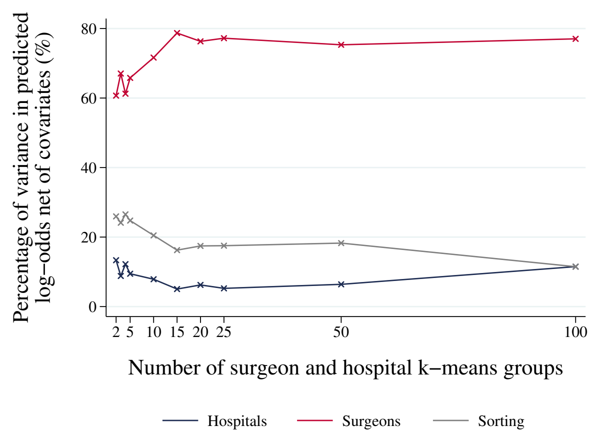
(a) Vary L, K=5



(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 (12). 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 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 B.1: 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-days 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-days 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 B.2: 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 B.3: 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 A.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 B.4: 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 B.5: Monte Carlo simulation results, assuming negative 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	95.3	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 B.6: 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 (6), i.e., weighting each interaction with each hospital group equally. Results are similar when using the estimated group effects from equation (8), 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 B.7: 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 (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 (8). 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 B.7: 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 (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 (8). 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 B.8: 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 (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 (6). Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 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 B.9: Limited evidence of "triaging" within surgeons and within hospitals using patient's predicted survival net of provider fixed effects

	(1)	(2)
Predicted 30-day mortality:		
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 (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 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 B.10: Variance decomposition for 30-day survival

	Percentage of variance (%)	
	Selection on observables	Control function
Hospitals	0.24	0.18
$Var(\psi_{k(h)})$		
Surgeons	1.81	2.23
$Var(\alpha_{l(j)})$		
Sorting	0.68	0.66
$2 \times cov(\alpha_{l(j)}, \psi_{k(h)})$		
Patients covariates	1.85	2.39
$Var(\beta X_{it})$		
Year	0.20	0.20
$Var(\lambda_t)$		
Covariance FEs-patients covariates	-0.33	-0.70
$2 \times cov(\alpha_{l(j)} + \psi_{k(h)} + \lambda_t, \beta X_{it})$		
Covariance surgeon and hospital-year	-0.01	-0.01
$2 \times cov(\alpha_{l(j)} + \psi_{k(h)}, \lambda_t)$		
Residuals	95.56	95.06
$Var(\epsilon_{ijht})$		
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 (9). 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 B.11: The relationship between patient outcomes and distance to the chosen hospital is similar when including patient observables

	(1)	(2)	(3)
30-days survival			
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 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 in the twelve months prior to the surgery for years 2012-2017, and in the year of surgery for 2011. 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 B.12: 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 (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 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.