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

Pauline Mourot
Boston University

September 2025

Abstract

I examine the roles of physician sorting across hospitals and physician-hospital complementarities in driving aggregate patient outcomes. I estimate 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 perform surgeries at multiple hospitals within a year. I leverage this variation in a two-way fixed-effects strategy with interactions between hospital and surgeon quality 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 exhibit higher returns from practicing at a high-survival hospital than high-survival surgeons do. I use the recovered parameters to examine the impact of provider sorting across and within regions on aggregate survival.

*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, Santiago Franco, Alex Frankel, Luis Garicano, Arjun Gopinath, Atul Gupta, Anders Humlum, Thibaut Lamadon, Jacob Leshno, Magne Mogstad, Jack Montjoy, Raffaele Saggio, Karthik Srinivasan, Chad Syverson, Nicholas Tilipman, and Walter Zhang. I am grateful to seminar and conference participants at the University of Chicago, the Cornell Brooks School of Public Policy, PUC Chile Business, Boston University, the University of Rochester Simon School of Business, Boston College, the Management group at the London School of Economics, Imperial College London, the University of Maryland College Park, the BFI Health Economics Initiative Annual Conference, the University of Illinois at Chicago, the Annual Health Econometrics Workshop, the Society for Economics Dynamics Winter meeting, the 2025 ASSA meetings, Yale University, the University of Indiana Bloomington, and the 2025 ASHEcon meetings for helpful feedback. I thank the Becker Friedman Institute at the University of Chicago for enabling me to access the Medicare claims data. All errors are my own.
Author's email address: pmourot@bu.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, a highly regulated industry, 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 common heart surgery. Second, using the estimated production function, I evaluate the impact of current and counterfactual allocations of surgeons to hospitals to examine the impact of surgeon sorting 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 survival, which has also been the focus of the literature investigating provider quality.[2]

I first show that surgeon sorting and non-perfect separability between surgeon and hospital value-added likely play a role in the context of CABG surgery. Using risk- and noise-adjusted measures of surgeon and hospital quality computed separately on different samples and dimensions, I find that higher-survival surgeons for CABG surgery tend to practice at hospitals with better quality rankings. Furthermore, I find suggestive evidence that lower-survival surgeons exhibit higher marginal returns to operating at higher-ranked hospitals. This latest fact rejects perfect separability between surgeon and hospital value-added but also suggests that surgeon and hospital value-added are substitutes in the production function of survival from CABG surgery.

There are two main limitations to this reduced form analysis. First, surgeon and hospital performance measures do not separately identify surgeons from hospitals effects. Second, these measures also do not capture surgeon and hospital effects on the relevant dimension of heterogeneity, i.e. both on their ability to achieve patient survival for CABG surgery. For example, a hospital that is highly ranked overall may not necessarily be equally ranked across all procedures, including CABG surgery. To remedy both limitations, I develop an empirical strategy to separately identify surgeon and hospital effects and their interactions

[1] See for example Chade, Eeckhout, and Smith (2017) and Eeckhout (2018) for reviews of this literature.
[2] In the case of CABG surgery specifically, see for example Huckman and Pisano (2006); Kolstad (2013).

on the relevant dimension of heterogeneity, survival for CABG surgery.

I estimate the joint production of patient survival between surgeons and hospitals using a two-way fixed-effects strategy with interactions between 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". Since I observe output at the surgeon-level directly, I can address identification issues for sorting which have been outlined in two-way fixed-effect models when using worker earnings as a proxy for worker productivity in the presence of frictions (Eeckhout and Kircher, 2011).[3]

Estimating individual provider value-added and their interactions in a model with two-sided heterogeneity raises two main challenges. First, identifying interaction effects when the network is sparse, i.e. when physicians do not practice at all hospitals, is a well-known identification problem when one is interested in examining the impact of alternative allocations of surgeons to hospitals. Additionally, individual value-added estimates are noisy when the outcome is a relatively rare event like mortality. To address these issues, 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) but extending it by grouping *both* sides of heterogeneity. I classify surgeons and hospitals using a proxy for their individual quality: provider-level risk-adjusted survival. This measure is used in practice and in the literature to describe providers' individual quality for CABG surgery. Second, patient sorting into providers may violate the exogenous network assumption required for identification of the fixed effects in a two-way fixed-effects model with a tri-partite network. I use a control function approach leveraging distance to hospitals as an excluded instrument to identify patient selection on unobservables as in Einav, Finkelstein, and Mahoney (2022). In the context of two-sided heterogeneity models, this acts as a partial endogenization of network formation by modeling the choice of hospitals by patients.

I find that surgeon and hospital quality are 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. Examining sorting, I find positive assortative matching: high-survival surgeons sort into high-survival hospitals. Exploring potential mechanisms, I find suggestive evidence that a mechanism outlined in the medical literature–"failure-to-rescue"–likely drives the substitutability result (Silber et al., 1992; Ghaferi, Birkmeyer, and Dimick, 2009). High-survival hospitals achieve higher survival rates

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

by rescuing their patients from complications, so that they have more role in raising the performances for lower-survival surgeons, shown to exhibit higher complication rates in the literature. Consistent with this mechanism, patients operated by lower-survival surgeons do exhibit higher-treatment intensity at higher-survival hospitals, measured using the length of hospital stay and hospital spending.

I evaluate the robustness of these results along several dimensions. Results are very similar when allowing for patient selection into hospitals on unobservables using a control function approach. Since I extend the grouped fixed-effect estimator to grouping both surgeons and hospitals in the classification step, I examine the robustness of my results to correlated classification error with three distinct exercises. I first use two split-sample strategies where surgeons and hospitals are classified using risk-adjusted survival on distinct samples. Both approaches deliver results in line with positive assortative matching and substitutability but results are noisy since the sample size drops substantially. Therefore, I use alternative outcomes in a third exercise which are much less noisy than 30-day survival: length of stay and hospital spending. I run the main specification and the split-sample strategy with these alternative outcomes. Results for positive assortative matching and substitutability between surgeon and hospital value-added are robust to using these alternative outcomes, and the split-sample strategy confirms that correlated classification error does not drive both results. I also show that results are robust to alternative number of surgeon and hospital groups, alternative classification strategies, using a logit production function, and focusing on non-emergent CABG surgery.

I use the estimated value-added and interaction effects to examine the impact of surgeon sorting on aggregate patient survival. Partial equilibrium reallocation exercises show that aggregated patient survival is highly sensitive to the sorting of surgeons to hospitals. A random reallocation of surgeons to hospitals nationally 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 25%. To put these numbers in perspective, this would amount to approximately 800 lives saved within Medicare every year for CABG surgery alone. Therefore, the magnitude of the substitutability between surgeon and hospital value-added for CABG surgery is large. Surgeon sorting constitutes a crucial dimension to the design of healthcare policy promoting both aggregate and individual provider quality.

Furthermore, I find that positive assortative matching is driven by both providers locations across space and surgeon sorting across hospital within their hospital referral regions. Using estimated effects, I find evidence that high-survival providers tend to locate in the same hospital referral regions. At the same time, positive assortative matching remains substantial when evaluated *within* regions. To shed light on the respective contributions of each explanation for the national-level positive assortative matching result, I perform partial equilibrium reallocations of surgeons across hospitals within their baseline hospital referral regions. Such reallocations lead to substantial mortality gains representing a little over 50% of the national reallocation mortality gains. This indicates that both provider locations across space and surgeon sorting within regions contribute substantially to the aggregate positive matching result.

Finally, I examine the role of surgeon locations in driving the heterogeneity in predicted survival across more and less populated regions. I find that high-survival providers tend to co-locate in more populated hospital referrals regions. Accordingly, using partial equilibrium exercises that reallocate surgeons nationally, I find that less populated regions benefit more from these reallocations than more populated regions. This is driven by moving high-survival surgeons to less populated regions who also tend to have lower-survival hospitals. Interestingly, within-region reallocations lead to very different regional gains along the population gradient. More populated regions would tend to benefit *more* from such reallocations, since they exhibit both high- and low-survival providers.

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). Team formation has been shown to be a substantial driver of individual patient outcomes, through shared experience by team members for cardiac surgery in Chen (2021) for example. This paper is most closely related to Huckman and Pisano (2006), who find evidence that surgeon performance is hospital-specific. They show that surgeon performance is correlated with their volume at the specific hospital, but not to their volume at other hospitals. This paper is, to the best of my knowledge, the first to separately identify surgeon from hospital effects and identify the match effects between them in the production of patient risk-adjusted survival. Thanks to these estimates, this paper shows the crucial impact of surgeon sorting on aggregate and individual provider quality.

Relatedly, this paper contributes to the literature examining the drivers for healthcare differences across regions. Finkelstein, Gentzkow, and Williams (2016) and Finkelstein, Gentzkow, and Williams (2021) have emphasized the importance of location effects in driving geographic variation in utilization and life expectancy, respectively. Investigating the role for physicians, Badinski et al. (2023) shows that physicians are a key driver of supply-side geographic variation in healthcare utilization, though it varies across specialties, consistently with results from Molitor (2018) for cardiologists. Using a different approach, Dingel et al. (2023) emphasize the key role of economies of scale in the production of medical services that make more populated regions net exporters of medical services of higher quality. This paper contributes to this literature in two main dimensions. First, it provides additional evidence that high-survival providers tend to co-locate in space, in particular in larger regions. Second, it sheds light on the role for surgeon sorting across but also within regions in driving predicted mortality differences across regions. This paper also highlights the potential consequences for healthcare provider payment regulations, in line with recent evidence from Gottlieb et al. (2025) who highlight the role of health policy on the geographical patterns of physician earnings in the U.S.

This paper also contributes to a longstanding literature on worker sorting across firms. Recent work by Kline, Saggio, and Sølvsten (2020) and Bonhomme et al. (2023) has shown substantial positive assortative matching between workers and firms in Europe and the U.S. using worker earnings with fixed effects and random effects approaches. Bonhomme, Lamadon, and Manresa (2019) estimate the existence of complementarities directly and find evidence for (weak) substitutability between workers and firms in the presence of positive assortative matching when using worker earnings.[4] This paper has two main contributions to this literature. First, it extends the application of the grouped fixed-effect estimator to classifying both workers and firms into groups in a tri-partite network with partially endogenous network formation (Bonhomme and Manresa, 2015; Bonhomme, Lamadon, and Manresa, 2022). Second, it documents positive assortative matching in a specific yet important labor market using a direct measure of output and shows how interactions between worker and firm effects matters when focusing in this specific setting.

Beyond healthcare, this paper is closely related to the education literature that seeks to separate the impact of teachers and schools on student outcomes. An abundant literature has focused on estimating teacher and school effectiveness.[5] Most closely related is Jackson (2013), who also finds evidence that teacher effects are not fully portable across schools, and estimates that a substantial portion of teacher effects corresponds to teacher-school match

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

[5]See for example the review on teacher value-added from Koedel and Rockoff (2015) and Angrist, Hull, and Walters (2023) for a review on estimating school effectiveness. Abdulkadiroğlu et al. (2020) notably use a control function approach to allow for selection on unobservables when estimating school effects, extended to rank-ordered list choices in their setting.

quality. In my paper, I explore match effects that correspond to the existence of complementarity or substitutability between surgeon and hospital value-added in the production function of health.

The rest of the paper is organized as follows. I describe the institutional setting, provide an overview of the data, and show reduced form evidence for surgeon sorting across hospitals and the existence of significant interactions between proxies for surgeon skill and hospital quality in Section 2. I delineate the empirical strategy in Section 3. Section 4 evaluates the sensibility of estimated parameters. I detail the 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 describe the main sample, and motivate the empirical strategy with suggestive evidence related to surgeon sorting and match effects between surgeons and hospitals in the CABG surgery context.

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 insurers 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 that accepts Medicare, which the vast majority do. 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.

CABG 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 expensive: 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.[6] 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 for provider quality. Hospital 30-day risk-adjusted mortality rates after CABG surgery are publicly reported yearly by CMS and integrated in their hospital five-star rating measure. For surgeons, 30-day risk-adjusted mortality calculated at the surgeon level for CABG surgery started being publicly reported in the state of New York in 1991, followed by other states including Pennsylvania and Massachussetts.[7]

Cardiac surgeons operate on their patients at multiple hospitals within the same year without an actual change of employment, which I leverage as additional variation in my empirical strategy. A large fraction of cardiothoracic surgeons are not employed by hospitals, but are rather independent in private practices, like freelancers (Huckman and Pisano, 2006; Kolstad, 2013). To get access to an operating room to perform surgery, they need to obtain operating privileges at hospitals. Obtaining such privileges is relatively low cost, usually

[6] I detail the processes involved during CABG surgery in Appendix A.1.

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

a one time administrative cost, and there is no limit in the number of hospitals at which they can obtain operating privileges. Operating at multiple hospitals notably allows for scheduling flexibility for surgeons, as detailed in Appendix A.1.

2.2 Data

The final sample includes a total of 110,672 patients treated by 2,892 surgeons across 1,167 hospitals between 2011 and 2017.[8] I detail the sample construction in Appendix A.2. Patient covariates are reported in Table 1. The Charlson score is a measure of health: it aggregates seventeen comorbidities based on severity from diagnoses listed in all claims in the past twelve months prior to surgery into a score from 0 to 24, with a larger score indicating poorer health. Patients undergoing CABG surgery exhibit an average Charlson score of 3.41, indicating moderate to high health risk. 40% of patients in the sample have had an acute myocardial infarction (AMI) and 42% of them received a diagnosis of congestive heart failure (CHF) in the year prior to surgery. Consequently, short-term mortality after CABG surgery is non-null: the average mortality after CABG surgery is 4% within 30 days and 5% within 60 days in the final sample.

The fact that surgeons operate on patients at multiple hospitals within the same year, which I call "multi-homing," is a sizeable addition to the usual variation provided by job movers in employer-employee matched data. Panel A.1a of Appendix Figure A.1 shows that multi-homers represent 19% of surgeries in the final sample, as compared to 25% for traditional movers. Overall, 45% of surgeries in the final sample are performed by "traditional movers" or "multi-homers". Furthermore, multi-homing happens for a susbtantial share of a surgeon's activity as depicted in Appendix Table A.3, such that it is not a marginal practice at the surgeon level.

Multi-homers are more likely to treat younger, lower-income patients in large-population ZIP codes, who tend to be sicker at baseline as reported in Appendix Table A.2. 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.

[8]The total number of observations is larger than the total number of patients for two reasons. First, 602 patients received CABG surgery more than once in the 2011-2017 final sample. Second, 16.3% of surgeries exhibit more than one performing surgeon in the final sample. This happens when the primary surgeon is assisted by another surgeon during the surgery, and the latest bills as such for the procedure. When more than one surgeon operates on a patient, I assign the patients' outcome from the surgery to both surgeons assuming these observations are independent.

2.3 Motivating facts

Panel 1a of Figure 1 illustrates what we already know from the literature: even after risk- and noise-adjustment, hospitals and surgeons vary substantially in their average 30-day survival rates for CABG surgery. The standard deviation across hospitals represents about 2 percentage points of 30-day survival, which represents about half of the average 30-day mortality for CABG surgery. The same holds for surgeons, who also exhibit substantial variation in their 30-day survival rates for CABG surgery. The standard deviation across surgeons also amounts to about 2 percentage points of 30-day survival.

Panel 1a of Figure 1 also illustrates what we *do not* know in the healthcare literature: whether surgeons sort across hospitals and how much this sorting drives the depicted heterogeneity across providers. In this figure, the average 30-day risk-adjusted survival at a hospital includes the effects of surgeons who operate there. Conversely, the average 30-day risk-adjusted survival for a surgeon includes the effects of hospitals they practice at. Panel 1a of Figure 1 does not allow to *separate* the effects of surgeons from hospitals, though these measures are the standard measures of provider "quality" used in practice and in the literature. Separating these effects is crucial to understand the determinants of this heterogeneity, the same way separating worker and firm effects as well as worker sorting across firms is to the labor literature examining wage inequality. The direction and magnitude of sorting will impact the depicted heterogeneity in Panel 1a of Figure 1 and will inform our understanding of the determinants of the observed variation in patient outcomes across providers. For example, if high-survival surgeons sort into high-survival hospitals, a determinant of hospital "quality" becomes the skills of the performing surgeons they are able to attract.

Leveraging external measures of quality for hospitals, I find evidence that surgeons do sort substantially across hospitals. Panel 1b of Figure 1 shows the percentage of surgeries performed by each surgeon quartile across hospitals ranked using the CMS overall five-star ratings.[9] Surgeon quartiles are formed using the risk- and noise-adjusted measures depicted in Panel 1a of Figure 1. A greater proportion of surgeries are performed by high-survival surgeons at higher-ranked hospitals, suggesting positive assortative matching. Appendix Table A.5 summarizes this result, with a correlation between surgeon's estimated risk- and noise-adjusted survival and CMS overall five-star rating of about 0.11. Appendix Table A.5 also extends this reduced form result across alternative external measures of hospital quality. Across all hospital quality measures, high-survival surgeons for CABG surgery tend to perform at better performing hospitals, for CABG surgery when measured for single-homers only for hospitals in a "split-sample" strategy in column (2) with a correlation of about 0.2,

[9] The CMS overall five-star rating aggregates quality measures across five areas of quality into a single rating for each hospital going from one (lowest) to five (highest).

and for five alternative diagnoses as measured in Chandra, Dalton, and Staiger (2023) in columns (3) to (7).[10] Therefore, sorting is likely to play a role in determining differences in performances across hospitals.

The existence of sorting is intricately linked to the existence of complementarity or substitutability between surgeon and hospital effects. Positive assortative matching would tend to maximize patient survival if surgeon and hospital effects are complements in the production of survival, but it would minimize it if surgeon and hospital effects are substitutes. This question is all the more important in healthcare since provider payments are regulated, in the U.S. but also in many other countries, and providers are paid separately.[11] Whether surgeon and hospital effects are complements, substitutes, or perfectly separable is an empirical question, which determines the direction and magnitude of the impact of surgeon sorting across hospitals on patient outcomes.

Contrary to the labor literature that finds little evidence for "match effects" in wages between workers and firms (Card, Heining, and Kline, 2013), I find suggestive evidence that these match effects play a more substantial role in the context of CABG surgery in Panel 1c of Figure 1.[12] I group surgeons in above- versus below-median 30-day survival groups using the risk- and noise-adjusted measured depicted in Panel 1a of Figure 1. Importantly, I use the CMS overall five-star rating as measure of hospital quality, which is computed using different measures and data than the one I use to classify surgeons. If surgeon and hospital effects are perfectly separable or additive, the slope for the high- and low-survival surgeons should be identical.[13] The slope is greater visually for low-survival surgeons in Panel 1c of Figure 1, suggesting that the marginal return of a higher-ranked hospital is larger for low-survival surgeons. Table 2 summarizes this result for alternative measures of hospital quality, all measured "out-of-sample": the slope for low-survival surgeons is consistently greater and statistically significant. Overall, surgeon and hospital effects not only appear to be non-additive, but they also point toward the existence of substitutability between surgeon and hospital effects.

This evidence motivates both the research question and the empirical strategy delineated

[10] The hospital quality measures from Chandra, Dalton, and Staiger (2023) are risk-adjusted, noise-adjusted, and account for differences in hospital volume and quality drift over time.

[11] In the U.S., hospitals and surgeons are paid separately a pre-agreed reimbursement rate. Hospitals are also constrained in their ability to financially attract surgeons by the federal anti-kickback statute.

[12] Recent advances in the literature find some evidence of substitutability between worker and firm effects, but they remain quantitatively modest (Bonhomme, Lamadon, and Manresa, 2019).

[13] I do not depict event-studies because the identifying variation also comes from "multi-homers" who practice at multiple hospital in the same time period and because 30-day survival is a noisier outcome than wages. A better visual representation of non-separability in my context is to compare the slopes across surgeon types as in panel 1c of Figure 1, which identifies the cross-partial derivative of the production function as clearly delineated in Section 3.

in the next Section. The sorting of surgeons across hospitals appears to be substantial but it is not commonly taken into account in performance measures for healthcare providers. Furthermore, there appear to exist match effects between surgeon and hospital effects which need to be estimated to evaluate the impact of surgeon sorting on patient outcomes. This paper develops an empirical strategy to estimate the direction and magnitude of sorting as well as the direction and magnitude of the match effects between surgeon and hospitals effects directly on the relevant dimension of heterogeneity, namely 30-day survival for CABG surgery for both surgeons and hospitals.

3 Empirical strategy

I leverage the existence of multi-homers and traditional movers in a two-way fixed-effects approach including interactions between surgeon and hospital unobserved heterogeneity. Using k-means clustering in the spirit of Bonhomme and Manresa (2015); Bonhomme, Lamadon, and Manresa (2019, 2022) but extending it to group on *both* sides of heterogeneity, I first classify hospitals and surgeons into groups.[14] I next show how to separately identify surgeon group, hospital group effects as well as match effects in a second step using a non-parametric production function of survival. 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

[14] Such a two-sided grouped fixed-effect estimator is notably investigated with simulations in Bonhomme (2020).

not a latent variable model. The function g describes how surgeon and hospital heterogeneity and patient observables combine to produce patient survival. Assuming $\mathbb{E}[\epsilon_{ijht}|\alpha_j, \psi_h, X_{it}] = 0$, the conditional expectation of patient survival is equal to the production function g.

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

I seek to estimate, rather than assume, the existence and magnitude of complementarities between surgeons and hospitals. 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 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 2 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 2a, 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 2b 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 substitutes in Figure 2a: the slope is larger for lower-α_j surgeons, such that the return to allocating surgeons to high-ψ_h hospitals is larger for low-α_j surgeons.

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

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

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

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

There are two main challenges to recover complementarities between surgeon and hospital quality as well as the sorting of surgeons across hospitals. First, surgeons are not observed at all hospitals, which poses an identification problem when evaluating counterfactual allocations of surgeons to hospitals. Furthermore, 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 4% in the sample, while the mean and median number of surgeries per surgeon is 45 and 37, respectively, against 95 and 69 for hospitals, as depicted in Appendix Table A.4. 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.[15] I address these issues by grouping surgeons and hospitals in a first step in the spirit of Bonhomme, Lamadon, and Manresa (2022), but extending it to grouping *both* sides. I cluster providers using their average risk-adjusted survival as a proxy for their individual quality using a k-means algorithm, as detailed in the next subsection. Using this classification, I can next recover grouped fixed effects estimates for surgeons' and hospitals' types as well as their interactions in a second step.

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

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

This assumption implies that $\mathbb{E}[\epsilon_{ijht}|D_{ijht}, \alpha_j, \psi_h, X_{it}] = \mathbb{E}[\epsilon_{ijht}|\alpha_j, \psi_h, X_{it}] = 0$ in equation (2). This is the exogenous network assumption, common in two-way fixed-effects model as in Abowd, Kramarz, and Margolis (1999). The probability for a patient to be treated at a hospital h by a surgeon j can depend on the individual hospital and surgeon heterogeneity and patient observables, but it cannot depend on unobservables at the patient level that have an impact on their survival Y^*_{ijht}. In the case of patient outcomes, this assumption may be violated if patients select into hospitals or surgeons on unobservables. I address this concern using distances to providers as instruments to identify selection on patient unobservables. I delineate this approach in subsection 3.3.

[15] Noise in the individual fixed effects also magnifies the bias on the sorting parameter, called limited mobility bias, which has been shown to be quantitatively important in the literature (Kline, Saggio, and Sølvsten, 2020; Bonhomme et al., 2023).

3.2 Classifying hospitals and surgeons

To both identify match effects and reduce the dimensionality of the fixed effects in the two-sided heterogeneity model, I group individual surgeons and hospitals.[16] This consists in a two-step grouped fixed-effects approach similar to Bonhomme, Lamadon, and Manresa (2019, 2022), but it extends it by clustering both sides, i.e., hospitals *and* surgeons.

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

$$\mathbb{E}[Y_{ijht}|\psi_h] = \int \underbrace{g(\alpha_j, \psi_h)}_{\text{Production function}} \underbrace{f(\alpha_j|\psi_h)}_{\text{Sorting}} d\alpha_j \tag{3}$$

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

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

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

With positive assortative matching, $\mathbb{E}[\alpha_j|\psi_h]$ is increasing in ψ_h, so average survival at hospital h is increasing in its quality and identifies the individual hospital type. With negative assortative matching, $\mathbb{E}[\alpha_j|\psi_h]$ is decreasing in ψ_h, which adds an "offsetting" effect to the individual quality of the hospital. In that case, average survival is not necessarily increasing in ψ_h and identification of the individual type from average survival may fail. Two different hospital types may exhibit the same average survival, which prevents from separating their individual types from their average survival. Consequently, identification of the individual types from individual moments requires that these individual moments are

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

increasing in the provider's type. This may not always be true under negative assortative matching. I will use several alternative moments to evaluate the robustness of the results to this assumption.

The goal of this classification is to cluster hospitals and surgeons into groups capturing their individual quality. Average risk-adjusted survival from CABG surgery is used as a measure of quality in practice in report-cards for surgeons and hospitals. It is publicly reported and used in CMS quality ratings for hospitals, for example. It is also used in the literature evaluating hospital and surgeon quality (Huckman and Pisano, 2006; Ghaferi, Birkmeyer, and Dimick, 2009; Kolstad, 2013). I use provider-level risk-adjusted survival as a proxy for individual quality to group surgeons and hospitals into quality groups. I compute it following the methodology used in the literature as the ratio of observed over predicted survival for the provider as delineated in Appendix A.4.

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. I use both standardized 30- and 60-day average risk-adjusted survival to form the groups. 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.6. 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.[17]

When the production function is monotonic and positive assortative matching exists, I can recover individual surgeon and hospital types from the grouped fixed effects. Using Monte Carlo exercises delineated in Appendix A.7, I show how I can recover or bound the sorting parameter using the two-sided grouped fixed effect estimator in my data. Because the classification for surgeons is based on moments that includes the effect of hospitals, and vice versa, correlated classification error may lead to over-estimate the true sorting parameter, as shown in Appendix Table A.6.

[17] Grouping surgeons and hospitals also addresses biases in the variance and covariance estimates in two-way fixed-effects models (Bonhomme, Lamadon, and Manresa, 2019, 2022). 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.

Leveraging the structure of the data, in which a bit more than half of surgeries are performed by single-homers while the rest are performed by surgeon-movers, I can form moments for hospitals using single-homer surgeries only while considering only surgeon movers in the surgeon classification and estimation. This classification breaks the correlation in the classification error between surgeons and hospitals, since the groups are formed using different observations. I show in Monte Carlo exercises in Appendix Table A.6 how such classifications tend to under-estimate the true sorting parameter due to increased noise, since we exclude single-homers surgeries, hence forming a lower bound for the sorting parameter. In the case of negative assortative matching, the estimated sorting parameter will be biased toward zero due to classification error, since the moments do not identify provider types. This robustness check will play a crucial role below to test whether the direction of sorting is driven by correlated classification error.

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 depict interactions 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:[18]

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

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 5.2.

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

[18]Note that equation (6) assumes perfect separability between provider value-added and patient health, hence estimating surgeon and hospital effects on the average patients. Dahlstrand (2021) estimates the non-separability between physician value-added and patient health and finds substantial complementarities in the context of general medicine. Complementarities between patient health and provider value-added are beyond the scope of this paper that focuses on surgeon sorting across hospitals.

$$D_{ijht} = \mathbb{1}\{u_{ih} \geq u_{ih'},\ \forall h'\}$$
$$\text{with } u_{ih} = \delta_h - \tau \ln(d_{ih}) + \eta_{ih} \tag{7}$$

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.[19] The outside option is defined as choosing a hospital outside of the patient's hospital referral region (HRR) of residence.

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

$$\mathbb{E}[\epsilon_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, X_{it}, \eta_{i1}, ..., \eta_{iH}, D_i = h] = \sum_{s\in\mathcal{H}} \phi_s(\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta) \tag{8}$$

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

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

$$\begin{aligned}\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, &\kappa_{l(j)k(h)}, \gamma_t, X_{it}, \eta_{i1}, ..., \eta_{iH}, D_i = h] = \\ &\alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \beta X_{it} + \gamma_t + \sum_{s\in\mathcal{H}} \phi_s(\eta_{is} - \mu_\eta) + \varphi(\eta_{ih} - \mu_\eta)\end{aligned} \tag{9}$$

Integrating equation (9) over the unobserved demand shocks $\eta_{i1}, ..., \eta_{iH}$ delivers the estimating equation

$$\begin{aligned}\mathbb{E}[Y_{ijht}|\alpha_{l(j)}, \psi_{k(h)}, &\kappa_{l(j)k(h)}, \gamma_t, X_{it}, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \\ &\alpha_{l(j)} + \alpha_h + \kappa_{l(j)k(h)} + \beta X_{it} + \gamma_t + \sum_{s\in\mathcal{H}} \phi_s \theta_{is}(h) + \varphi\theta_{ih}(h)\end{aligned} \tag{10}$$

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

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

and $\hat{p}_{is}$ is the predicted probability for patient i to choose hospital s from the demand model in equation (7). Derivations are included in Appendix A.8. Note that the control function

[19] 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.

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

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

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

I estimate the demand model market by market, using hospital referral regions (HRRs) as market definitions.[20] 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:

$$\begin{aligned} Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}, \kappa_{l(j)k(h)}, \gamma_t, \ln d_{i1}, ..., \ln d_{iH}, D_i = h] = \\ \alpha_{l(j)} + \psi_{k(h)} + \kappa_{l(j)k(h)} + \sum_p \beta_p X_{it,p} + \gamma_t + \sum_{s \in \mathcal{H}} \phi_s \hat{\theta}_{is}(h) + \varphi \hat{\theta}_{ih}(h) \end{aligned} \tag{11}$$

Distance to the hospital is an excluded instrument here, since it is excluded from the last step: distance to the hospital can only have an impact on patient survival through the choice of hospital. Since CABG surgery is usually performed in non-emergency settings, distance to the hospital as time to treatment should have no impact on patient survival.[21] 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 5.2.

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

[20] 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.

[21] 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.

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.

4 Estimated parameters

I summarize the estimated parameters from the empirical model. I first describe the provider groups formed in the first step. Next, I show that recovered parameters from risk-adjustment are sensible, and that recovered provider effects are correlated with external observable measures of quality.

4.1 Grouped types of surgeons and hospitals

In the first step, I group surgeons and hospitals using k-means clustering on average risk-adjusted survival as delineated in Section 3.2. For the main specification, 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. In Subsection 5.2, I show that results are robust to alternative number of groups. After the first step, the relevant source of variation to identify provider group effects now comes from surgeon *groups* practicing at different hospital *groups*. The number of surgeons observed at multiple hospital groups remains large, as depicted in Panel A.1b of Appendix Figure A.1.

There exist a substantial amount of variation in survival across groups. Appendix Figure A.2 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. Groups are also of varying sizes. Note that the ordering of the groups displayed in Appendix Figure A.2 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. Importantly, note that the average risk-adjusted survival for hospitals and surgeons still includes the combination of the hospital and surgeon effects in Appendix Figure A.2.

4.2 Estimated parameters are sensible

Risk-adjustment parameters. The coefficients on patient observables for risk adjustment are sensible. Appendix Table A.7 reports coefficients on patient covariates from estimating equations (6) and (11). In both specifications, older and sicker patients are less likely to survive 30 days after surgery. Women are also less likely to survive after CABG surgery, consistent with several studies in the medical literature (Zwischenberger, Jawitz, and Lawton, 2021).

Correlations with external measures of quality. Estimated effects for surgeon and hospital groups are correlated with external measures of quality for these providers. Figure 4a reports correlations between estimated surgeon effects and individual surgeon observables. High-survival surgeons tend to exhibit greater Medicare volumes: I find that high-survival surgeons tend to exhibit larger volume in CABG surgery, in surgery, and total activity overall within Medicare. This 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). "Tenured" experience measured as the number of years since medical school graduation is not correlated with the estimated survival effects, which is consistent with previous evidence from Birkmeyer et al. (2013) notably. In terms of observable place of practice, high-survival surgeons tend to practice at teaching hospitals and high-volume hospitals in terms of Medicare discharges, overall and for CABG surgery.

Correlations of estimated hospital group effects are reported in Figure 4. Panel 4a correlates the estimated hospital group effects for CABG surgery with risk-, noise- and drift-adjusted hospital 30-day mortality effects estimated in Chandra, Dalton, and Staiger (2023) for five other diagnoses. High-survival hospitals for CABG surgery tend to exhibit lower mortality in these other diagnoses, though some are not statistically significant.[22]

Panel 4b correlates the estimated hospital ranks with observables and external rankings. High-rank hospitals using my estimated effects for CABG surgery tend to be high-volume hospitals, measured using Medicare discharges. High-rank hospitals tend to be teaching hospitals, though this is not statistically significant, but hospital size measured as the number of beds is not correlated with my estimated rankings. Using external rankings, high-rank hospitals are more likely to be mentioned in the top hospitals according to the 2017 U.S. World News report, both overall and for cardiology, though the latest is not statistically

[22]Though consistent in sign, there are many reasons why hospital performances may not perfectly align across these measures and diagnoses. Hospitals may be better at treating some diseases or diagnoses than others. Additionally, these diagnoses tend to be emergency diagnoses, which likely involve different teams and departments. Hip & knee replacement, the elective diagnosis available, indeed exhibits a stronger relationship. I also use the Chandra, Dalton, and Staiger (2023) measures for the latest year available in their data, 2014, while my estimates span 2011 to 2017.

significant at 5%. High-ranks hospitals also tend to exhibit higher CMS five-star overall ratings.

No evidence of triaging. 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 Appendix Figure A.3, I find no systematic relationship between patient covariates and hospital rankings within surgeons. The evidence is similar for hospitals: there is no evidence that hospitals systematically "triage" their sickest patients toward higher-survival surgeons as shown in Appendix Figure A.4.

5 Substitutability and sorting results

In this Section, I describe the main results of the paper: surgeon and hospitals are substitutes in the production function of survival for CABG surgery and high-survival surgeons sort into high-survival hospitals. I evaluate the robustness of these results along several dimensions, notably allowing for selection of patients on observables and exploring results using out-of-sample validations. I find suggestive evidence in favor of "failure-to-rescue", a mechanism emphasized in the medical literature, as a mechanism for substitutability between surgeons and hospitals.

5.1 Main specification

To investigate the differential returns to allocating high- and low-survival surgeons to alternative hospitals, I report the average predicted 30-day survival for the average patient for each hospital-surgeon group interaction separately in Figure 5a. This figure is the empirical counterpart to theoretical graphs in Figure 2. As delineated in Section 3.1, 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.

Surgeon and hospital quality are substitutes in the production function of 30-day survival for CABG surgery. As depicted in Figure 5a, 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 3. 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 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.[23]

I find positive assortative matching where high-survival surgeons sort into high-survival hospitals. Figure 5b 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 tend to be performed by high-survival surgeons, while surgeries at low-survival hospitals tend to be performed by low-survival surgeons. Table 3 reports the correlation between surgeon and hospital effects, which is positive at about 0.4. Variance decompositions in Appendix Table A.8 suggest that sorting explains between 20 and 40% of the explained variance in survival net of covariates.

5.2 Robustness

5.2.1 Patient selection on unobservables

I report results from the empirical strategy allowing for selection on unobservables 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. I examine whether the distance to the hospital is predictive of hospital choice in Panel A.8b of Appendix Figure A.8. 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.[24] 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

[23]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.

[24]Patients tend to be treated at hospitals close to their ZIP code of residence, as depicted in Appendix Figure A.8a. 21% of surgeries are performed outside of the patient's HRR. 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). The relationship between the chosen hospital and distance also appears log-linear, supporting the functional form assumption in equation (7).

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 A.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 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 A.8c of Appendix Figure A.8 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.[25]

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 Appendix Table A.7. The estimated parameter for selection on gains $\hat{\psi}$ is negative but 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 A.8d.

Results are robust when using the control function approach. As shown in Appendix Figure A.9, surgeon groups and hospitals groups are substitutes in the production function of survival. Estimated slopes in column (2) of Table 3 for each surgeon group are similar to the selection on observables approach. Positive assortative matching is also similar in magnitude, with a correlation between hospital and surgeon group effects of 0.47. The variance decomposition of a model without interactions between surgeon and hospital groups also leads to similar conclusions, as reported in Appendix Table A.8. Overall, the main results are not altered by allowing for patient selection into hospitals on unobservables.[26]

[25] 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.

[26] Note that, even though the sorting and substitutability results remain extremely similar, the control function approach does capture additional selection into providers. Selection into providers likely does not change the *rankings* of providers, crucial to my coefficients of interest here.

5.2.2 Correlated classification error

Split-sample approaches.

To alleviate correlated classification error concerns from the classification stage, I adopt a split-sample strategy that leverages the existence of single-homers and surgeon movers.[27] I classify hospitals using average survival using only surgeries performed by single-homers, which includes about 55% of surgeries in the main sample. I next classify surgeon-movers only, and estimate the two-way fixed effect strategy for surgeries performed by surgeon-movers only. In this approach, observations used to classify hospitals are different observations from the ones used to classify surgeons and to estimate the provider effects, hence breaking any mechanical spurious correlation between surgeon and hospital group assignment.

Results are reported in the third column of Table 3. While results are noisier due to the (large) reduction in the estimation sample size, I still find evidence for substitutability between surgeon and hospital effects and positive assortative matching. In light of the Monte Carlo exercises reported in Appendix Table A.6, this result validates positive assortative matching.

To alleviate any remaining concerns related to a correlation between estimated provider effects and the error term, I next estimate a more drastic split-sample strategy. I add two more years of data to the sample, 2018 and 2019, and split the sample in two sub-samples: 2011-2015 and 2016-2019. I classify hospitals using single-homers only on the 2011-2015 subsample, and classify surgeons and estimate the second step using movers only for 2011-2015. I then use the estimated effects and rankings from the 2011-2015 subsample to evaluate the slopes and correlation between provider effects in the 2016-2019 subsample.

Results are extremely noisy since this approach reduces the sample size to about twelve thousand observations. In spite of this, results still suggest substitutability: the slope for the two bottom groups is much larger than for the top three ones, though noise prevents from making sound statistical inference. Furthermore, the correlation between surgeon and hospital effects also points toward positive assortative matching. These results are reassuring about correlated classification error not driving the main results, though they should be taken with caution due to the large drop in sample size.

Other outcomes. I explore results when using alternative health outcomes, much less noisy than 30-day survival, to additionally explore the impact of correlated classification error on the main results. Length of stay and hospital spending are alternative measures

[27] Results from Monte Carlo analyses reported in Appendix Table A.6 show the intuition behind this robustness check.

of surgeon and hospital performances, and we expect them to be negatively correlated with survival rates for surgeons based on prior literature (Birkmeyer et al., 2013): high-survival surgeons would achieve shorter length of stay with lower spending for patients.

For both additional outcomes, conclusions are similar.[28] I find evidence of substitutability as reported in Panel B of Table 3.[29] Low-performing surgeons, with longer length of stay and higher hospital spending, tend to perform better at high-performing hospitals, and more so than better-performing surgeons. The correlation between hospital and surgeon effects also point toward positive assortative matching.

I also perform the first split-sample strategy delineated above but on these other outcomes, so that different observations are used to classify surgeons and hospitals. This exercise leads to a similar drop in sample size, but noise has a lesser impact due to the nature of these outcomes. Results reported in Appendix Table A.13 show that the substitutability and sorting results are robust to this classification error concern for these other outcomes.

Finally, I also estimate sorting using an alternative method altogether on these other outcomes, namely using individual fixed effects and the leave-out correction from Kline, Saggio, and Sølvsten (2020). This approach excludes interactions between surgeon and hospital fixed effects. These results are reported in Appendix Table A.10. As expected, the plug-in estimate as in Abowd, Kramarz, and Margolis (1999), i.e. the covariance of estimated individual surgeon and hospital fixed effects, is negative. Recent work has demonstrated the sensitivity of these estimates to the incidental parameter problem and shown the importance of bias correction when evaluating worker sorting across firms using wages (Kline, Saggio, and Sølvsten, 2020; Bonhomme et al., 2023). When using the leave-out correction in my context, the estimated covariance becomes positive and points to strong positive assortative matching. These results are consistent with results using the grouped fixed effect estimator without interactions in Appendix Table A.8. Overall, these results further support the existence of positive assortative matching.

5.2.3 Additional robustness checks

Alternative number of groups. I investigate the robustness of results when varying the number of surgeon and hospital groups in Appendix Table A.12. The slope for low-survival surgeon remains larger than for high-survival surgeons, whether one increases the number of

[28]Note that there are many dimensions of skill and quality, and I find that they do not perfectly align across these outcomes. In particular, the correlation between the fixed effects when using survival versus length of stay or spending are negative for both surgeons and hospitals, so that high-survival providers tend to be short length of stay and low spending. However, these correlations remain small: they are about -0.03 for hospitals and between -0.07 and -0.12 for surgeons.

[29]Note that better performing providers should now exhibit *lower* length of stay and spending, and the slopes are now negative. The conclusion for the hypothesis tests are reversed.

surgeon groups, hospital groups, or both. Positive assortative matching remains of similar magnitude.

Alternative classifications. I examine the robustness of results to using alternative classifications in Panel D of Appendix Table A.13. In column (4), I explore results when adding other outcomes to the k-means algorithm: length of stay and hospital spending. In column (5), I use simple quintiles of risk-adjusted survival for hospitals and surgeons, instead of the k-means algorithm. In column (6), I use the noise-adjusted estimates for surgeon and hospital effects as shown in Figure 1 and delineated in Appendix A.5 to group surgeons and hospitals in the k-means algorithm. Results remain robust to these alternative classification methods: the slope is decreasing in the surgeon's rank and provider effects are positively correlated.

Alternative production function. I examine results when using a logit model, instead of a linear probability model, in column (7) of Appendix Table A.13.[30] Again, results remain virtually unchanged by this change of functional form.

Excluding emergencies. Because emergency CABG surgery may be different from elective CABG surgery, I investigate robustness of results when excluding CABG surgery performed in an emergency setting in column (8) of Appendix Table A.13. Results remain extremely similar on this subsample.

5.3 Mechanism: failure-to-rescue

While substitutability between surgeons and hospitals in terms of patient survival may appear surprising at first, 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. 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

[30] The logit production function would assume substitutability between surgeon and hospital effects when omitting interactions. I include interactions to flexibly estimate the cross partial derivative.

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.

To gain insights in the plausibility of "failure-to-rescue" to be driving the susbstitutabiltiy result, I run the model for two alternative patient outcomes, length of stay and hospital spending per patient, but keeping the surgeon and hospital groups based on risk-adjusted survival. Length of stay and hospital spending are measures of the intensity of care here across the hospital ranks, within surgeon group. Results are reported in Appendix Figure A.5. Low-survival surgeons tend to exhibit longer length of stay and higher spending per patient, so that low-survival surgeons also exhibit worse performances when examining alternative patient outcomes such as length of stay or hospital spending. In addition, when examining how surgeon performances on these other outcomes vary across hospital ranks, I find that lower-survival surgeons tend to exhibit longer length of stay and higher spending in higher-survival hospitals. This evidence is consistent with more intensive care given to patients of low-survival surgeons at high-survival hospitals, which then translates into higher survival for these patients. Overall, these findings are consistent with prior literature on physician skill and "failure-to-rescue".

Generalizability. Whether results for CABG surgery are generalizable to other surgeries, in particular in terms of 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.

Discussion: payments in healthcare. These results highlight that the current sorting of surgeons across hospitals does not maximize aggregate patient survival. Considering the amount of regulation in this industry, including how providers are paid and how their relationships are regulated, this result appears less surprising. Hospitals and surgeons are paid separately a predetermined amount. In addition, hospitals cannot offer financial incentives to surgeons if they are not employed by the hospital, which is often the case for cardiac surgeons. The federal anti-kickback statute prohibits hospitals from paying doctors for referrals, with the goal of removing financial incentives from doctors' clinical decisions. Overall, this means that low-survival hospitals will not be able to offer higher compensations to attract high-survival surgeons that they do not employ. This suggests that payments that incorporate the match effects, value-added based payments, could be a potential avenue for policies aiming at maximizing patient survival. This idea is not new: it has notably been proposed in the context of teacher payments in the education literature (Hoxby, 2014).[31] [32]

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. More work is needed to determine the respective role of provider payments and these other dimensions on the current sorting of surgeons across hospitals.

6 Counterfactual allocations of surgeons to hospitals

I use the estimated provider effects and production function estimates to learn about the impact of surgeon sorting on both aggregate patient survival but also the geography of healthcare provider quality. First, I document simple facts about provider sorting across

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

[32] 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.

regions using the estimated effects and regions' baseline observables. Motivated by these facts, I perform two simple counterfactual exercises. I first reallocate surgeons across hospitals nationally, irrespective of their current geographic location, to learn about the impact of surgeon sorting on aggregate patient survival. I also use this exercise to illustrate that such reallocations would benefit less populated areas much more than more populated areas. Second, I reallocate surgeons across hospitals but keeping their region constant, i.e. I reallocate surgeons across hospitals *within* their current region. This counterfactuals shows that a substantial driver of positive assortative matching is due to surgeon sorting across hospitals within their region. It also illustrates baseline differences across regions: more populated areas would benefit more from such reallocations since they tend to have both low- and high-survival providers at baseline.

6.1 Surgeon sorting within versus across regions

I use the estimated provider effects to gauge to what extent positive assortative matching reflects provider location decisions across space. High-survival providers tend to be located in more populated and higher-income locations. Examining the correlation between estimated provider rankings and patient covariates in Appendix Figure A.6, I find that high-survival surgeons and hospitals appear to treat patient from ZIP codes that are higher income and more populated. 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 Panel A.6b of Appendix Figure A.6.

Note that two forces could be driving these facts. First, high-ability surgeons may sort into more populated regions that tend to have high-survival hospitals. Second, larger cities tend to produce higher-quality medical services by leveraging economies of scale, through division of labor or spillovers between providers for example, such that both hospitals and surgeons that locate there will tend to perform better than in smaller cities, as emphasized in Dingel et al. (2023).[33]

While separating these two explanations is beyond the scope of this paper, partial equilibrium national reallocations of surgeons across hospitals are informative about the current sorting of surgeons on two main dimensions. First, it allows to evaluate the sensivity of aggregate patient survival to surgeon sorting, i.e. to quantify the magnitude of the substi-

[33]The empirical strategy in this paper uses a fixed-effects approach, such that provider effects are fixed over time in the sample. It does not capture changes in provider effects over time after a move, which could allow to separate both explanations. General equilibrium effects of surgeon reallocations across regions may be very different when the second force dominates. Taking into account the "drift" in the two-way fixed effects strategy proposed here would be interesting but is beyond the main goal of this paper.

tutability between surgeons and hospitals in the production function of survival for CABG surgery. Second, by comparing gains across regions along the population gradient, we can further explore the impact of surgeon sorting on differences in predicted patient survival across regions. I study these counterfactuals in the next subsection.

In addition, 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 Appendix Figure A.7. I find that a substantial fraction of HRRs exhibit strong positive assortative matching, suggesting that surgeon sorting is substantial within regions too. This motivates a second partial equilibrium counterfactual exercise that reallocates surgeons across hospitals, but keeping the surgeon's geographical location constant, detailed in Subsection 6.3.

6.2 National reallocations of surgeons

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.[34] For the random reallocation, I randomly assign surgeon groups to patient-hospital pairs, keeping the number of surgeries performed by each surgeon group constant.[35] For the negative assortative matching reallocation, I reallocate surgeons to patient-hospital pairs so that surgeries at the lowest-survival hospital group are performed by surgeons from the highest survival group until all surgeries available at this hospital group are taken or until all surgeries performed by the highest-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 newly assigned surgeon groups.

[34] The main goal of this exercise is to evaluate the strength of the 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.

[35] Note that reallocating patients across surgeon and/or hospital groups would have no impact on aggregate survival since I assume perfect separability between surgeon and hospital value-added and patient observables and unobservables in equations (6) and (11).

As reported in Table 4, randomly reallocating surgeon groups to hospital-patient pairs nationally decreases average 30-day mortality by about 3 deaths per thousand patients, a 7% decrease in 30-day mortality compared to the current positive sorting. It also reduces the dispersion in 30-day mortality across patients by 8%. Consistent with the existence of substitutability between surgeons and hospitals, moving away from positive assortative matching is beneficial for patients in terms of 30-day mortality. 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 26% decrease in 30-day mortality compared to the current positive sorting. Furthermore, the dispersion in 30-day mortality across patients would be reduced by 31% compared to the current sorting.[36]

There are two main takeaways from this simple reallocation exercise. First, the production function of 30-day survival for CABG surgery exhibits a strong substitutability between surgeon and hospital quality. This implies that aggregate patient survival is highly sensitive to the sorting of surgeons across hospitals. Incorporating the impact of various regulations that impact physician sorting, such as provider payments, should be at the heart of healthcare policies aiming at promoting aggregate patient outcomes and individual provider quality.

Second, the baseline location of surgeons is a substantial driver of the inequality in predicted survival across regions along the population gradient. As shown in column (1) of Table 5, more populated regions exhibit lower predicted mortality rate at baseline. National reallocations tend to reallocate high-survival surgeons from higher- to lower-population regions. Combined with substitutability, smaller regions in terms of total population face greater mortality reductions with these reallocations as depicted in Panel 6a of Figure 6. These forces are strong enough to eliminate the baseline population gradient for predicted mortality with random reallocations and negative matching reverses it, as shown in columns (2) and (4) of Table 5.

[36]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 and 800 lives saved per year under random sorting and negative assortative matching respectively. Once again, these numbers should be taken with caution, since they assume away general equilibrium effects. However, they indicate that the substitutability of surgeons and hospitals in the production function of survival for CABG surgery is quantitatively significant.

6.3 Reallocations of surgeons within regions

Since I find evidence that positive assortative matching is substantial within regions, I now 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. Surgeon groups are allocated to patient-hospital pairs based on the number of surgeries performed by this surgeon group at hospitals within the same HRR. This ensures that each surgeon group performs the same number of surgeries within an HRR as they did at baseline.

A substantial fraction of the positive assortative matching result is driven by surgeon sorting *within* regions. Randomly reallocating surgeons within HRRs leads to a decrease of about 1.6 deaths per thousand patients, representing 55% of the mortality gains from a national reallocation. When imposing negative assortative matching, the reallocation leads to about 5 deaths per thousand patients, representing 51% of the gains from a national reallocation. Reductions in the standard deviation follow similar patterns, with between 54% and 65% of the reduction in inequality from a national reallocation being achieved by reallocations within HRRs.

However, gains across regions along the population gradient have a very different profile compared to the national reallocation counterfactuals. As shown in Panel 6b of Figure 6, regions that benefit the most from reallocations within regions are the more populated ones. High-survival providers tend to be located in more populated regions, but these regions also exhibit the full range of providers, from low- to high-survival, so that they offer more scope for gains from reallocations of surgeons across hospitals.[37] Consequently, the negative relationship between predicted mortality and HRR population becomes even stronger than at baseline, as shown in columns (3) and (5) of Table 5. Interestingly, while substantial mortality gains are achieved by reallocating surgeons within regions and while all regions tend to benefit, these gains exacerbate baseline differences in predicted mortality across HRRs in favor of more populated regions.

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

[37] More populated HRRs tend to include more alternative surgeon and hospital groups on average.

and hospitals in the context of CABG surgery, and its consequences for aggregate patient outcomes and the geography of healthcare provider quality.

In the context of survival from CABG surgery, I find that surgeon and hospital quality are substitutes, so that the return to allocating low-survival surgeons to high-survival hospitals is larger than for high-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. Yet, I find positive assortative matching where high-survival surgeons sort into higher survival hospitals.

This detailed quantification of the joint production function between surgeons and hospitals allows to learn about the impact of surgeon sorting on patient outcomes. Using partial equlibrium reallocations, I show that surgeon sorting across hospitals has a large impact on patient 30-day survival from CABG surgery. Furthermore, these estimates allow to learn about the geographic distribution of providers. High-survival providers tend to be co-locate in highly populated regions, but there remain substantial heterogeneity and sorting *within* regions. While reallocating surgeons across hospitals within regions can achieve substantial survival gains, such reallocations would exacerbate the differences in average quality across regions along the population gradient.

In outlining the importance of understanding the production function of hospitals for aggregate outcomes, this paper is in line with the recent literature showing the importance of endogenizing firms' internal organization to explain aggregate market outcomes. Using highly detailed micro-data on worker-level output, Adhvaryu et al. (2020) and Metcalfe, Sollaci, and Syverson (2023) highlight the importance and productivity consequences of firms' internal matching of managers to line workers and managers to retail stores respectively. More work is needed to bring actionable insights for policy in the healthcare setting, such as estimating and incorporating learning from co-workers, taking into account team formation within the hospital, and understanding the drivers behind the estimated sorting of surgeons across hospitals.

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.

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. "Do parents value school effectiveness?" *American Economic Review*, 110(5): 1502–39.

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms." *Econometrica*, 67(2): 251–333.

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.

Angrist, Joshua, Peter Hull, and Christopher Walters. 2023. "Methods for measuring school effectiveness." *Handbook of the Economics of Education*, 7: 1–60.

Badinski, Ivan, Amy Finkelstein, Matthew Gentzkow, Peter Hull, and Heidi Williams. 2023. "Geographic Variation in Healthcare Utilization: The Role of Physicians." National Bureau of Economic Research.

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. 2020. "Econometric analysis of bipartite networks." In *The Econometric Analysis of Network Data*. 83–121. Elsevier.

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, 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*, 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.

Chandra, Amitabh, Maurice Dalton, and Douglas O Staiger. 2023. “Are hospital quality indicators causal?” National Bureau of Economic Research.

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.

Dahlstrand, Amanda. 2021. “Defying distance? The provision of services in the digital age.” *Job Market Paper, London School of Economics and Political Science.*

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 sorting—in 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.”

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. “Sources of geographic variation in health care: Evidence from patient migration.” *The Quarterly Journal of Economics*, 131(4): 1681–1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. “Place-based drivers of mortality: Evidence from migration.” *American Economic Review*, 111(8): 2697–2735.

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.

Gottlieb, Joshua D, Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2025. "The earnings and labor supply of us physicians." *The Quarterly Journal of Economics*, 140(2): 1243–1298.

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*.

Jackson, C Kirabo. 2013. "Match quality, worker productivity, and worker mobility: Direct evidence from teachers." *Review of Economics and Statistics*, 95(4): 1096–1116.

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.

Koedel, Cory, and Jonah E Rockoff. 2015. "Value-added modeling: A review." *Economics of Education Review*, 47: 180–195.

Kolstad, Jonathan T. 2013. "Information and quality when motivation is intrinsic: Evidence from surgeon report cards." *American Economic Review*, 103(7): 2875–2910.

Kolstad, Jonathan T, and Michael E Chernew. 2009. "Quality and consumer decision making in the market for health insurance and health care services." *Medical Care Research and Rreview*, 66(1_suppl): 28S–52S.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2016. "Matching, sorting and wages." *Review of Economic Dynamics*, 19: 63–87.

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

Metcalfe, Robert D, Alexandre B Sollaci, and Chad Syverson. 2023. "Managers and Productivity in Retail." National Bureau of Economic Research.

Molitor, David. 2018. "The evolution of physician practice styles: evidence from cardiologist migration." *American Economic Journal: Economic Policy*, 10(1): 326–56.

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*.

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*, 105–130.

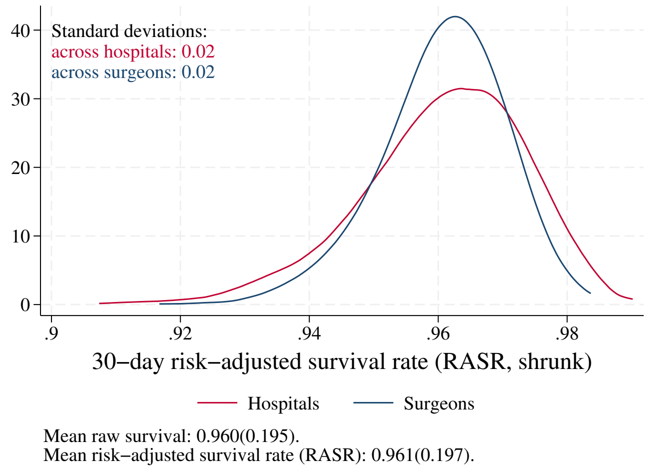
Ward, Sarah T, Justin B Dimick, Wenying Zhang, Darrell Campbell, and Amir A Ghaferi. 2019. “Association between hospital staffing models and failure to rescue.” *Annals of Surgery*, 270(1): 91.

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.

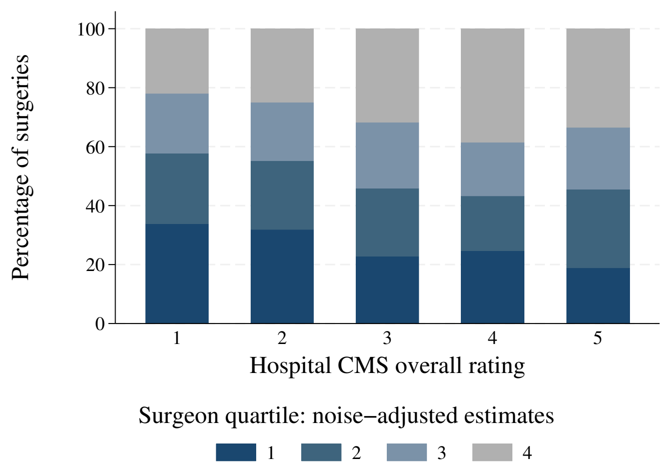
Zwischenberger, Brittany A, Oliver K Jawitz, and Jennifer S Lawton. 2021. “Coronary surgery in women: how can we improve outcomes.” *JTCVS techniques*, 10: 122–128.

Figure 1: Motivation: heterogeneity, non-separability, and sorting

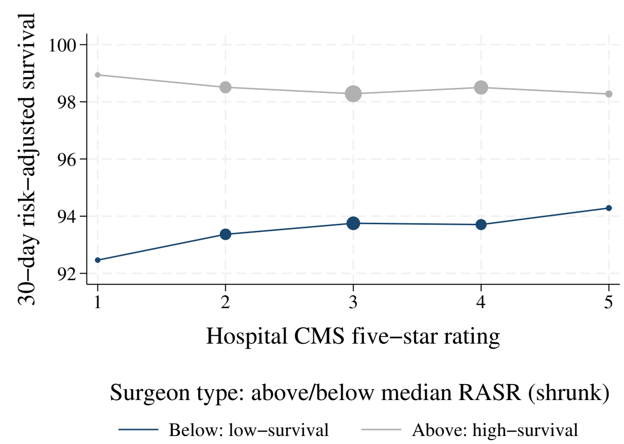
(a) Distribution of 30-day risk-adjusted survival



(b) Suggestive evidence of positive assortative matching

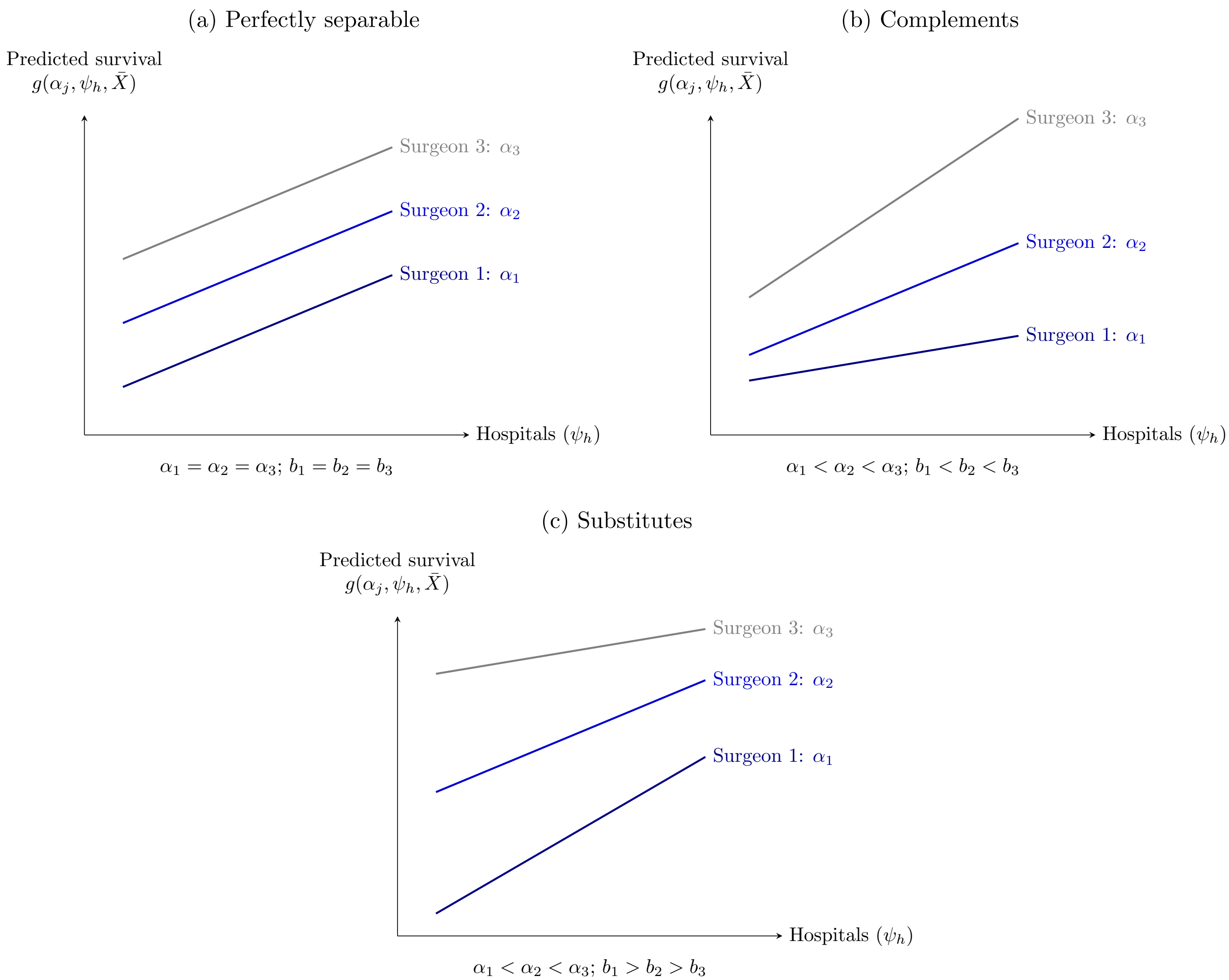


(c) Suggestive evidence of non-separability



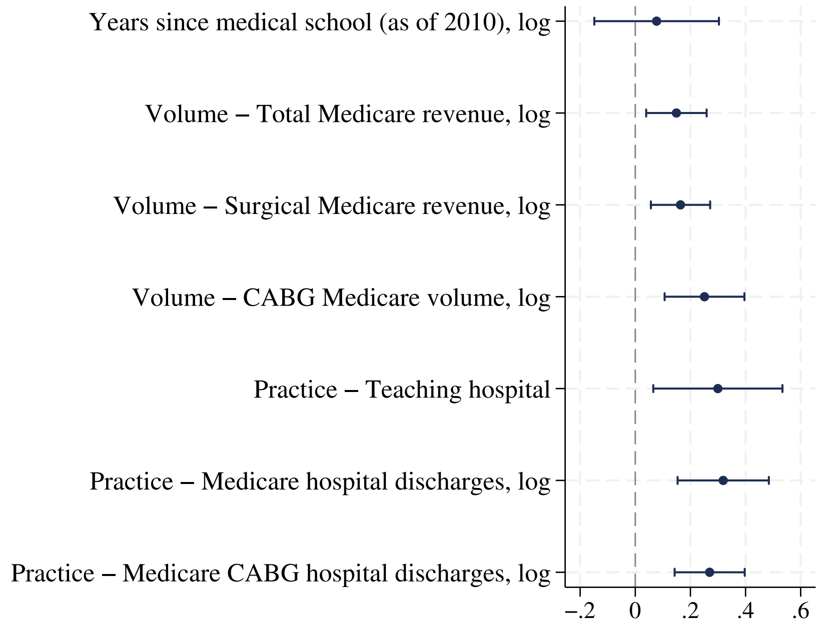
Notes: Panel 1a depicts the distribution of average risk-adjusted 30-day survival (RASR) across surgeons and hospitals, weighted by the number of patients at each provider. These averages are adjusted for measurement error using empirical Bayes shrinkage as detailed in Appendix A.5, which "shrinks" noisily estimated averages toward the mean. Risk-adjustment is performed by predicting 30-day survival using a logit model as delineated in Appendix A.3. Panel 1c displays average 30-day risk-adjusted survival for above- and below-median risk-adjusted survival surgeons across hospitals' CMS five-star rating. Panel 1b describes the percentage of CABG surgeries performed by each surgeon quartile at hospitals in the corresponding CMS five-star rating. Surgeons are classified using the computed risk- and noise-adjusted 30-day survival for CABG surgery displayed in panel 1a, while hospitals are ranked using an externally constructed measure obtained from the CMS Hospital General Information and Complications and Deaths datasets for 2017. 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: 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 2a 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 2b 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 2c details the 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.

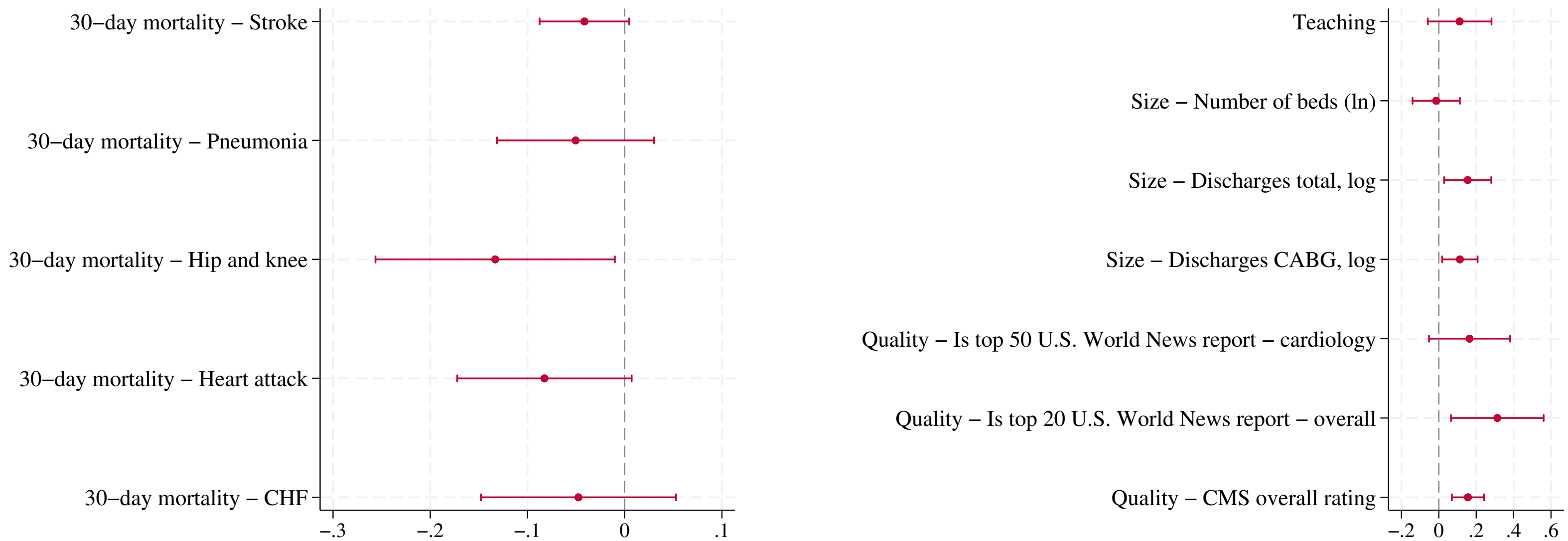
Figure 3: Correlation of estimated surgeon group effects with external measures of skill



Notes: This graph reports the point estimates and 95% confidence intervals from regressions of the surgeon-group estimated effects on surgeon-level covariates. Estimated surgeon-group survival effects are positively correlated with surgeon-specific volumes in Medicare. Surgeons in higher estimated survival groups tend to practice in teaching hospitals, and in hospitals with more total discharges overall and for CABG specifically. The correlation with tenured experience, measured as the time since medical school graduation as of 2010, is not statistically significant. 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, to capture surgeons' average ability "purged" from their sorting across hospitals. 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. Hospital volumes are obtained using the CMS 2017 discharges for fee-for-service Medicare Part A by provider and service. Hospital teaching status comes from the 2017 Dartmouth Atlas data. Confidence intervals are constructed using clustered standard errors at the surgeon level.

Figure 4: Correlation of estimated hospital group effects with external measures quality

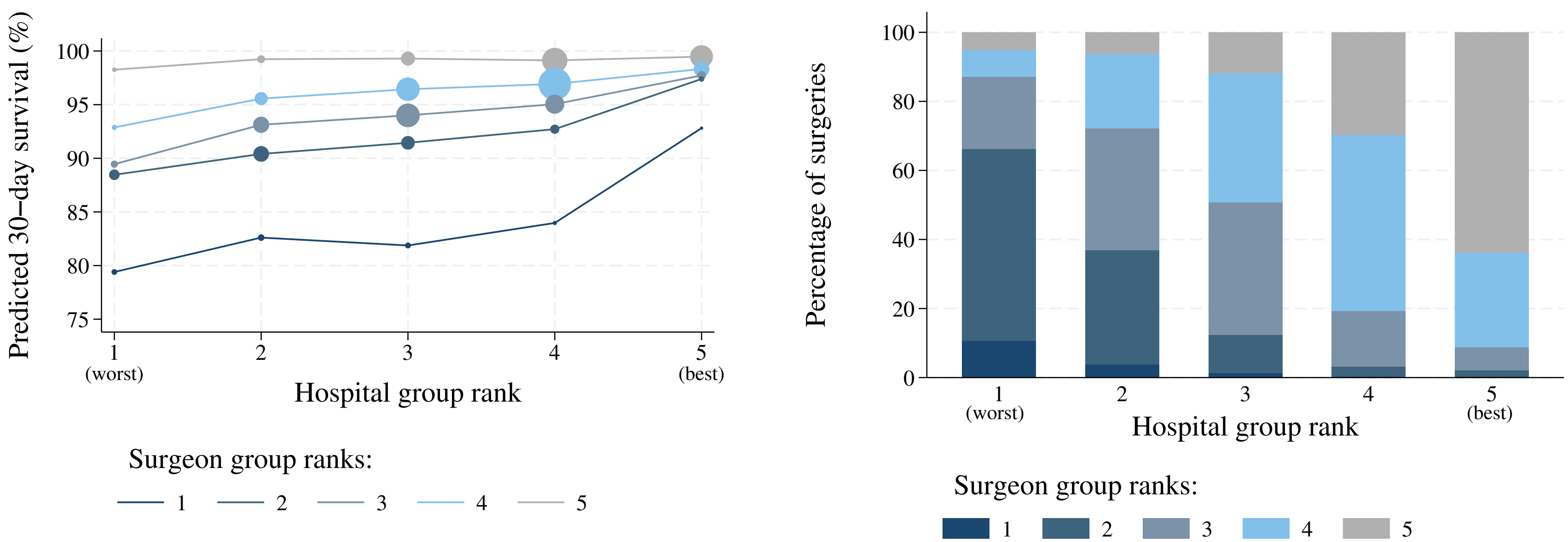
(a) Group effects on 30-mortality (Chandra, Dalton, and Staiger, 2023) (b) Group ranks on observables and external rankings



Notes: These graphs report the point estimates and 95% confidence intervals from regressions of the hospital-group estimated effects or ranks on hospital-level covariates. Panel 4a shows that higher-survival hospital groups tend to be associated with lower 30-day mortality for other diagnoses as measured by Chandra, Dalton, and Staiger (2023). Panel 4b shows that higher-rank hospital groups tend to be higher-volume hospitals when measured by discharges but not by the number of hospital beds. Higher-ranked hospitals using estimated effects also tend to be ranked in the *U.S. News* list of best hospitals for both overall care and cardiology, though the later is not statistically significant at 5%. Teaching hospitals tend to be higher-ranked hospitals using estimated effects, though this relationship is not statistically significant. The hospital-group estimates in panel 4a 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, to capture hospitals' average effect "purged" from the composition of surgeons that operate with them. Hospital rankings in panel 4b use these average effects to rank hospitals. 30-mortality measures from Chandra, Dalton, and Staiger (2023) are risk-adjusted and use empirical Bayes estimation to account for differences in hospital volume and quality drift over time. I use their 2014 measures, latest year available in their data. Hospital volumes are obtained using the CMS 2017 discharges for fee-for-service Medicare Part A by provider and service. Hospital teaching status and number of beds comes from the 2017 Dartmouth Atlas data, and the CMS five-star rating comes from the CMS Hospital General Information and Complications and Deaths datasets for 2017. Confidence intervals are constructed using clustered standard errors at the surgeon level.

Figure 5: Substitutability and sorting

(a) Substitutability (b) Sorting

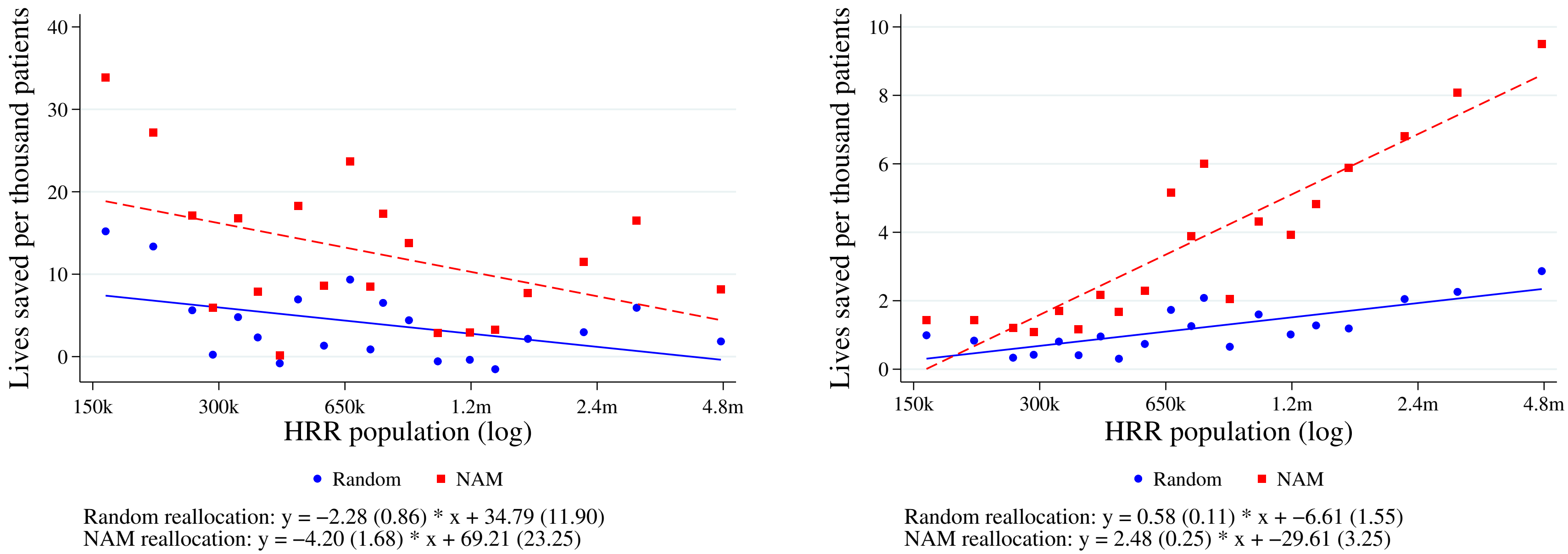


Notes: These graphs show results when assuming selection on observables as delineated in equation (6). Panel 5a displays the predicted 30-day survival for the average patient in the data across hospital and surgeon groups where groups are described by their relative rankings. The production function 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 3: 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 5b 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. Weighting each interaction equally allows to capture the average provider effect "purged" from sorting. 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 6: Regions benefit differently from reallocations as a function of population

(a) Reallocations *across* HRRs

(b) Reallocations *within* HRRs



Notes: These graphs illustrate the relationship between average mortality gains from partial equilibrium reallocation exercises where surgeons are reallocated to alternative types of hospitals and region size measured with total population. The graphs summarize these relationships using binned scatter plots with twenty equally sized bins. Two types of reallocations are reported: random reallocation and negative assortative matching. Patients and surgeons are reallocated either nationally in Panel 6a or within hospital HRRs only in Panel 6b. When surgeons are reallocated nationally, smaller HRRs gain more on average from reallocations than larger HRRs, for both random and negative assortative reallocations, with a steeper gradient with negative assortative matching. When surgeons are reallocated within their regions, larger HRRs gain more from reallocations than smaller HRRs, and this relationship is stronger for negative assortative reallocations. Overall, whether surgeons are reallocated across versus within regions will benefit different regions differently; in the first case, inequality in mortality across regions would disappear or favor smaller regions while in the later case, inequality across regions would increase in favor of larger regions. In the random reallocation, patient-hospital pairs are randomly reallocated to surgeon groups conditional on the number of surgeries available per surgeon groups nationally. For the negative assortative matching reallocation, surgeons from the highest-survival group are allocated to the lowest-survival hospital group until no surgeries are available at this hospital group, and so on. For reallocations within HRRs, surgeon groups operating in an HRR are reallocated to alternative hospital groups within the same HRR. 30-day mortality is predicted using parameter estimates from equation (6). Results per region are averaged over 100 simulations, and bootstrap standard errors are in parentheses (computed using 200 replications). 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.

Table 1: 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)	33.39	13.93
ZIP code population (x1,000)	25.27	18.97
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	0.04	0.20
60-days mortality	0.05	0.22
Length of stay	10.32	7.50
N	110,672	

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

Table 2: Suggestive evidence of non-perfect separability between surgeon skill and hospital quality

	(1) 30-day survival AMI quartiles	(2) RASR quartiles w/ single-homers Est.: movers only	(3) CMS five-star rating
Hospital group rank	-0.036	0.055	-0.033
	(0.039)	(0.081)	(0.049)
Hospital group rank * low-survival surgeons	0.447	0.690	0.336
	(0.084)	(0.167)	(0.108)
Observations	129,185	35,129	104,519
R-squared	0.026	0.028	0.026
Patient covariates	X	X	X
Surgeon group FE	X	X	X

Notes: This table reports the results of the regression $y_{ijht} = \beta_1 Q_h + \beta_2 Q_h \times 1\{\text{below median RASR}_j\} + \gamma X_{it} + \lambda_{g(j)} + \epsilon_{ijht}$ where y_{ijht} is an indicator for 30-day survival from CABG surgery, Q_h are hospitals group rankings (higher rank means higher survival or quality rating), $1\{\text{below median RASR}_j\}$ is an indicator for surgeons below the median risk- and noise-adjusted 30-day survival rate for CABG as depicted in panel 1a of Figure 1, X_{it} are patient observables listed in Table 1, and $\lambda_{g(j)}$ are fixed-effects for surgeons above versus below the median risk- and noise-adjusted 30-day survival rate for CABG. The estimate for β_2 is statistically different from zero, suggesting non-perfect separability between surgeon and hospital group value-added. Furthermore, it is negative, which indicates that surgeon and hospital value-added are substitutes in the production of 30-day CABG survival. In all columns, hospital groups are formed "out of sample", i.e. using data that does not include the outcome of patients included in the regression. The first column groups hospitals in quartiles using the outcome of a different diagnosis computed by CMS in 2017: 30-day AMI risk-adjusted survival. The second column groups hospitals in quartiles of computed risk- and noise-adjusted 30-day CABG survival as in panel 1a of Figure 1, but using the outcomes of patients from surgeons practicing at a single hospital in my sample. The regression is then run for movers only, i.e., surgeons who practice at more than one hospital in my sample. The third column uses the CMS five-star rating to group hospitals in five distinct groups, five being the highest grade. The AMI survival measure and the CMS five-star rating are obtained from the CMS Hospital General Information and Complications and Deaths datasets for 2017. 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 clustered at the surgeon level.

Table 3: Greater returns from higher-hospital group for lower-survival surgeon groups (substitutability)

			A. Split-sample Groups: single-homers v. movers only		B. Other outcomes	
	Baseline	Control function	Ranks & est. 2011-2017	Ranks 2011-2015 on 2016-2019	Length of stay (days)	Hospital spending (x1,000)
Slope surgeon						
Rank 1 (worst)	1.85	1.38	1.64	0.59	-1.24	-10.31
	(1.11)	(1.28)	(1.34)	(0.68)	(0.43)	(5.28)
Rank 2	1.58	1.45	0.21	0.33	-0.55	-7.13
	(0.40)	(0.65)	(0.41)	(0.36)	(0.11)	(1.03)
Rank 3	1.37	1.30	0.33	0.03	-0.56	-4.96
	(0.30)	(0.41)	(0.21)	(0.27)	(0.08)	(0.97)
Rank 4	0.84	0.71	0.13	0.01	-0.50	-4.41
	(0.20)	(0.24)	(0.16)	(0.29)	(0.09)	(0.93)
Rank 5 (best)	0.15	0.01	-0.02	0.04	-0.39	-4.15
	(0.06)	(0.14)	(0.06)	(0.28)	(0.11)	(1.44)
Test p-values						
Equality of slopes	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Slope rank 5 $\geq$ 1	<0.01	<0.01	<0.01	0.30	0.99	0.97
Slope rank 4 $\geq$ 2	<0.01	<0.01	0.19	0.30	0.99	0.99
Correlation surg.-hosp.	0.46	0.45	0.15	0.20	0.68	0.73
	(0.04)	(0.05)	(0.05)	(0.01)	(0.05)	(0.09)
Observations	130,075	130,075	35,129	12,341	130,075	130,075

Notes: This table reports the estimated slope coefficient per group of surgeons for alternative specifications. Larger slopes for low-survival surgeon groups indicate substitutability between surgeon and hospital group value-added. The slopes $\hat{\beta}^L$ are obtained from the regression $\hat{y}_{ijht} = \sum_{L=1}^{5} \mathbb{1}\{j \in L\}\, \beta^L \text{rank}_{k(h)} + \lambda_L + \epsilon_{ijht}$ where $\hat{y}_{ijht}$ is the predicted 30-day risk-adjusted survival from models delineated in equation (6) expressed in percentage points of survival, 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 $k(h)$, and λ_L are surgeon group fixed effects. Provider groups are ranked assuming interaction terms are equally likely, to capture providers' average effect "purged" from sorting. The control function column reports results from estimating equation (11). In panel A, hospital groups are formed using single-homers while surgeons are movers only. In the third column, the group ranks and the slopes are estimated using data from 2011-2017 for movers-only. In the fourth column, the group ranks are estimated using data from 2011-2015 and the slopes are estimated using data from 2016-2019. For the specifications using other outcomes, both the groupings and the outcome of the second-step main regression are based on the alternative outcome. Standard errors in parenthesis are bootstrap standard errors using 1,000 replications.

Table 4: 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	0.19	-0.75	-0.26
Change in deaths per 1,000 (reallocated - baseline)				
Aggregate	-2.97	-1.62	-10.52	-5.39
	(0.10)	(0.09)	(0.11)	(0.10)
% change from current allocation	-7	-4	-26	-14
% of national change	-	55	-	51
Standard deviation	-3.14	-1.97	-12.12	-6.51
	(0.13)	(0.12)	(0.12)	(0.12)
% change from current allocation	-8	-5	-31	-17
% of national change	-	63	-	54

Notes: This table reports results of a partial equilibrium reallocation exercise where surgeons are reallocated to alternative hospital groups. Two types of reallocations are reported: random reallocation and negative assortative matching. In the random reallocation, patient-hospital pairs are randomly reallocated to surgeon groups conditional on the number of surgeries available per surgeon groups nationally. For the negative assortative matching reallocation, surgeons from the highest survival group are allocated to the lowest survival hospital group until no surgeries are available at this hospital group, 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.

Table 5: Region-level predicted mortality and population across alternative allocations of surgeons

	Predicted mortality (lives per thousand)				
		Random		NAM	
	Baseline	Across HRRs	Within HRRs	Across HRRs	Within HRRs
	(1)	(2)	(3)	(4)	(5)
Population (log)	-2.05	0.23	-2.63	2.15	-4.53
	(1.36)	(0.58)	(1.37)	(0.50)	(1.36)
Observations	305	305	305	305	305
R-squared	0.01	0.00	0.01	0.05	0.04

Notes: This table illustrates the relationship between predicted CABG mortality, expressed in deaths per thousand patients, and region size across alternative allocations of surgeon types to hospital types. With the baseline allocation of surgeons to hospitals, larger HRRs in terms of total population tend to exhibit lower mortality rates. This relationship disappears or is reversed with reallocations of surgeons across HRRs, because high-survival surgeons are reallocated to lower-population HRRs. However, the negative baseline relationship becomes more negative with reallocations within regions. In the random reallocation, patient-hospital pairs are randomly reallocated to surgeon groups conditional on the number of surgeries available per surgeon groups nationally. For the negative assortative matching reallocation, surgeons from the highest-survival group are allocated to the lowest-survival hospital group until no surgeries are available at this hospital group, 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 per region are averaged over 100 simulations, and bootstrap standard errors are in parentheses (computed using 200 replications). 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.

Appendix – For Online Publication

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

Pauline Mourot

September 2025

A Data and theory 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

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). 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 this 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 Appendix Table A.1, I am able to match the claims of more than 95% patients identified in the Carrier file.

I restrict the sample in four main dimensions as reported in Appendix Table A.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.

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.

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 1 - 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 Risk-adjusted survival at the hospital and surgeon level

I compute risk-adjusted survival at the hospital or surgeon level following the methodology used in the CABG report-card literature (Huckman and Pisano, 2006; Ghaferi, Birkmeyer, and Dimick, 2009; Kolstad, 2013). In particular, the predicted probability of survival for each patient is estimated using

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

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, and OSR is the national average survival rate. I use the same methodology for surgeons.

A.5 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 1.

A.6 *K*-means algorithm

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

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

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

I follow the same strategy to partition the J surgeons into a pre-specified number of ‘'“types” L such that

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

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

A.7 Monte-Carlo simulations

I keep the network as given and simulate survival using the results from the main model estimates defined in equation (6), also assuming five groups on both sides a sin the main specification. I next estimate the two-step grouped fixed effect estimator using the same model delineated in equation (6). Patient covariates are omitted for simplicity here. These exercises show how to leverage single-homers versus surgeon movers to break correlated classification error between surgeons and hospitals, to form a lower bound on the true positive assortative matching.

Table A.6 show the results from Monte Carlo exercises when varying the true sorting parameter. The estimated sorting is largelly over-estimated for relatively weaker positive assortative matching and for negative assortative matching. Correlated classification error may lead to such over-estimates of the true sorting parameter since the classification for surgeons is based on moments that includes the effect of hospitals, and vice versa.. This is illustrated by results that break this correlation in the classification step: when hospitals are grouped on average survival using surgeries performed by single-homers only, and I estimate the model using surgeon movers only, the estimated sorting now tends to be under-estimated due to increased noise. Increasing the number of observations allows ot get closer to the true sorting parameters for positive assortative matching. In the case of negative assortative matching, provider-level average survival does not identify provider types, so the estimated sorting parameter remains zero. The intuition for this is that, when the classification error is uncorrelated between surgeons and hospitals, negative assortative matching misclassifies providers such that one cannot detect sorting at all.

Overall, by breaking any correlation in classification error between surgeons and hospitals, the single-homers versus movers approach constitutes a key robustness check to ascertain the existence of positive assortative matching. If I were to find zero sorting with such a robustness check, I would not be able to rule out negative assortative matching. However, if positive assortative matching still emerges, this confirms the existence of positive assortative matching and provides a lower bound for its true value. The trade-off is that this robustness check increase noise drastically by excluding three fourth of the data, which is why I keep it as a key robustness check and not as main specification.

A.8 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

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

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

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

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

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

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

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

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

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

To derive the control functions, note that

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

Using Small and Rosen (1981), we have

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

so that

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

with $\hat{p}_{ih}$ the predicted probability for i to choose hospital h obtained from the demand model. Now, assuming the choice of hospital is $s \neq h$, we have

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

Use

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

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

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

Therefore, the control function can be written as:

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

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

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

A.9 Variance decompositions

Linear case. I estimate a model without interactions such that

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

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

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

Table A.8 and Appendix Table **??** shows results from such decompositions, 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.

Logit case. I alternatively estimate a logit production function where surgeon and hospital group quality enter additively such that

$$Pr[Y_{ijht} = 1 | X_{it}, \alpha_{l(j)}, \psi_{k(h)}] = \frac{\exp\big(\alpha_{l(j)} + \psi_{k(h)} + \sum_s \beta_s X_{it,s}\big)}{1 + \exp\Big(\alpha_{l(j)} + \psi_{k(h)} + \sum_p \beta_p X_{it,p}\Big)} \tag{A.3}$$

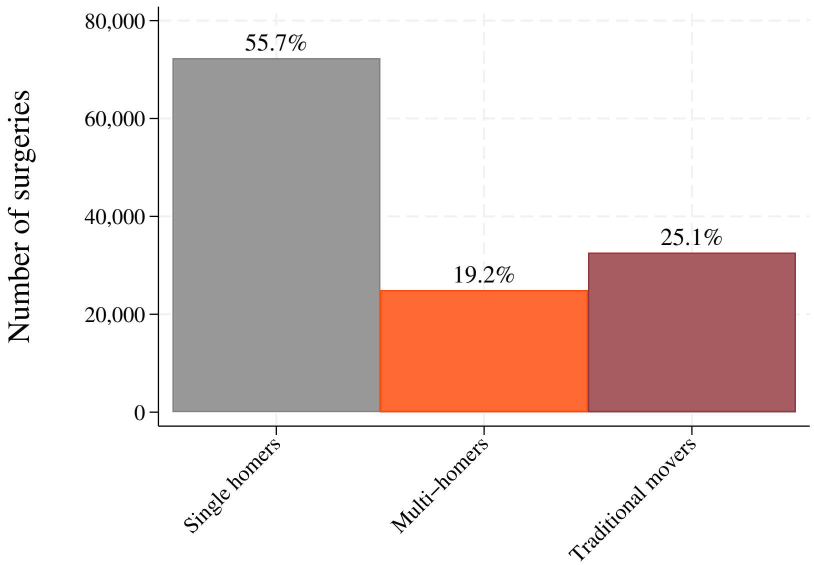
Since the predicted log odds of survival are linear in the hospital and surgeon group fixed effects and patient covariates, I decompose the variance as

$$Var\Big(\ln\Big(\frac{\hat{p}_{ijht}}{1-\hat{p}_{ijht}}\Big) - \sum_p \hat{\beta}_p X_{it,p}\Big) = Var(\hat{\alpha}_{l(j)}) + Var(\hat{\psi}_{k(h)}) + 2 \times cov(\hat{\psi}_{k(h)}, \hat{\alpha}_{l(j)}) \tag{A.4}$$

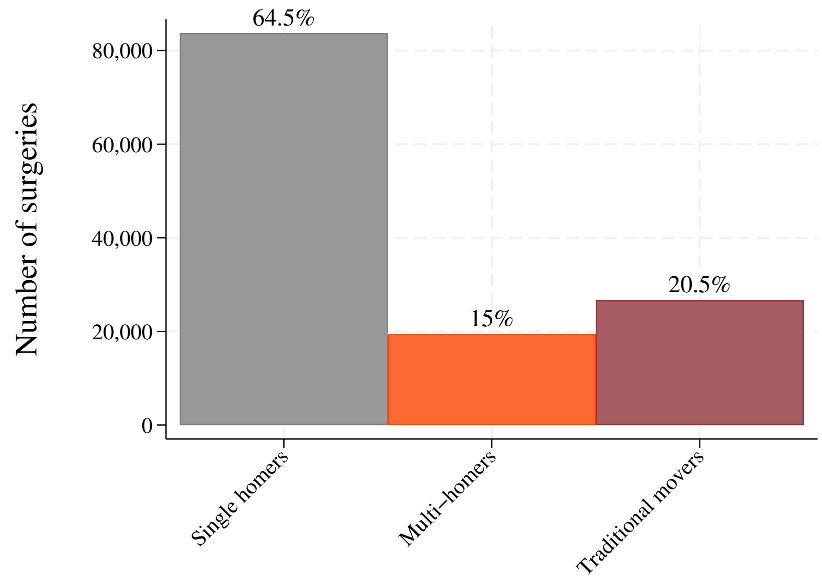
where $\hat{p}_{ijht}$ corresponds to the predicted 30-day survival from the estimated logit model. Panel B of Table A.8 reports results from this decomposition.

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

(a) ’A “move” is a different hospital



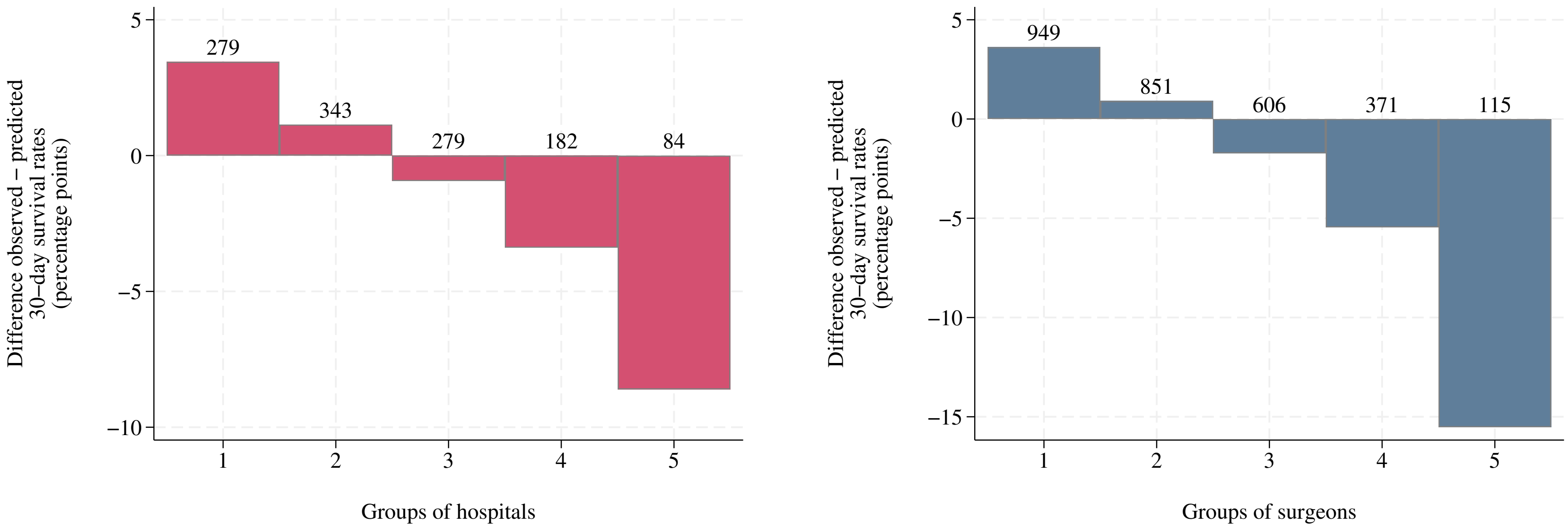
(b) A “move” is a different hospital group



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, while “traditional movers” are all the other surgeons observed at more than one hospital in the main sample time frame. These definitions are imperfect because I do not observe all surgeries performed by a surgeon in a given year. The definition for multi-homers is restrictive to get as close to a lower bound on the fraction of surgeries performed by multi-homers. Note that these labels do not impact the main analyses and results in the paper. Panel A.1a depicts the share of surgeries performed by each “type” of surgeon, where the type describes how many hospitals surgeons practice at in the main sample. Panel A.1b depicts the share of surgeries performed by each “type” of surgeon, but “moves” are now defined as moves across hospital *groups* and not individual hospitals. 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.

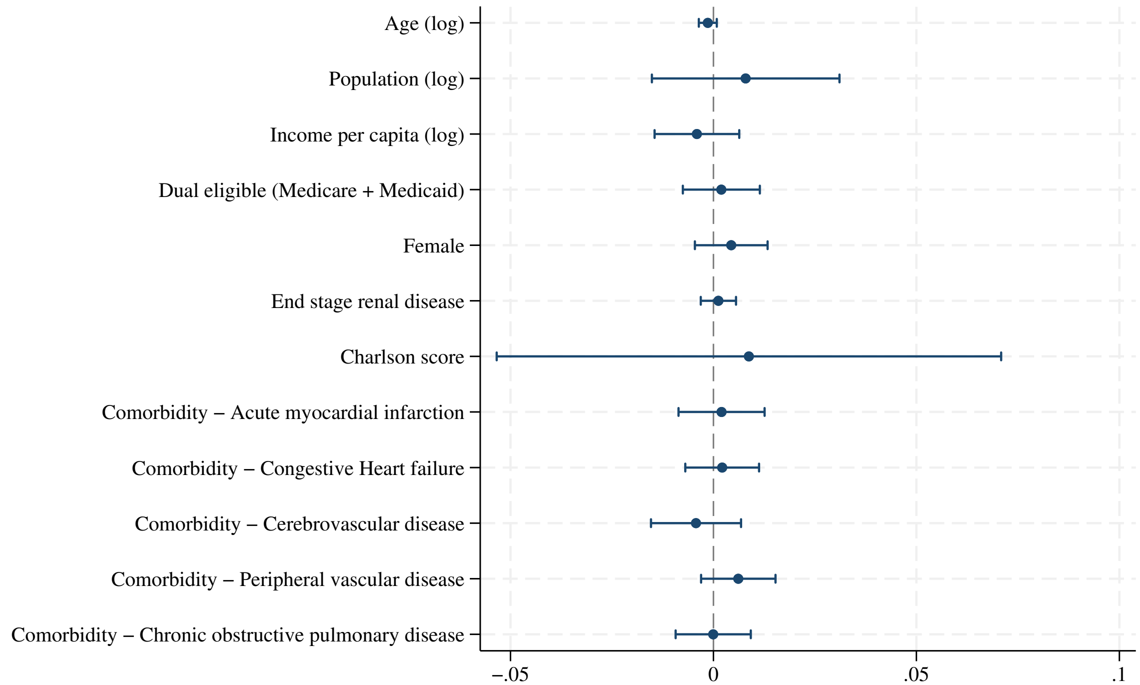
Figure A.2: Risk-adjusted survival rate variation across k-means groups

(a) Hospitals (b) Surgeons



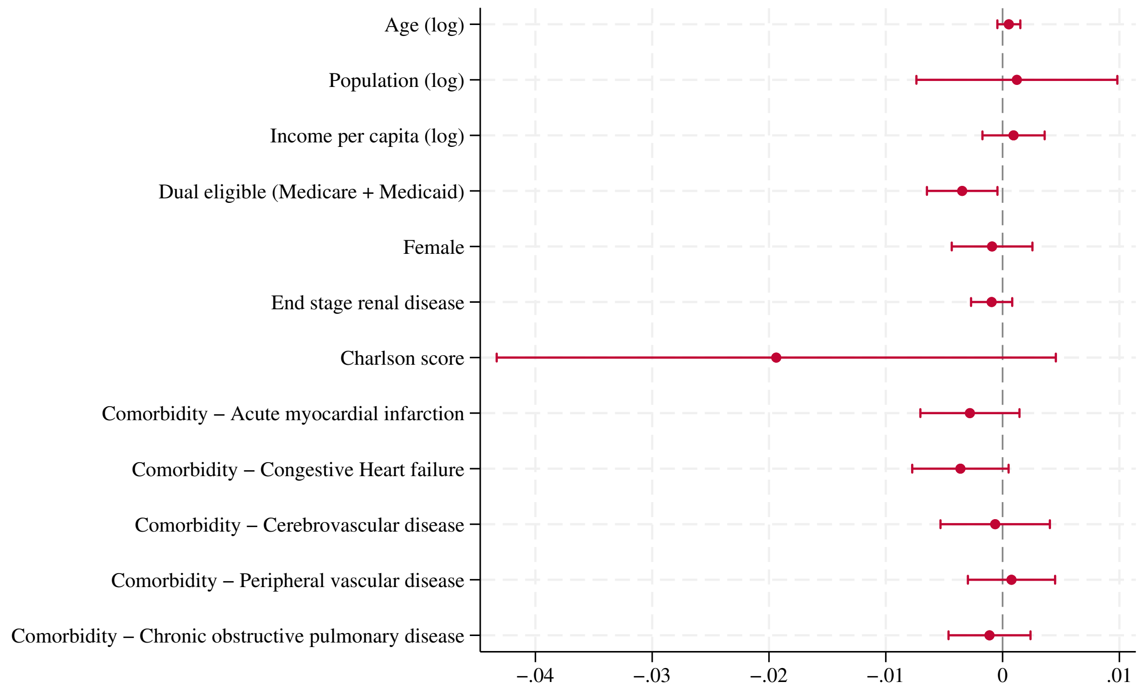
Notes: The variation in risk-adjusted 30-day survival across groups of hospitals and surgeons resulting from k-means clustering is large. Risk-adjusted survival is expressed as the difference between the average observed and average predicted 30-day survival. Numbers on top of bars indicate the number 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 A.3: 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 A.4: No evidence that hospitals systematically triage sicker patients into higher-survival surgeons

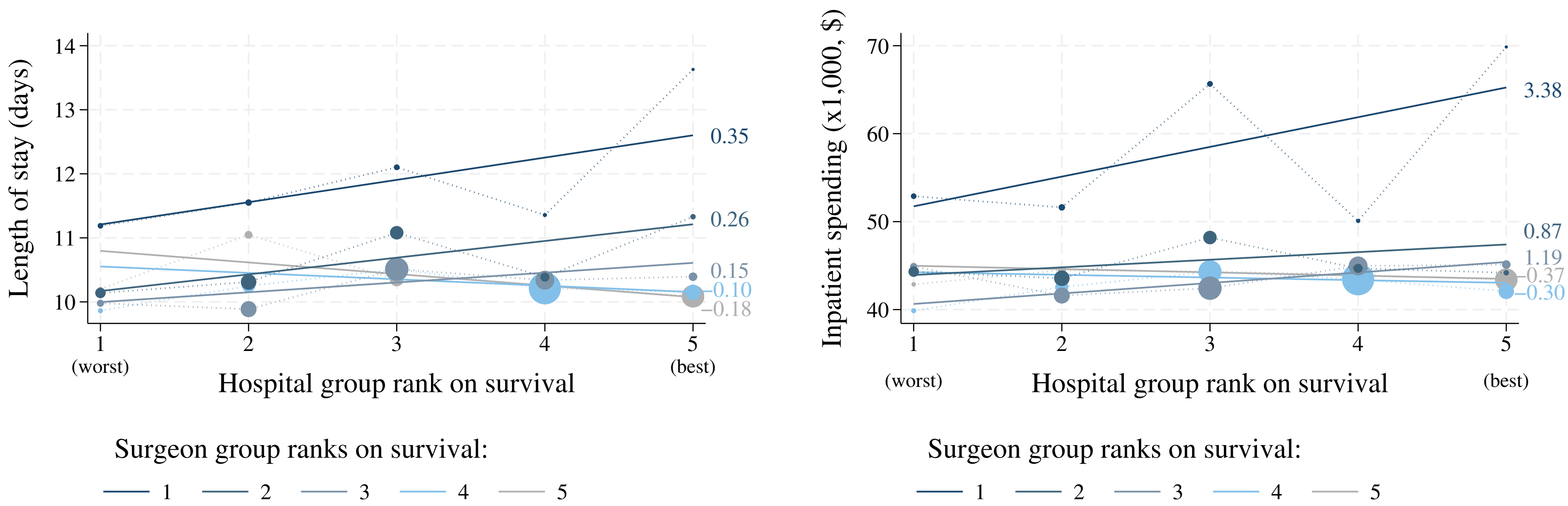


Notes: This graph examines the existence of "triaging" within hospitals, i.e., whether higher-survival surgeons tend to operate on sicker patients within a hospital using patient observables. All coefficients are close to zero and statistically insignificant, suggesting a limited role for triaging into 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 A.5: Length of stay and hospital spending across surgeon and hospital groups (groups formed using survival)

(a) Length of stay

(b) Hospital spending

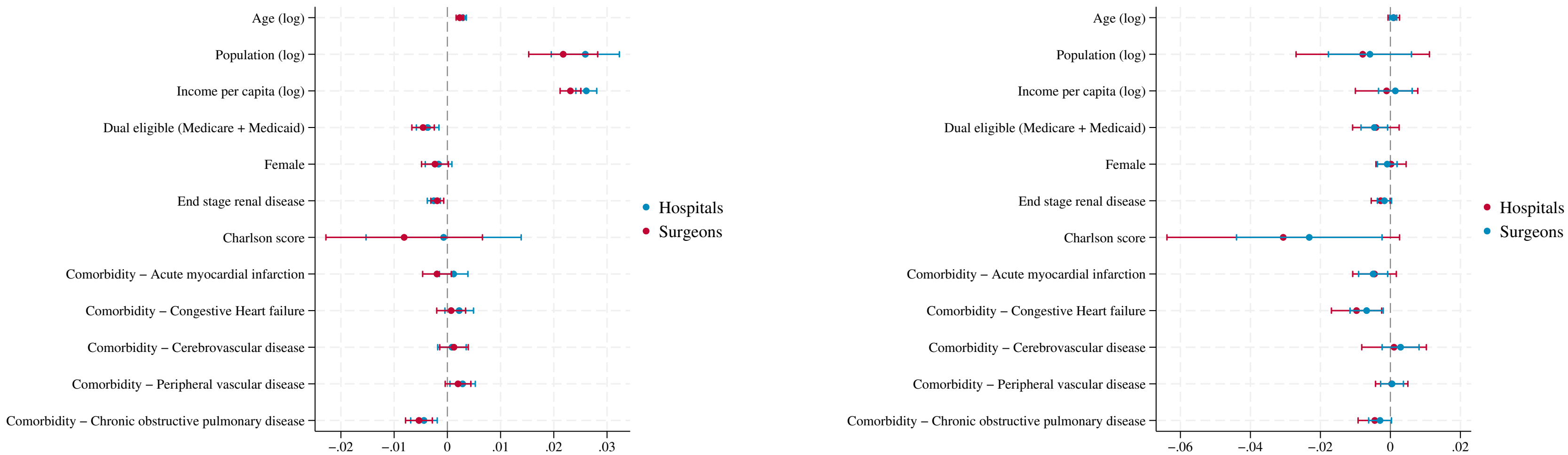


Notes: These graphs show how predicted length of stay and hospital spending for the average patient varies with hospital and surgeon group ranks, when surgeon and hospital groups are still determined using average risk-adjusted survival. These results are obtained from regressions as delineated in equation (6) but when the patient outcome is alternatively length of stay or hospital spending. Groups are formed using k-means clustering on average risk-adjusted survival as delineated in Section 3. The slopes of fitted lines across hospital rankings for each surgeon group are indicated to the right of each line. Lower-survival surgeons tend to exhibit longer length of stay and higher hospital spending than high-survival surgeons. However, examining the variation across hospital ranks, we find that lower-survival surgeons tend to achieve *worse* length of stay and hospital spending at better hospitals. This is consistent with the “failure-to-rescue” mechanisms. Provider groups are ranked assuming interaction terms are equally likely, to capture providers’ average effect “purged” from sorting. 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 A.6: Heterogeneity in patients treated across provider groups captured by provider location

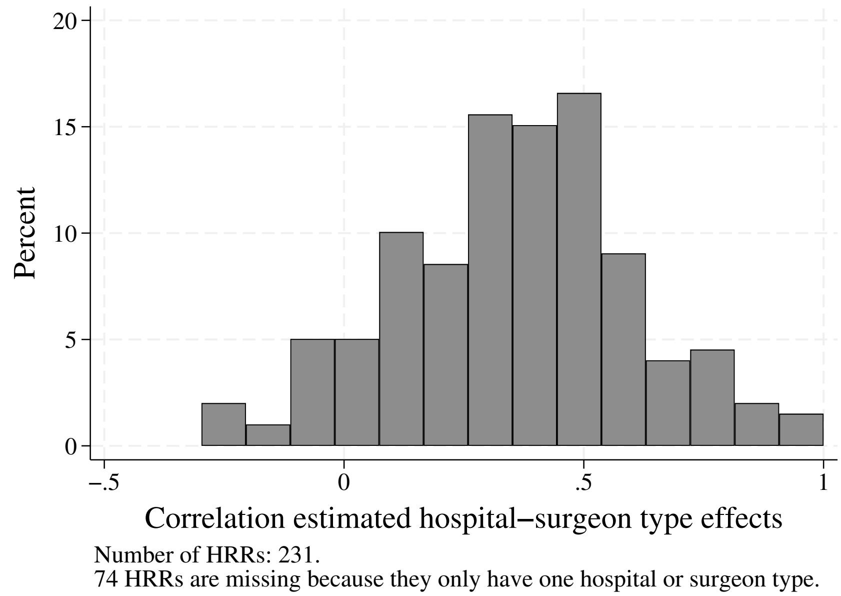
(a) No provider-location control

(b) With provider-HRR fixed effects



Notes: These graphs show 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. Indeed, the relationships with patient socioeconomic status such as ZIP code income, population and age become statistically insignificant after controlling for the provider locations. However, there exist marginally statistically significant evidence of advantageous selection into provider types using health measures, such as the Charlson score for surgeons or the fraction of patients with congestive heart failure for hospitals. Coefficients reported in panel A.6a correspond to the estimated $\hat{\beta}$ from the regression $x_{il(j)} = \alpha + \beta \text{rank}_{l(j)} + \epsilon_{il(j)}$, where $x_{il(j)}$ correspond to the covariates of patients treated by provider group $l(j)$. In panel A.6b, provider-location fixed effects $\lambda_z(j)$ are added using the ZIP code of their primary practice such that $x_{il(j)} = \alpha + \beta \text{rank}_{l(j)} + \lambda_z(j) + \epsilon_{il(j)}$ Coefficients reported in this graph correspond to the estimated $\hat{\beta}$ from the regression . 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.

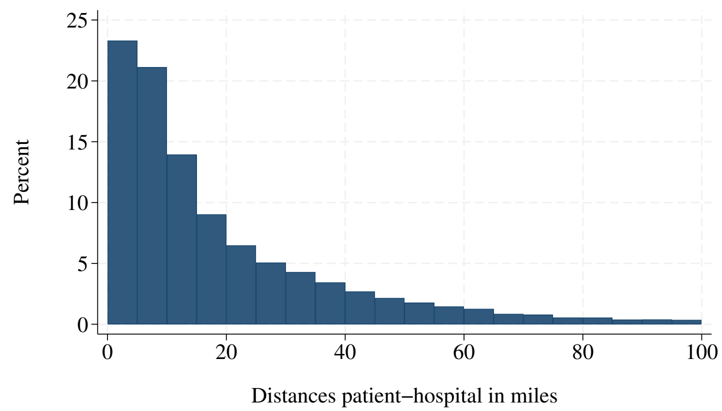
Figure A.7: Sorting within hospital referral regions (HRRs)



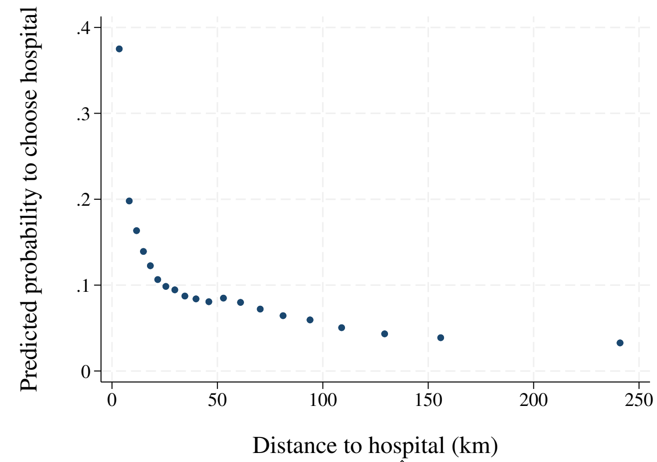
Notes: This graph shows the distribution of the estimated correlation between surgeon and hospital group effects computed for each HRR separately. There exist substantial positive assortative matching *within* HRRs for a substantial fraction of HRRs. Predictions come from estimating the model described by equation (6). Provider group effects are computed assuming interaction terms are equally likely, to capture providers' average effect "purged" from sorting. 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 A.8: Distance to the hospital is a strong predictor of hospital choice within an HRR

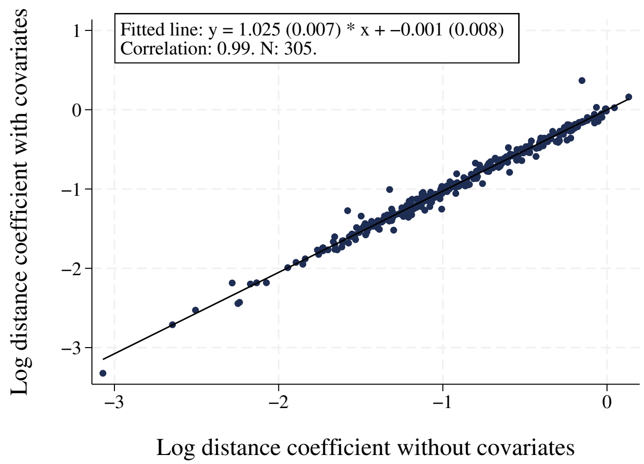
(a) Distances to chosen hospital



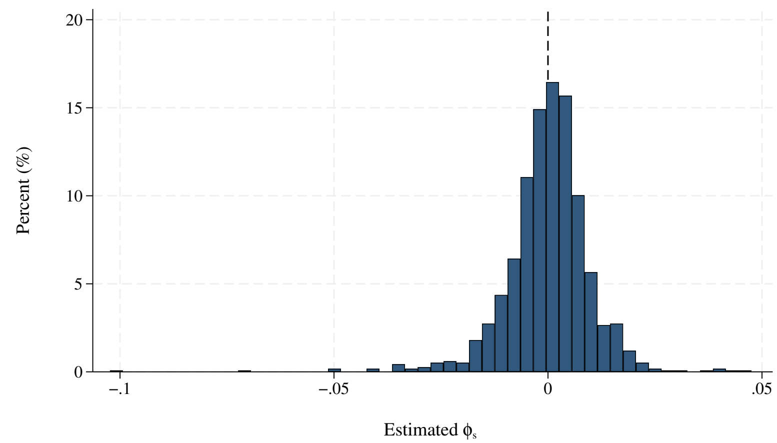
(b) Probability to choose a hospital and distance



(c) Distance parameters per HRR



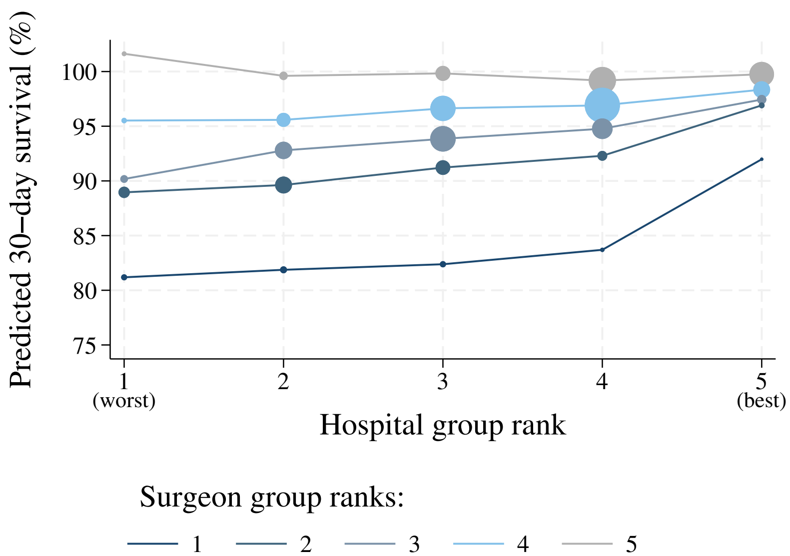
(d) Distribution of $\hat{\phi}_s$



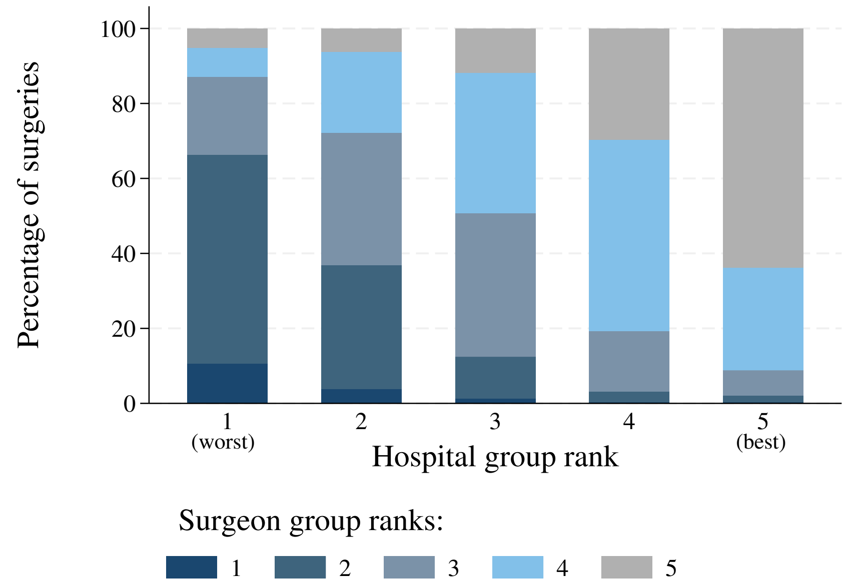
Notes: Panel A.8a depicts the distribution of distances between the patient's residential ZIP code and their chosen hospital's ZIP code, for patients treated at hospitals within their residential hospital referral region (HRR). The relationship looks log-linear. Panel A.8b depicts the relationship between 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 with a binned scatter plot with twenty equally sized bins. Only predicted probabilities for hospitals within a patient's residential HRR are included. Panel A.8c 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. Panel A.8d illustrates the distribution of the estimated control function parameters $\hat{\phi}_s$ from equation (11), with s denoting a specific hospital. Results suggest that some hospitals face adverse selection while other hospitals face advantageous selection. 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. Years 2011 to 2017 are included.

Figure A.9: Substitutability and sorting with a control function approach

(a) Substitutability



(b) Sorting



Notes: These graphs show results when using the control function approach delineated in equation (11). Panel A.9a 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 **??**: 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 A.9b 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. Provider groups are ranked assuming interaction terms are equally likely, to capture providers' average effect "purged" from sorting. 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 A.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	135,818	114,320	3,508	1,321
	(88%)	(93%)	(92%)	(100%)
Exclude surgeries performed end of 2010	135,701	114,220	3,508	1,321
	(88%)	(93%)	(92%)	(100%)
Exclude less than five surgeries per surgeon at a hospital	130,443	110,981	2,904	1,174
	(84%)	(91%)	(76%)	(88%)
Exclude patients or providers located outside of mainland U.S.	130,075	110,672	2,892	1,167
or in an HRR where all patients chose a hospital outside the HRR	(84%)	(90%)	(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. 602 patients received CABG surgery more than once in the 2011-2017 final sample. 16.3% 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 A.2: 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	72.43	72.23	72.14	-0.21***	-0.29***	0.08
	(8.24)	(8.09)	(8.28)	(0.06)	(0.06)	(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)	33.73	32.51	33.31	-1.22***	-0.43***	-0.79***
	(14.04)	(13.65)	(13.86)	(0.10)	(0.09)	(0.12)
ZIP code population (x1,000)	24.38	27.77	25.31	3.38***	0.93***	2.45***
	(18.64)	(19.83)	(18.84)	(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)	(0.21)	(0.00)	(0.00)	(0.00)
Charlson score	3.38	3.43	3.46	0.05***	0.08***	-0.02
	(2.65)	(2.69)	(2.67)	(0.02)	(0.02)	(0.02)
30-days mortality	0.04	0.04	0.04	0.00***	0.00***	0.00
	(0.19)	(0.20)	(0.20)	(0.00)	(0.00)	(0.00)
60-days mortality	0.05	0.06	0.05	0.01***	0.00*	0.00
	(0.22)	(0.23)	(0.23)	(0.00)	(0.00)	(0.00)
Length of stay	10.36	10.20	10.34	-0.15***	-0.02	-0.14***
	(7.82)	(6.76)	(7.34)	(0.06)	(0.05)	(0.06)
Years since medical school graduation, as of 2010	23.81	24.61	21.17	0.80***	-2.63***	3.43***
	(9.10)	(9.19)	(8.32)	(0.07)	(0.06)	(0.07)
Number of patients	72,439	25,004	32,632			
Number of surgeons	1,797	366	729			

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 3 years later 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 A.3: 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	349	154	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 A.4: Summary of observations per provider: surgeons and hospitals

		Mean	Median	Standard deviation
Per surgeon				
	Number of unique patients	45	37	34
	Number of unique hospitals	1.5	1	.81
Per hospital				
	Number of unique patients	95	69	91
	Number of unique surgeons	3.8	3	2.7

Notes: This table summarizes the number of unique patients and unique hospitals or surgeons per type of provider in the final 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.

Table A.5: Suggestive evidence of positive assortative matching: alternative proxies for surgeon skill and hospital quality

	Hospital quality						
			30-day risk-adjusted mortality (Chandra et al., 2023)				
	CMS five-star rating	RASR w/ single-homers Corr.: movers only	AMI	CHF	Hip & Knee	Pneumonia	Stroke
Surgeon skill (risk- & noise-adjusted)							
30-day survival for CABG	0.108	0.197	-0.150	-0.084	-0.097	-0.101	-0.009
N	103,349	35,023	127,443	127,833	122,438	125,002	127,613

Notes: This table displays the correlations between surgeons' skill and alternative proxies for hospital quality. In all columns, hospital groups are formed "out of sample", i.e. using data that does not include the outcome of patients included to compute the correlations nor the surgeons' 30-day survival rate. Surgeons' 30-day CABG survival is computed using the risk- and noise-adjusted 30-day survival measures for CABG surgery displayed in panel 1a of Figure 1. The CMS five-star ratings come from the CMS Hospital General Information and Complications and Deaths datasets for 2017. The second column proxies hospital quality using the risk-adjusted survival rate computed as detailed in Appendix A.3 *using single-homers only* while the correlations are estimated for surgeons performing surgery at more than one hospital in the main sample. The risk-adjusted mortality measures for other diagnoses come from Chandra, Dalton, and Staiger (2023). 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 A.6: Monte Carlo simulations - sorting estimates

True groups	Estimated groups	Classification	Number of observations	Outcome noise	Mean survival	Sorting: True	Sorting: AKM	Sorting: GFE	Correlation true vs estimated FE: Surgeons	Correlation true vs estimated FE: Hospitals
5	5	-	130,075	-	0.95	0.46	-0.90	0.48	0.63	0.59
							(0.03)	(0.04)	(0.01)	(0.05)
5	5	-	130,075	-	0.94	0.28	-0.89	0.47	0.63	0.61
							(0.02)	(0.03)	(0.01)	(0.03)
5	5	-	130,075	-	0.91	-0.47	-0.90	0.35	0.52	0.22
							(0.02)	(0.03)	(0.01)	(0.03)
5	5	Single-homers vs. movers	35,129	-	0.95	0.36	-0.90	0.10	0.63	0.24
							(0.02)	(0.07)	(0.02)	(0.16)
5	5	Single-homers vs. movers	35,129	-	0.94	0.19	-0.89	0.05	0.61	0.26
							(0.02)	(0.04)	(0.02)	(0.16)
5	5	Single-homers vs. movers	35,129	-	0.91	-0.35	-0.89	0.01	0.60	0.04
							(0.02)	(0.04)	(0.02)	(0.06)
5	5	Single-homers vs. movers	70,258	Lower noise	0.95	0.36	-0.86	0.18	0.73	0.41
							(0.03)	(0.05)	(0.02)	(0.11)
5	5	Single-homers vs. movers	70,258	Lower noise	0.94	0.19	-0.86	0.08	0.71	0.37
							(0.02)	(0.04)	(0.02)	(0.13)
5	5	Single-homers vs. movers	70,258	Lower noise	0.91	-0.35	-0.87	0.00	0.69	0.04
							(0.02)	(0.03)	(0.02)	(0.05)

Notes: This table shows Monte Carlo simulation results when generating survival based on the estimated model from equation (6) assuming five groups on each side and assuming away patient covariates. The estimated model uses the two-sided grouped fixed effect estimator using a linear probability model with interactions as in the main specification of the paper. When the true sorting is strong positive assortative matching, we recover the sorting, slightly overestimated. With weaker PAM and NAM, the sorting is largelly over-estimated due to correlated classification error, since surgeon and hospital groups are formed on moments that include the effect of each other. However, these results also show how we can recover positive assortative matching by classifying surgeons and hospitals on different observations. When the classification step is performed on single-homers surgeries for hospitals and movers surgeries for surgeons, hence uncorrelating classification error in the classification step, we recover an under-estimate of the true sorting parameter. Noise is now uncorrelated in the classification step, such that it biases the estimated sorting toward zero. Increasing the number of observations allows to get closer to the true parameter for positive assortative matching. Since provider types are not identified from average survival with strong negative matching, the estimated sorting will remain close to zero. Estimated parameters and standard deviations are calculated across 100 simulations.

Table A.7: Estimated coefficients on patient observables for risk adjustment

	Selection on observables	Control function
Age - [65;70)	0.0053***	0.0051***
	(0.0021)	(0.0021)
Age - [70;75)	0.0002	0.0001
	(0.0021)	(0.0021)
Age - [75;80)	-0.0085***	-0.0087***
	(0.0021)	(0.0022)
Age - [80;85)	-0.0215***	-0.0218***
	(0.0023)	(0.0023)
Age - [85;90)	-0.0193***	-0.0193***
	(0.0031)	(0.0031)
Age - [90;95)	-0.0675***	-0.0674***
	(0.0078)	(0.0078)
Age - [95;100)	0.0132	0.0159
	(0.0409)	(0.0413)
Female	-0.0163***	-0.0164***
	(0.0012)	(0.0012)
Dual eligible (Medicare + Medicaid)	0.0063***	0.0064***
	(0.0016)	(0.0016)
Income per Capita (log)	0.0047***	0.0041***
	(0.0016)	(0.0018)
Population (log)	-0.0002	-0.0000
	(0.0005)	(0.0005)
End stage renal disease	-0.0256***	-0.0256***
	(0.0027)	(0.0027)
Charlson score	-0.0012***	-0.0012***
	(0.0003)	(0.0003)
Comorbidity - Acute myocardial infarction	-0.0093***	-0.0095***
	(0.0012)	(0.0012)
Comorbidity - Congestive Heart failure	-0.0224***	-0.0225***
	(0.0012)	(0.0013)
Comorbidity - Peripheral vascular disease	-0.0117***	-0.0116***
	(0.0013)	(0.0014)
Comorbidity - Cerebrovascular disease	0.0000	0.0000
	(0.0012)	(0.0013)
Comorbidity - Chronic obstructive pulmonary disease	0.0004	0.0000
	(0.0013)	(0.0013)
Roy selection		-0.0000
		(0.0007)
N	130,075	130,075

Notes: This table reports the estimated coefficient for patient covariates from running regressions delineated in equations (6) and (11). Income per capita and population are computed from the patient's ZIP code of residence and come from the American Community Survey (ACS) 2015-2019 from the U.S. Census Bureau. Standard errors are clustered at the year, surgeon and hospital group levels. 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 A.8: 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	65.33	9.56	25.11	0.50
Control function	72.44	6.09	21.46	0.51
Split-sample (groups: single-homers, est.: movers only, 2011-2017)				
30-day survival	93.16	2.45	4.38	0.15
Length of stay	82.45	3.76	13.79	0.39
Hospital spending	49.42	15.39	35.19	0.64
Other outcomes				
Length of stay	57.30	9.75	32.95	0.70
Hospital spending	35.48	21.18	43.34	0.79
Alternative groupings				
K-means on survival, length of stay, and hospital spending	72.35	8.23	19.42	0.40
Quintiles risk-adjusted survival	66.60	8.15	25.25	0.54
K-means on noise-adjusted estimates (empirical Bayes)	64.45	8.40	27.15	0.58
Logit production function	65.23	9.93	24.84	0.49
No emergencies	67.81	8.55	23.65	0.49
Alternative number of groups				
K=5, L=10	69.82	7.12	23.06	0.52
K=10, L=5	61.56	11.66	26.78	0.50
K=10, L=10	66.20	8.94	24.86	0.51
K=20, L=20	68.35	8.00	23.65	0.51
K=50, L=50	68.46	8.18	23.37	0.49

Notes: This table reports the variance decomposition as delineated in equation (A.2) for alternative specifications. The split-sample specifications use single-homers to form hospital groups while surgeons are movers only, and both are estimated on 2011-2017. For other outcome results, the grouping is performed on the specified outcome. K-means on all outcomes includes standardized 30- and 60-day survival, length of stay, and hospital spending. Quintiles include the same number of surgeries. K-means on the noise-adjusted estimates groups using the risk-adjusted surgeon and hospital effects with empirical Bayes shrinkage computed as in Figure 1. The variance in predicted log-odds of 30-day survival is used for the logit model. The sample without emergencies excludes all hospital claims with non-zero emergency department amounts. Hospital spending corresponds to the facility payment made to the hospital. 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 A.9: Variance decomposition for 30-day survival

	Percentage of variance (%)	
	Selection on observables	Control function
Hospitals	0.25	0.17
$Var(\psi_{k(h)})$		
Surgeons	1.69	2.04
$Var(\alpha_{l(j)})$		
Sorting	0.65	0.60
$2 \times cov(\alpha_{l(j)}, \psi_{k(h)})$		
Patients covariates	1.33	1.85
$Var(\beta X_{it})$		
Year	0.14	0.14
$Var(\lambda_t)$		
Covariance FEs-patients covariates	-0.23	-0.48
$2 \times cov(\alpha_{l(j)} + \psi_{k(h)} + \lambda_t, \beta X_{it})$		
Covariance surgeon and hospital-year	-0.00	-0.01
$2 \times cov(\alpha_{l(j)} + \psi_{k(h)}, \lambda_t)$		
Residuals	96.18	95.68
$Var(\epsilon_{ijht})$		
N patients	110,672	110,672
N surgeons	2,892	2,892
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 (A.1). 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 4%, 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 A.10: Individual fixed effects corrections: plug-in (AKM) and leave-out (KSS)

	Fixed effects (AKM)	Corrected (KSS)
Risk-adjusted length of stay		
Covariance surgeon-hospital	-5.59	0.92
Percentage of explained variance (%)	-117.34	19.23
Percentage of total variance (%)	-8.53	1.40
Risk-adjusted hospital spending		
Covariance surgeon-hospital	-75,686,104	49,466,240
Percentage of explained variance (%)	-44.15	28.86
Percentage of total variance (%)	-6.37	4.17
N	33,845	33,845
Percentage of full set (%)	26.02	26.02

Notes: This table reports the estimated covariance between individual hospital and surgeon fixed effects in a model without interactions when using the plug-in estimator as in Abowd, Kramarz, and Margolis (1999) (AKM) and when using the leave-out correction from Kline, Saggio, and Sølvsten (2020), referred to as KSS here. They are estimated using the Kline, Saggio, and Sølvsten (2020) Matlab package `LeaveOutTwoWay`. While the AKM estimator points to strong negative matching, corrected estimates indicate strong positive assortative matching. The corrected results are consistent with the grouped fixed-effects results for both outcomes reported in Panel F of Table A.8. Note that the leave-one-out connected set represents about a quarter of the full sample: this connected set represents the largest set of hospitals connected by surgeons after any surgeon from the graph is removed. The reported leave-out corrections are robust to heteroskedasticity and serial correlation of the error term within a surgeon-hospital pair. The estimates are performed on risk-adjusted outcomes, computed as the ratio of observed over predicted at the patient level multiplied by the average outcome in the sample, without additional controls. Predictions for risk-adjustment are obtained using Poisson regressions and including all patient covariates and year fixed effects. 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 A.11: The relationship between patient outcomes and distance to the chosen hospital is similar when including patient observables

30-day survival	(1)	(2)	(3)
Log distance (km)	0.00119	0.00126	0.00148
	(0.00059)	(0.00060)	(0.00066)
Observations	102,589	102,589	102,589
R-squared	0.00587	0.01209	0.01771
Patient's HRR FE	Yes	Yes	Yes
Patients' observables		Health-income-age	All

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

Table A.12: Robustness of the substitutability result to alternative number of groups

	(1) Baseline	(2) $K = 10; L = 5$	(3) $K = 5; L = 10$	(4) $K = 10; L = 10$
Slope surgeon				
Rank 1 (worst)	1.85	1.28	1.48	1.27
	(1.11)	(0.60)	(2.34)	(1.27)
Rank 2	1.58	0.85	1.56	0.94
	(0.40)	(0.21)	(0.81)	(0.43)
Rank 3	1.37	0.62	1.41	0.84
	(0.30)	(0.16)	(0.52)	(0.27)
Rank 4	0.84	0.37	1.25	0.63
	(0.20)	(0.11)	(0.40)	(0.21)
Rank 5	0.15	0.10	0.87	0.43
	(0.06)	(0.04)	(0.39)	(0.21)
Rank 6			0.80	0.37
			(0.53)	(0.26)
Rank 7			1.11	0.57
			(0.49)	(0.23)
Rank 8			0.48	0.23
			(0.30)	(0.15)
Rank 9			0.29	0.15
			(0.21)	(0.10)
Rank 10 (best)			-0.03	-0.01
			(0.03)	(0.02)
Test p-values				
Equality of slopes	<0.01	<0.01	<0.01	<0.01
Slope rank 5 $\geq$ 1	<0.01	<0.01		
Slope rank 4 $\geq$ 2	<0.01	<0.01		
Slope rank 10 $\geq$ 1			<0.01	<0.01
Slope rank 8 $\geq$ 3			<0.01	<0.01
Correlation surg.-hosp.	0.46	0.41	0.47	0.45
	(0.04)	(0.05)	(0.05)	(0.05)
Observations	130,075	130,075	130,075	130,075

Notes: This table reports the estimated slope coefficient per surgeon group for alternative specifications. The slopes $\hat{\beta}^L$ are obtained from the regression $\hat{y}_{ijht} = \sum_{L=1}^{5} \mathbb{1}\{j \in L\}\, \beta^L \text{rank}_{k(h)} + \lambda_L + \epsilon_{ijht}$ 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 $k(h)$, and λ_L are surgeon group fixed effects. Provider groups are ranked assuming interaction terms are equally likely, to capture providers' average effect "purged" from sorting. 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 in parenthesis are bootstrap standard errors using 1,000 replications.

Table A.13: Additional robustness of the substitutability result

		C. Split-sample, other outcomes *Single-homers v. movers only*		D. Alternative groupings				
	Baseline	Length of stay of (days)	Hospital spending (x1,000)	All outcomes	Quintiles	Noise-adjusted	Logit	No ER
Slope surgeon								
Rank 1 (worst)	1.85	-0.41	-6.32	0.97	1.67	1.45	1.84	2.32
	(1.11)	(0.35)	(4.96)	(0.41)	(0.18)	(0.33)	(1.22)	(1.52)
Rank 2	1.58	-0.27	-4.78	0.62	0.98	1.21	1.45	2.12
	(0.40)	(0.10)	(1.14)	(0.27)	(0.13)	(0.20)	(0.38)	(0.55)
Rank 3	1.37	-0.28	-3.30	0.38	0.70	0.66	1.23	1.41
	(0.30)	(0.08)	(0.96)	(0.38)	(0.10)	(0.16)	(0.28)	(0.39)
Rank 4	0.84	-0.09	-2.59	0.20	0.28	0.37	0.75	0.75
	(0.20)	(0.07)	(0.81)	(0.19)	(0.08)	(0.12)	(0.17)	(0.24)
Rank 5 (best)	0.15	0.05	-3.06	0.13	-0.01	0.16	0.11	0.08
	(0.06)	(0.09)	(0.89)	(0.17)	(0.02)	(0.11)	(0.05)	(0.04)
Test p-values								
Equality of slopes	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Slope rank 5 $\geq$ 1	<0.01	1.00	1.00	<0.01	<0.01	<0.01	<0.01	<0.01
Slope rank 4 $\geq$ 2	<0.01	1.00	1.00	<0.01	<0.01	<0.01	<0.01	<0.01
Correlation surg.-hosp.	0.46	0.35	0.64	0.39	0.48	0.58	0.44	0.45
	(0.04)	(0.10)	(0.14)	(0.03)	(0.02)	(0.04)	(0.02)	(0.04)
Observations	130,075	35,129	35,129	130,075	130,075	130,075	130,075	100,329

Notes: This table reports the estimated slope coefficient per group of surgeons for alternative specifications, as in Table 3. Provider groups are ranked assuming interaction terms are equally likely, to capture providers' average effect "purged" from sorting. In panel C, hospital groups are formed using single-homers only while surgeons are movers only. The all-outcomes specification for other groupings includes standardized 30- and 60-day survival, hospital spending, and length of stay in the k-means algorithm to group both surgeons and hospitals. Quintiles include the same number of surgeries. K-means on the noise-adjusted estimates groups using the risk-adjusted surgeon and hospital effects with empirical Bayes shrinkage computed as in Figure 1. The logit column estimates equation (6) using a logit. The sample without emergencies excludes all hospital claims with non-zero emergency department amounts. 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 bootstrap standard errors using 1,000 replications.