MICHIGAN ROSS

Working Paper

Using Patient-Centric Quality Information to Unlock Hidden Health Care Capabilities

Guihua Wang
Stephen M. Ross School of Business
University of Michigan

Wallace J. Hopp
Stephen M. Ross School of Business
University of Michigan

Jun Li
Stephen M. Ross School of Business
University of Michigan

Steven F. Bolling
Cardiovascular Center
University Hospitals

Franco L. Fazzalari
Cardiovascular Center
University of Michigan Hospitals

Ross School of Business Working Paper
Working Paper No. 1291
August 2016

UNIVERSITY OF MICHIGAN

Using Patient-Centric Quality Information to Unlock Hidden Health Care Capabilities

Guihua Wang
Ross School of Business, University of Michigan, guihuaw@umich.edu

Jun Li
Ross School of Business, University of Michigan, junwli@umich.edu

Wallace J. Hopp
Ross School of Business, University of Michigan, whopp@umich.edu

Franco L. Fazzalari
Cardiovascular Center, University of Michigan Hospital, flf@med.umich.edu

Steven Bolling
Cardiovascular Center, University of Michigan Hospital, sbolling@med.umich.edu

We document a wide variation in quality among 188 surgeons at 35 hospitals in New York state that perform mitral valve surgery. Our analysis shows that patients of different demographics and levels of acuity benefit differently from elite surgeons. However, existing healthcare provider quality information is based on population averages and so does not differentiate patients of different medical conditions. This implies that patient-centric quality information, which calibrates outcome statistics by patient demographics and acuity, can increase the ability of patients to choose the most appropriate surgeon. In this paper, we develop an approach for computing patient-centric information from outcome data and evaluate the potential health benefits from using such information to guide patients to surgeons. We estimate that the total societal benefits from using patient-centric information are comparable to those achievable by enabling the best surgeons to treat 40% more patients under population-average information.

Key words: Healthcare outcome analysis, provider quality, patient-centric quality information

1. Motivation

How to "fix health care" is one of the most hotly debated topics in all of American society. Academic articles, media programs, legislative debates and water cooler conversations are generating scores of recommendations on how to provide patients with better and more cost effective health care. The vast majority of these proposals, from reimbursement bundling and accountable care organizations to patient care paths and lean transformations, are aimed at changing health care delivery and/or payment structure. However, a widely overlooked reality is that better and cheaper health care is available right now within the American system. Unfortunately, many patients just can't find it.

To provide a sense of this opportunity, Table 1 summarizes the risk-adjusted mortality rates for Medicare patients with six different medical conditions based on Healthgrades analysis of MedPAR data for years 2011 through 2013. The Healthgrades methodology uses a multivariate logistic regression to adjust for patient demographics and clinical risk factors that affect patient outcomes in significant and systematic ways (Healthgrades, 2015). It then ranks hospitals into three tiers with ascending quality: "1-star", "3-star" and "5-star". The comparison between 1-star and 5-star hospitals in Table 1 shows a substantial quality gap in mortality rate across providers.

Table 1 Average Mortality Rates between 5-Star and 1-Star Hospitals

	5-Star Hospital	1-Star Hospital	Relative Gap
Heart Attack	5.4%	10.3%	47.5%
COPD	0.6%	3.1%	81.8%
Pneumonia	2.6%	7.8%	67.2%
Stroke	4.5%	9.9%	54.6%
Colorectal Surgeries	2.6%	8.7%	70.7%
Sepsis	13.4%	23.1%	42.1%

Source: Healthgrades analysis of MedPAR data 2011-2013.

Healthgrades estimates that, if all hospitals as a group performed similarly to hospitals receiving 5 stars as a group, roughly 228,426 lives would have been saved between 2011 and 2013. But this simple analysis is only suggestive of the potential for better health care through hospital quality improvement. Mortality rates are not the only, or even the best, measure of the quality of health care. Also, it is not assured that the 1-star and 3-star hospitals can replicate the performances of the 5-star hospitals. Furthermore, even all hospitals receive 5 stars, if these hospitals have both high-performing and low-performing surgeons, the treatment outcomes will depend on case allocation between high-performing and low-performing surgeons in a hospital. Finally, estimates of outcome improvements are based on population averages and do not differentiate the potential gains by patients of different demographics and levels of acuity. To get a more accurate sense of what is possible and how it might be achieved, we need to delve into the details of a specific medical condition at the surgeon level.

In this paper, we do this by focusing specifically on patients with mitral valve disease and addressing three key questions: (1) How does quality vary among surgeons, (2) How are differences in quality affected by patient characteristics, and (3) How can outcome data be used to increase social value (i.e., sum of patients' utility)? Unfortunately, while simple to state, these questions are not straightforward to analyze using currently available data.

To answer the first question, we need to characterize the performance of the surgeons that treat mitral valve patients. There are various consumer-oriented healthcare rating systems that attempt to do this by providing provider quality information. For example, the Centers for Medicare and Medicaid Services maintain the Hospital Compare website that reports on over 4,000 Medicare-certified hospitals across the country with regard to quality of care, safety measures and patient satisfaction. Various non-profit organizations, including the Leapfrog Group, Consumer Reports and the California Health Care Foundation, and private companies such as US News and Healthgrades, also share self-reported hospital quality information and rankings via websites.

While these sources provide useful information, they fall well short of providing the data mitral valve patients need to make accurate comparisons of providers because: (1) they generally aggregate ratings into broad categories such as heart surgery, rather than reporting them for individual procedures such as mitral valve surgery; (2) most of these analyses are performed at the hospital level and do not provide information about individual surgeons within a hospital; and (3) most ratings do not indicate the magnitude of the difference between levels (e.g., between "1-star" and "3-star", or between "average" and "above average").

Some states, such as New York, address these issues by compiling risk-adjusted mortality rates for individual hospitals and surgeons that perform CABG and/or mitral valve surgery. They also indicate whether a hospital or a surgeon is statistically significantly better or worse than the state average.[1] However, there are still some remaining issues. First, they focus primarily on mortality rates (and sometimes on complication and readmission rates), where the low probability of events can make it difficult to discern statistical differences among providers. When all or most providers are not statistically different from the state average, patients don't have a basis for identifying better health care. However, if we focus on a specific medical condition, there may be other metrics that are both important indicators of quality and that show more variation across providers. In this study, we focus on mitral valve surgery and introduce mitral valve repair rate as an alternative quality measure in addition to the conventional measures (i.e., rates of mortality, complication and readmission). As we will show later in the paper, repair rate is an informative measure of surgeon quality and patient benefit, and a metric that shows significant variation among providers.

Second, all existing quality ratings are based on population-average outcome measures,[2] and therefore do not provide personalized guidance to patients of different demographics and levels of acuity. Population-average quality information has two major issues. First, patients care more about quality information specific to their procedure types and medical conditions than about information about an average patient who may not even exist. The lack of relevance makes patients less likely to act upon such information. Second, population-average information indicates that

[1] https://www.health.ny.gov/statistics/diseases/cardiovascular

[2] Population-average outcome measures refer to those focusing on the "average effect" or "homogeneous effect" of a given treatment (Kravitz et al., 2004).

all patients will benefit the same from an elite surgeon. However, it isn't reasonable to expect all patients to go to the single best surgeon, since this would create an impossible capacity imbalance. But, with patient-centric quality information, as we will show for the case of mitral valve surgery, it is possible to achieve substantial improvements in outcomes without overloading any single surgeon. Evaluating patient-centric information enables us to answer the second of our key questions.

To address the third key question of how outcome data can be used to improve social value, we examine two potential approaches for leveraging patient outcome data: (1) using population-average information and increasing surgeon capacity so that more patients can be treated by the best surgeons, and (2) providing patient-centric information so that patients can better balance benefits with costs of traveling to an elite surgeon for treatment. The second approach is motivated by the fact that existing quality information is almost always based on population averages and so does not indicate differences in patient benefits from elite surgeons. Consequently, patients who are guided to these surgeons may not be those who benefit the most. If patients, and the cardiologists who refer them, have access to patient-centric quality information and use it in selecting a surgeon, the patients with the most to gain will be those most inclined to incur additional costs associated with being treated by an elite surgeon. Whether the effect of offering patient-centric quality information is stronger or weaker than the effect of capacity increase is an open question we will address.

The remainder of the paper is organized as follows: Section 2 reviews the relevant literature. Section 3 summarizes the empirical setting and the data used in our study. Section 4 introduces the multilevel probit model that we use to evaluate surgeon quality. In Section 5, we analyze the estimation results from our quality model, document a wide quality gap among surgeons and discuss how the quality gap depends on patients demographics and levels of acuity. In Section 6, we evaluate policy interventions, including increasing surgeon capacity and providing patient-centric information. We estimate that providing patient-centric information offers societal benefits comparable to those achievable with a 40% increase in capacity. The paper concludes in Section 7 with a summary and a discussion of the challenges that must be addressed in order to take advantage of the vast opportunity to improve health care through better quality information and matching of patients with providers.

2. Literature Review

There has been a growing interest of studying hospital quality since 1989 when the Agency for Health Care Policy and Research was created by Congress in response to a report of wide geographic variations in practice patterns among hospitals in the US (see for example, Chassin et al., 1987). In a seminal paper, Keeler et al. (1992) compared 297 US hospitals for congestive heart failure, acute myocardial infarction, pneumonia, stroke or hip replacement, and found that quality varied from state to state, but that quality was generally better in teaching, large, and urban hospitals than in non-teaching, small, and rural hospitals. Subsequent studies have also found that high-volume hospitals tend to perform better than low-volume hospitals (Birkmeyer et al., 2002, Gammie et al., 2009, Vassileva et al., 2012) and that high-volume surgeons tend to perform better than low-volume surgeons (Birkmeyer et al., 2003, Bolling et al., 2010, Kilic et al., 2013).

In the operations management literature, a number of studies have also examined factors affecting health care quality. Some of these have focused on surgeon experience and its impact on surgical outcome. For example, KC and Staats (2012) investigated the differential effects of focal and related experience, and found that surgeon focal experience has a greater effect than related experience on surgeon performance. KC et al. (2013) examined how surgeons learn from their own and others' experiences, and found that individuals learn more from their own successes but also from others' failures. Ramdas et al. (2014) studied how learning and forgetting affect surgical outcomes by analyzing a surgeon's experience with specific surgical device versions and the time between their repeated uses. Other studies have analyzed the impact of workload on quality and patient outcome. For instance, Kim et al. (2015) examined the impact of ICU congestion on a patient's care pathway and the subsequent effect on patient outcomes, and found that the impact of ICU admission is highly variable for different patients and different outcomes. Jaeker and Tucker (2015) studied the relationship between workload and patient length of stay (LOS), and found that the effects of inpatient workload on LOS propagate across patient types. Freeman et al. (2015) show that gatekeeper providers (midwives in their study) ration resource-intensive discretionary services and also increase the rate of specialist referrals when workload increases. In addition to surgeon experience and workload, queue management (Song et al., 2015) and secure messaging between patients and physicians via patient portals (Bavafa et al., 2013) have been found to affect productivity and patient outcome as well. However, none of these studies have compared quality among health care providers or studied the impact of patient-centric information on outcomes.

Even though the aforementioned studies have vastly different focuses, the findings from this line of research suggest that: (1) both hospitals and surgeons play pivotal roles in determining healthcare quality; (2) experience at both institutional and individual levels significantly affects quality; and (3) besides experience, there are many other nuanced factors that may affect provider quality.

We follow this line of literature in our evaluation of healthcare provider quality by incorporating hospital and surgeon volume effects, as well as hospital and surgeon specific effects, in our model. Unlike the aforementioned studies, our focus is not identifying the effect of a specific factor on provider quality, but rather examining the quality gap among providers, which allows us to offer insights on how to best utilize the capabilities of existing healthcare system.

To accomplish this, we need to first identify elite surgeons who produce better quality outcomes. More importantly, we need to quantify the quality gap between an elite surgeon and an average surgeon for patients of different demographics and levels of acuity. Most prior studies have focused on measuring provider quality based on the population average (i.e., risk adjusted) outcomes, and assume away heterogeneity in outcomes among patients of different demographics and levels of acuity. As a result, their assessments of provider quality apply to average patients (who may not even exist) but may not be useful for a given patient. Recognizing this flaw in population average information, a number of observers have called for a patient-centered focus in both patient care and in quality assessment (see, e.g., FDA, 2013, Gerteis, 1993, IOM, 2011, Kattan and Vickers, 2004, Kent and Hayward, 2007, Kravitz et al., 2004). Our study contributes to the literature on patient-orientated care by proposing patient-centric quality information as a channel to better match patients with care.

3. Empirical Setting and Data

We choose mitral valve surgery as the empirical setting for our analysis of health care provider quality for several reasons. First, mitral valve disease is the most common form of heart valve disease in US. It affects 5% of the population and results in over 500,000 hospital admissions per year.[3] Second, mitral valve repair is a relatively new and complicated procedure. Because of the high level of skill required, surgeons may differ substantially in their outcomes. Third, there are many extant medical studies that provide data on the clinical outcomes of treatments available to mitral valve patients.

3.1. Mitral Valve Disease

The mitral valve is located between the left chambers of the heart. Its main function is to allow blood to flow from the left atrium to the left ventricle but not in the other direction. Mitral valve disease refers to conditions that compromise the ability of the mitral valve to seal against the backflow of blood.

There are two clinical options for the correction of mitral valve disease — mitral valve repair and mitral valve replacement. Mitral valve repair restores the function of the original valve, and is therefore the preferred option (Bolling et al., 2010). Table 2 compares the risks of mortality and complications associated with both procedures for a 60-year old male patient without major comorbidities (Society of Thoracic Surgeons, 2016). We see that the risks associated with replacement are 44.8% to 94.3% higher than those associated with repair.

Consequently, surgeons always strive to repair a mitral valve if possible. However, since it is impossible to guarantee a repair, whenever a surgeon operates on a mitral valve patient, he/she

[3] http://heartvalvedisease.nm.org/mitral-valve-disease.html

Table 2 Comparison of Mitral Valve Repair and Replacement

	Mitral Valve Repair	Mitral Valve Replacement	Relative Gap
Operative Mortality	0.4%	0.7%	94.3%
Prolonged Ventilation	2.7%	5.1%	85.8%
Renal Failure	0.9%	1.5%	63.5%
Reoperation	4.7%	6.8%	44.8%

Source: Society of Thoracic Surgeons risk evaluator, 2016.

has either a biological valve (from a cow or pig) or a mechanical valve (made of special carbon compounds and titanium) ready as a backup. If, during the procedure, visual inspection reveals the valve is not repairable, or a repair is attempted but fails (e.g., leaks), a replacement valve will be installed. The likelihood of a repair depends on patient characteristics and surgeon skill. Hence, repair rate (fraction of patients whose valves are repaired) is an indicator of surgeon quality.

3.2. Data Description

We used data from New York state that describe 10 million in- and out-patient discharges from all hospitals in New York from 2009-2012. These data contain patient-level clinical and resource-use information, including admission status (e.g., elective, emergent and urgent), patient demographics and comorbidities, hospital and physician identifiers, and principal and secondary diagnoses. For each discharge, the data record whether a patient received a mitral valve repair or replacement. They also indicate whether a patient died or experienced other complications during the procedure or post-surgery hospitalization. Because they record all visits, we are able to identify readmissions for the same patient across hospitals and time. Finally, the data include 5-digit zip codes of patients' home and hospital addresses, which allow us to estimate travel distance from each patient's home to any hospital.

3.3. Data Preparation

We identified discharges related to mitral valve surgery by using the clinical codes 35.12, 35.23 and 35.24 in the International Classification of Disease (9th revision). To focus on *isolated* mitral valve surgery, we followed previous studies (e.g., Vassileva et al., 2012) in excluding patients who were less than 30 years old, had coronary revascularization, congenital heart disease, excision of ventricular aneurysm, replacement of thoracic aorta, aortic fenestration procedure, closed heart valvuloplasty, heart transplant, or other valvular repair.

Because the ultimate objective of this study is to match patients with care, we focused on elective patients only, as opposed to emergent or urgent patients whose choice of providers may have been constrained by the urgency of their medical condition (Batt and Terwiesch, 2015). An elective mitral valve patient can wait for a year or more from diagnosis to treatment (Carroll et al., 1995), which provides considerable flexibility in the choice of providers.

Lastly, we focused on New York patients who were treated in New York hospitals. We do not directly observe New York residents who were treated outside New York because we lack data to compute patient-to-hospital distances for these patients.[4] This is unlikely to cause a sampling concern in our context, because New York has 4 out of the 50 nationally ranked heart programs. If a patient decides to seek a better provider than those available locally, the best providers in New York are comparable to the best providers in the country. We also excluded patients who traveled from other states to New York for mitral valve surgery, because we do not have sufficient data on out-of-state hospitals to describe these patients' local treatment options. We believe this exclusion also does not bias the estimation of provider quality, because it is very unlikely that provider quality varies for in-state vs. out-of-state patients conditional on patient demographics and medical conditions.

4. The Empirical Model

While it is widely accepted that health care providers differ with regard to quality, the size and nature of the differences are not well understood. Estimation of provider quality can be challenging for several reasons. First, each health care provider admits a different mix of patients. What may look like a provider effect could be due to the provider treating healthier or sicker patients. Second, even if we control for patient demographics and common comorbidities, there are usually conditions (e.g., test result details) that medical professionals and/or patients themselves can observe but researchers and policy makers cannot, which may affect provider/patient selection and clinical outcome (Dranove et al., 2003). Third, there may be insufficient data to discern statistically significant gaps in quality among health care providers. This is particularly a problem for quality metrics that measure rare events. Finally, the magnitude of provider quality gap may vary by patient characteristics. In this section, we evaluate provider quality for mitral valve surgery while addressing these issues.

4.1. Measures of Provider Quality

Mortality, complication and readmission rates are commonly used measures of health care provider quality. Rating systems such as US News, the Leapfrog Group and Healthgrades use mortality rates to compare and rank hospitals. The CMS-maintained website, Hospital Compare, compares Medicare-certified hospitals based on complication, readmission and mortality rates, as well as patient satisfaction. Most relevant to our study is the New York State Cardiac Surgery Reporting

[4] Although many others states in the US make their inpatient and outpatient discharge data available, most do not contain patient-level zip code information without which we cannot estimate distances to out-of-state providers.

System, which reports risk-adjusted mortality rates of hospitals and surgeons for major cardiovascular surgeries.

Each quality metric has its own shortcomings. The main obstacle to using mortality rate as a quality measure is the rarity of mortality for many procedures, which makes it difficult to discern statistically significant differences among providers. This problem is exacerbated when provider quality is evaluated based on relatively small patient samples. Dimick, et al. (2004) studied the small sample bias for the seven procedures for which mortality has been advocated as a quality indicator by the Agency for Healthcare Research and Quality. They found that *none* of these procedures *except* for CABG are performed frequently enough for mortality to serve as a gauge of hospital quality. Not surprisingly, rating systems such US News and the Leapfrog Group, which use mortality as a quality metric, do not indicate whether a given hospital is statistically significantly different from the national average. Systems that do indicate significance show most hospitals or surgeons as not significantly different from the average (at a 5% significance level), even when data are pooled over a three year interval (see for example the risk-adjusted mortality rates provided by New York State Cardiac Surgery Reporting System). Readmissions and complications are more common than mortality. However, using readmission rate as a quality metric also poses problems, including a lack of data, statistical modeling and the usability of the measures for quality improvement and accountability (Lorch et al., 2014, Morgan et al., 2013). In addition, a readmission may not indicate a low-quality outcome, for instance when follow up visits are normal or even a positive indication that the patient has survived long enough to be readmitted (Gilman et al., 2014). Complication rate may be more preferable to mortality or readmission rate. However, the clinical definition of a complication is imprecise because it does not always distinguish between hospital-acquired complications and present-on-admission complications (Bastani, Goh and Bayati, 2015).

Because of these known problems, we propose an additional measure of quality, repair rate (fraction of patients that receive a repair vs. a replacement) specifically for mitral valve procedures. Studies have shown that repairing instead of replacing the mitral valve offers lower operative risk (Gammie et al., 2009), lower risk of short-term and long-term complications such as strokes (LaPar et al., 2010), and better long-term survival (Daneshmand et al., 2009, Vassileva et al., 2013). For these reasons, the medical literature advocates mitral valve repair for *isolated* mitral pathologies and all patient age groups. However, because it involves a more complex surgical procedure than does replacement, the fraction of patients whose valves are successfully repaired (repair rate) depends on the skill of the surgeon.[5]

[5] http://my.clevelandclinic.org/services/heart/disorders/valvetreatment/mitral-valve-repair

In this study, the quality measures are operationalized as follows. Mortality is measured as death during hospitalization.[6] Complication is measured as occurrence of one or multiple mitral valve related complications including stroke, wound infection, renal failure, renal dialysis and ventilation observed during hospitalization (Society of Thoracic Surgeons, 2016). To analyze readmission rate, we merged outpatient discharge data with the inpatient discharge data using the link provided by the Agency for Healthcare Research and Quality. We focus on 30-day readmission by identifying patients who visited the same or other hospitals within 30 days after discharge.[7] Lastly, we observe from the data whether a patient received mitral valve repair or replacement based on the clinical procedure codes.

4.2. Factors Affecting Surgical Outcomes

Surgical outcomes of general patients are often affected by patient, hospital and surgeon characteristics. For example, old age correlates with increased risks of mortality, complication and readmission (Gupta et al., 2014, Merkow et al., 2015, Society of Thoracic Surgeons, 2016). Studies have found that white patients are less likely to have complications and unplanned readmissions than are black and Hispanic patients, and that female patients are more likely than male patients to have these undesired events (Iribarne et al., 2014, Merkow et al., 2015, Society of Thoracic Surgeons, 2016). Such differences can be the result of medical (e.g., comorbidities) or behavioral (e.g., delay in undergoing cardiac surgery) differences between various patient groups (Fasken et al., 2001). Comorbidities that increase the risk of mortality, complication and readmission include diabetes, chronic obstructive pulmonary disease, hypertension and renal failure (Gupta et al., 2014, Iribarne et al., 2014, Merkow et al., 2015, Society of Thoracic Surgeons, 2016).

With respect to repair rate in mitral valve surgery, Bolling et al. (2010) and Vassileva et al. (2013) separately found that younger and white patients are more likely to receive a repair, whereas females are less likely to receive a repair. Presence of various comorbidities including atrial fibrillation, chronic obstructive pulmonary disease, diabetes, heart failure, renal disease and hypertension also reduces the likelihood of mitral valve repair (Daneshmand et al., 2009, Savage et al., 2003, and Vassileva et al., 2013).

Of more direct interest to us in this study is the impact of hospitals and the surgeons within these hospitals on surgical outcomes. Presumably a hospital with more skilled surgeons, as well as more experienced support teams and an organizational structure that promotes learning and quality improvement, will have better quality than a hospital without these assets. However, because

[6] The New York in- and out-patient discharge data do not track post-discharge death or complication.

[7] https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program.html

quality and its antecedents are challenging to measure, patients and researchers alike must often rely on proxies to gauge the provider (hospital and/or surgeon) effect on quality. One of the most common proxies is surgical volume. Birkmeyer et al. (2002) found that mortality rate decreases as hospital volume increases by studying 2.5 million Medicare patients who underwent one of the 14 types of cardiovascular and cancer procedures from 1994 to 1999. Gammie et al. (2009), Vassileva et al. (2012) and Vassileva et al. (2013) all found that mitral valve repair rate increases as a function of a hospital's mitral volume. At the surgeon level, Birkmeyer et al. (2003) studied the relative importance of the experience of the operating surgeons by examining mortality among all 474,108 Medicare patients who underwent one of the eight cardiovascular procedures or cancer resections from 1998 to 1999 and found that the observed association between hospital volume and operative mortality is largely mediated by surgeon volume. In a separate study, Kilic et al. (2013) evaluated the combined effect of hospital and surgeon volume on operative outcome of mitral valve surgery by examining 50,152 patients from the National Inpatient Sample from 2003 to 2008 and found that the effect of hospital volume is largely driven by the individual surgeon volume within that hospital.

There are two mechanisms by which volume can promote quality: (1) practice helps surgeons and support teams maintain or improve their skills, and (2) scale justifies investment in advanced technology and equipment. Because of the documented relationship between surgical volume and quality, we include mitral volume of both surgeon and hospital in our quality model.

Note, however, volume alone is not sufficient to capture all variations in surgeon skills or hospital effects. Some surgeons may have better training or higher innate ability. At the hospital level, initiatives focused on quality improvement have also proven to be very effective (Barr et al., 2006, Lindenauer et al., 2007). Therefore, it is imperative to account for variations in surgeon and hospital quality beyond those captured by surgical volume.

4.3. Quality Model

To evaluate the impact of surgeon and hospital on patient outcomes, we need an econometric model that can address the following three issues. First, our data has a nested structure, i.e., patients are grouped under different surgeons and surgeons are grouped under different hospitals. Second, outcomes of patients treated by the same surgeon/hospital may correlate with each other due to unobservable surgeon/hospital characteristics. Third, we want to capture the previously discussed volume effects at both surgeon and hospital levels. Because of the potential for selection bias in patients of both surgeons and hospitals, we need a model that allows us to properly correct for such selection bias. Note that a simple probit or logit model cannot address these three issues. Instead, we used a multilevel probit model (Gibbons and Hedeker, 1997).

Let Y^*_{ijk} denote the latent variable associated with the outcome measure (i.e., mortality, complication, readmission, or repair) of patient i treated by surgeon j at hospital k. Y^*_{ijk} can be measured as a function of patient, surgeon and hospital characteristics:

$$\begin{aligned}
Y^*_{ijk} &= \gamma_1 Age_i + \gamma_2 Gender_i + \gamma_3 Race_i + \gamma_4 Comorb_i \\
&\quad + \gamma_5 SurgVol_i + \gamma_6 HospVol_i + \alpha_k + \beta_{jk} + \epsilon_{ijk} \\
Y_{ijk} &= \mathbf{1}\{Y^*_{ijk} > 0\}
\end{aligned}$$

where $SurgVol_i$ and $HospVol_i$ are measures of surgeon and hospital volumes of the procedure received by patient i,[8] α_k represents the unobserved effect of hospital k, and β_{jk} represents the unobserved effect of surgeon j at hospital k.[9] We assume these unobserved effects are drawn from two normal distributions:

$$\alpha_k \sim N(\mu_\alpha, \sigma^2_\alpha), \beta_{jk} \sim N(\mu_\beta, \sigma^2_\beta)$$

where μ_α and μ_β represent the mean hospital and surgeon effects, and σ^2_α and σ^2_β represent between-hospital and between-surgeon variations after accounting for hospital volume, surgeon volume and patient conditions at admission. If there are no between-hospital or between-surgeon differences in the outcomes beyond those captured by surgical volume and patient characteristics, then $\sigma^2_\alpha = 0$ (i.e., $\alpha_1 = \alpha_2 = ... = \alpha_K$) or $\sigma^2_\beta = 0$ (i.e., $\beta_1 = \beta_2 = ... = \beta_J$).

For all models, we have robust standard errors clustered by surgeon to allow for differences in the variance/standard errors due to arbitrary intra-group correlation (KC and Terwiesch, 2011, Jaeker and Tucker, 2015).

4.4. Estimation of the Volume Effect

Because surgical volume has a direct impact on the quality of care, medical publications, such as those from the Leapfrog Group, suggest that surgical volume is a good indicator of provider quality. However, surgical volume can also be correlated with unobserved patient characteristics that affect surgical outcomes. For example, health conscious patients could be more likely to be influenced by such medical guidances to choose high-volume surgeons and hospitals. They are also more likely to receive better outcomes due to their healthier lifestyles. If so, the resulting correlation will introduce negative biases when estimating the volume effect on mortality, complication and readmission rate, and a positive bias when estimating the volume effect on repair rate. Similarly, if a low-volume

[8] We follow the notations in KC and Terwiesch (2011) and use i to associate surgeon and hospital volumes with patients.

[9] There are twelve surgeons practicing at different hospitals. Because surgeon performance is institution-specific and not fully transferable across hospitals (Huckman and Pisano, 2006), we assume that these surgeons have independent unobserved effects. This assumption allows us to estimate provider quality using the multilevel probit model.

surgeon is more likely to select healthier patients, then, since a bad outcome is likely to have a larger impact on the outcome statistics when the sample size is smaller, the resulting correlation will introduce positive biases when estimating the volume effect on mortality, complication and readmission rate, and a negative bias when estimating the volume effect on repair rate.

To correct for potential biases like these, we use distance to construct instruments for surgical volumes. Previous studies have used distance from a patient's home to the hospital as an instrumental variable to study medical outcomes (McClellan, McNeil and Newhouse, 1994, McConnell et al., 2005, Brooks et al., 2006, Pracht et al., 2007, KC and Terwiesch, 2011). As in these studies, distance is an appropriate instrument for our purposes because it is correlated with the likelihood of choosing a provider. That is, the closer a provider is to a patient's home, the more likely it will be chosen. At the same time, where a patient is located is unlikely to be correlated with the remaining unobserved characteristics after controlling for demographics (age, gender and race) and common comorbidities, and we will provide empirical evidence supporting this requirement in the Appendix.

Similar to KC and Terwiesch (2011), we take the following steps to construct instruments for surgical volumes. We first model the choice of a provider, either a hospital or a surgeon, as a function of the patient's distance from the provider. That is, the probability of patient i going to provider j, denoted as p_{ij}, can be written as,

$$p_{ij} = \frac{exp(\delta Dist_{ij})}{\sum_{j=1}^{J} exp(\delta Dist_{ij})},$$

where the parameter δ can be estimated using a multinomial logit model.[10] We then calculate the expected volume for each patient by summing up the products of choice probability and associated provider volume over all providers:

$$\widehat{Vol}_i = \sum_{j=1}^{J} Vol_j \times p_{ij}$$

$\widehat{Vol}_i$ can now be used as an instrument for Vol_i. Note that due to the nonlinearity of the probit model, the standard two-stage least squares (2SLS) method will not yield consistent estimates as explained in Amemiya (1990). To address such situations, Wooldridge (2002) developed a maximum-likelihood estimation approach to estimate parameters of a probit model with an instrumental variable. We make use of a similar approach by first running the OLS regression of surgical volume

[10] In the multinomial logit model, the choice is defined as the chosen provider (either surgeon or hospital). The choice set includes all providers in the state of New York who perform mitral valve surgeries. Distance is defined as the Euclidean distance between the centroid of a patient's zip code and that of the provider, which has been shown to be highly correlated with actual travel distance and travel time (Boscoe, Henry and Zdeb, 2012).

on IV and all exogenous variables, and estimating the residuals. Then we estimate the multilevel probit model of the outcome variable on surgical volumes, the estimated residuals and all exogenous variables.[11] We should note that, since there is no variation in distances to surgeons within the same hospital, the identification for surgeon volume is driven by the variations in estimated choice probabilities among surgeons across different hospitals.

Lastly, it is worth noting that using distance to instrument surgical volume addresses the potential selection bias introduced by patients selecting providers based on volumes. However, there may be other selection biases which cannot be perfectly addressed by this approach, such as those introduced by surgeons selecting patients. For example, if low-performing surgeons are more likely to select healthier patients, the unobserved surgeon specific effects will also be correlated with unobserved patient characteristics, which affect surgical outcome directly. Instrumenting volume will not address such selection bias, and it is also infeasible to instrument hundreds of surgeon specific effects. Fortunately, we note that such selection will lead to an underestimation of quality gap among high- and low-performing surgeons, as the low-performing surgeons will have stronger incentives to select healthier patients. In other words, our approach will provide a conservative estimate of provider quality gap if such selection exists.[12]

4.5. Estimation of Hospital and Surgeon Effects

Because our primary goal is to guide patients to better care, our main interests are the hospital specific effect α_k and the surgeon specific effect β_{jk} on patient outcomes. To estimate these effects, we use an orthogonal transformation as in Gibbons and Bock (1987). That is, we rewrite the hospital and surgeon specific effect as $\alpha_k = \mu_\alpha + \theta_k \sigma_\alpha$ and $\beta_{jk} = \mu_\beta + \theta_{jk}\sigma_\beta$, where θ_k and θ_{jk} follow the standard normal distribution.

Note that, conditional on hospital and surgeon specific effects θ_k and θ_{jk}, the outcomes of all patients treated by surgeon j at hospital k are independent; therefore, the marginal probability of observing the set of outcomes at a hospital k can be expressed as:

$$h(Y_k) = \int_{\theta_k} \left(\prod_{j=1}^{J_k} \int_{\theta_{jk}} \left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk}, \theta_{jk}, \theta_k, \mu, \sigma, \gamma) \right) \phi(\theta_{jk}) d\theta_{jk} \right) \phi(\theta_k) d\theta_k. \tag{1}$$

[11] The scaled coefficients we obtain from this two-stage estimation are greater than their unscaled counterparts, but they are suitable for post-estimation analyses which will be discussed later. It is impossible to estimate the true coefficients of probit/logit models as they are scaled by the variance of error term. Post-estimation analyses of these models are normally based on scaled coefficients.

[12] Dranove et al. (2003) found evidence of low-performing surgeons selecting healthier patients after initial releases of the quality report cards in New York and Pennsylvania. However, in a more recent study, Kolstad (2013) found that surgeons with differing intrinsic incentives to improve quality do not appear to have accomplished reductions in mortality by changing their patient mix either based observable or unobservable (to the regulator) patient severity.

The individual likelihood function $l(Y_{ijk}|X_{ijk},\theta_{jk},\theta_k,\mu,\sigma,\gamma)$ equals:

$$l(Y_{ijk}|X_{ijk},\theta_{jk},\theta_k,\mu,\sigma,\gamma) = \left[\Phi(z_{ijk}(X_{ijk},\theta_k,\theta_{jk};\mu,\sigma,\gamma))\right]^{Y_{ijk}}\left[1-\Phi(z_{ijk}(X_{ijk},\theta_k,\theta_{jk};\mu,\sigma,\gamma))\right]^{1-Y_{ijk}},$$

where

$$\begin{aligned} z_{ijk}(X_{ijk},\theta_k,\theta_{jk};\mu,\sigma,\gamma) &= \gamma_1 Age_i + \gamma_2 Gender_i + \gamma_3 Race_i + \gamma_4 Comorb_i, \\ &\quad + \gamma_5 SurgVol_i + \gamma_6 HospVol_i + \mu_\alpha + \theta_k\sigma_\alpha + \mu_\beta + \theta_{jk}\sigma_\beta. \end{aligned}$$

We can now estimate (γ,μ,σ) by maximizing the log likelihood of observing the outcomes at all hospitals, which is expressed as:

$$\log L = \sum_{k=1}^{K} \log h(Y_k).$$

Upon obtaining estimates $(\hat\gamma,\hat\mu,\hat\sigma)$, we calculate $\hat\theta_k$ and $\hat\theta_{jk}$ for each hospital and each surgeon using the expected a posteriori (EAP) value (Bayes estimate) of θ_j and θ_{jk} (Bock & Aitkin, 1981, Gibbon & Hedeker, 1997).

$$\begin{aligned} \hat\theta_k &= \frac{\int_{\theta_k}\theta_k\left(\prod_{j=1}^{J_k}\int_{\theta_{jk}}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}\right)\phi(\theta_k)d\theta_k}{\int_{\theta_k}\left(\prod_{j=1}^{J_k}\int_{\theta_{jk}}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}\right)\phi(\theta_k)d\theta_k}, \\ \hat\theta_{jk} &= \frac{\prod_{j=1}^{J_k}\int_{\theta_{jk}}\theta_{jk}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\hat\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}}{\prod_{j=1}^{J_k}\int_{\theta_{jk}}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\hat\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}}. \end{aligned}$$

These quantities can be evaluated using Gauss-Hermite quadrature as described in Gibbons and Bock (1987) or Bock and Aitkin (1981). Estimates of α_k and β_{jk} can be recovered by $\hat\alpha_k = \hat\mu_{\alpha_k} + \hat\theta_k\hat\sigma_{\alpha_k}$ and $\beta_{jk} = \hat\mu_{\beta_{jk}} + \hat\theta_{jk}\hat\sigma_{\beta_{jk}}$. Finally, the standard errors can be estimated using

$$\begin{aligned} \sigma(\hat\theta_k) &= \frac{\int_{\theta_k}(\theta_k-\hat\theta_k)^2\left(\prod_{j=1}^{J_k}\int_{\theta_{jk}}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}\right)\phi(\theta_k)d\theta_k}{\int_{\theta_k}\left(\prod_{j=1}^{J_k}\int_{\theta_{jk}}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}\right)\phi(\theta_k)d\theta_k}, \\ \sigma(\hat\theta_{jk}) &= \frac{\prod_{j=1}^{J_k}\int_{\theta_{jk}}(\theta_{jk}-\hat\theta_{jk})^2\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\hat\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}}{\prod_{j=1}^{J_k}\int_{\theta_{jk}}\left(\prod_{i=1}^{N_{jk}} l(Y_{ijk}|X_{ijk},\theta_{jk},\hat\theta_k,\hat\mu,\hat\sigma,\hat\gamma)\right)\phi(\theta_{jk})d\theta_{jk}}. \end{aligned}$$

5. Results and Discussions

In this section, we first summarize results from the above model applied to different quality metrics. Then we examine quality gaps between surgeons and conclude that mitral valve repair rate is an informative measure of surgeon quality, and a metric that shows significant variation among providers. Finally, we use our model to show how patients of different demographics and levels of acuity benefit differently from elite surgeons.

5.1. Summary of Patient Characteristics

Between 2009 and 2012, 2,718 patients of New York state underwent elective mitral valve surgery at New York hospitals. Table 3 summarizes their characteristics. These data reveal some insights into patient choices. For example, the average travel distance is longer for patients under 60 than for older patients. This may be because younger patients are better able to travel. However, patients over 80 travelled on average further than patients in their 60s and 70s. This could be because their medical condition is often too delicate for a local hospital to handle. Overall, the average observed mortality rate was 1%, complication rate was 11%, readmission rate was 4%, and repair rate was 57%. However, all of these metrics worsen as the age of the patient increases.

Table 3 Summary of Patient Characteristics

	Patients	Travel Dist. (miles)		Repair Rate		Mortality Rate		Complication Rate		Readmission Rate	
	%	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Age											
below 50	12%	19	29	67%	47%	0%	5%	5%	22%	3%	17%
50 to 60	22%	22	29	72%	45%	1%	7%	5%	22%	2%	15%
60 to 70	27%	18	26	60%	49%	2%	13%	9%	29%	4%	21%
70 to 80	27%	17	22	47%	50%	1%	12%	15%	36%	5%	21%
above 80	13%	19	22	39%	49%	3%	16%	20%	40%	9%	28%
Gender											
male	55%	20	26	63%	48%	1%	11%	11%	31%	4%	20%
female	45%	18	27	51%	50%	1%	12%	11%	31%	5%	22%
Race											
asian	2%	18	22	42%	50%	2%	14%	9%	30%	8%	27%
black	8%	8	19	52%	50%	1%	12%	16%	37%	4%	19%
hispanic	5%	8	17	44%	50%	0%	0%	11%	32%	5%	22%
others	13%	18	22	64%	48%	1%	12%	10%	31%	10%	30%
white	73%	21	28	58%	49%	1%	11%	10%	30%	3%	18%
Total	2718	19	26	57%	49%	1%	11%	11%	31%	4%	21%

5.2. Measures of Quality

There were 188 surgeons in NY who performed mitral valve surgeries between 2009 and 2012. The average number of procedures per surgeon was 15 with a std. dev. of 24. To assess and compare the quality of these surgeons, we first calculate their observed (unadjusted) rates of mortality, complication, readmission and mitral valve repair. Figures 1 and 2 show the distributions of these four measures of quality. We see that mortality and readmission are rare events, with at least 70% of the surgeons showing no incidence. Complications are more common; only 48% of surgeons had no events and 18% of surgeons had complication rates above 20%. As shown in Figure 2, mitral valve repair rate has even larger variations; 19% of surgeons had repair rates of 0%, 20% of surgeons had repair rates of more than 90%, and the remaining surgeons had rates spread between these extremes. Although these rates are not risk adjusted, they suggest the possibility of large quality differences among surgeons.

To compute risk adjusted quality metrics, we estimate the quality model from Section 4. Table 4 summarizes the results. We first examine how patient characteristics affect outcomes. Not

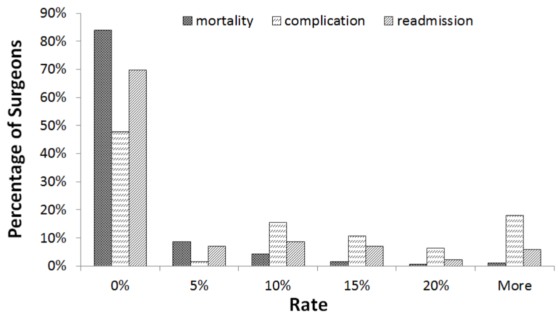


Figure 1 **Histogram of Observed Rates of Mortality, Complication and Readmission**

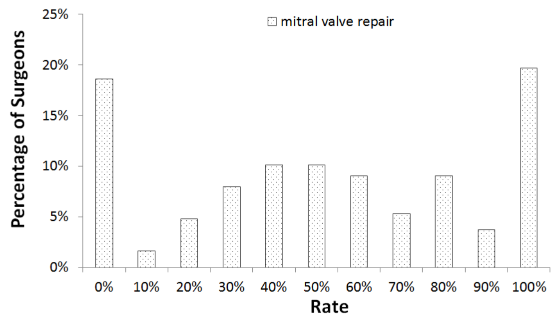


Figure 2 **Histogram of Observed Mitral Valve Repair Rate**

surprisingly, repair rate decreases and rates of mortality, complication and readmission increase as patient age increases. Compared with male patients, female patients are less likely to receive mitral valve repair, and are equally likely to have deaths, complications and readmissions. Compared with white patients, Hispanic patients are less likely to receive mitral valve repair, black patients are more likely to have complications and asian patients are more likely to have readmissions. Finally, mitral valve repair rate is lower for patients with comorbidities of atrial fibrillation, chronic lung disease, coagulopathy, diabetes or renal disease. These results are consistent with those of previous studies (see for example, Bolling et al., 2010 and Vassileva et al., 2013). Comorbidities such as atrial fibrillation and chronic heart failure affect other measures of quality as well, but the impact and significance level vary for different quality measures.

Surgical volume also influences outcomes. For mitral valve repair, we see that the coefficient of surgeon volume is positive and significant at the 5% level, suggesting that a surgeon's skill increases with volume. The coefficient of hospital volume is also positive, but is not significant at the 10% level. These results are consistent with those of previous studies (see for example, Birkmeyer et al., 2003 and Kilic et al., 2013). For quality metrics other than repair rate, we do not observe a statistically significant effect of volume, partly because the events measured are relatively rare. However, we do see that repair leads to a lower level of complication and mortality than replacement, as suggested by the medical literature (LaPar et al. 2010).[13]

5.3. Quality Gap among Surgeons

To compare surgeons using different measures of quality, we calculate the predicted rates of mitral valve repair, complication, readmission and mortality for a patient with population-median characteristics. We calculate the state average of each measure as the mean of all surgeons' rates for that

[13] We checked the quality model for these three measures without including repair as an explanatory variable, and obtained similar quality gaps between surgeons.

Table 4 Estimation Results of the Quality Model

	Repair		Mortality		Complication		Readmission	
	Coefficient	Std. Err.	Coefficient	Std. Err.	Coefficient	Std. Err.	Coefficient	Std. Err.
Surgical Volumes								
hosp_vol	0.05	0.17	0.08	0.39	0.16	0.20	−0.24	0.22
surg_vol	0.34**	0.14	0.06	0.19	−0.10	0.12	0.00	0.13
Patient Demographics								
age	−0.02***	0.00	0.03***	0.01	0.02***	0.00	0.02***	0.00
female	−0.30***	0.06	0.24	0.17	−0.03	0.08	0.13	0.10
black	−0.12	0.12	0.13	0.33	0.23*	0.14	0.14	0.20
hispanic	−0.24*	0.14	−4.81	839.01	−0.03	0.19	0.31	0.23
asian	−0.31	0.21	0.07	0.54	−0.18	0.30	0.60**	0.29
others	0.17	0.11	−0.06	0.31	−0.27*	0.16	0.65***	0.17
Comorbidities								
cm_af	−0.13**	0.06	−0.59***	0.19	−0.06	0.08	0.03	0.10
cm_alcohol	0.25	0.22	0.10	0.58	−0.07	0.29	−5.27	509.51
cm_anemdef	0.11	0.09	−0.82*	0.46	−0.23*	0.12	0.20	0.14
cm_arth	−0.27	0.18	−4.34	1651.95	−0.42	0.32	−0.13	0.35
cm_bldloss	0.02	0.41	−4.23	3341.71	−5.07	865.44	−4.31	1410.33
cm_chf	−0.06	0.42	0.53	0.51	1.77***	0.54	−4.30	1182.32
cm_chrnlung	−0.24***	0.07	−0.05	0.24	−0.01	0.10	0.27**	0.11
cm_coag	−0.13**	0.06	0.25	0.18	0.24***	0.08	0.15	0.10
cm_depress	0.07	0.11	−0.70	0.60	−0.20	0.17	0.06	0.19
cm_dm_all	−0.11	0.07	−0.41	0.30	0.17*	0.10	−0.17	0.13
cm_drug	−0.09	0.26	−4.44	2406.64	0.24	0.36	−4.50	677.26
cm_htn_c	0.05	0.06	−0.42**	0.19	−0.33***	0.08	0.01	0.10
cm_hypothy	0.04	0.10	−5.18	682.00	−0.08	0.14	−0.29	0.18
cm_liver	−0.06	0.23	1.35***	0.41	0.32	0.26	0.34	0.38
cm_lymph	−0.25	0.33	−4.17	2728.50	0.66**	0.34	0.06	0.53
cm_lytes	0.02	0.07	0.08	0.18	0.37***	0.08	0.17*	0.10
cm_mets	−0.49	0.66	−4.12	.	−4.69	2217.78	−4.25	7904.30
cm_neuro	0.08	0.15	−0.14	0.44	0.00	0.22	0.26	0.23
cm_obese	−0.17	0.11	−4.53	825.16	0.23*	0.14	0.04	0.18
cm_para	−0.12	0.23	0.67	0.43	0.84***	0.23	0.45	0.29
cm_perivasc	0.01	0.12	0.19	0.29	0.29**	0.13	0.03	0.19
cm_psych	−0.30	0.26	−4.11	2139.63	−0.20	0.45	−0.07	0.52
cm_pulmcirc	−0.15	0.66	1.43**	0.66	6.05	999.08	−4.61	2461.66
cm_renlfail	−0.25***	0.10	0.40*	0.23	1.01***	0.11	0.23	0.15
cm_tumor	−0.20	0.36	−4.34	3131.49	−0.16	0.49	0.05	0.58
cm_valve	−0.28	0.46	−0.03	0.65	1.84***	0.55	−4.21	1057.44
cm_wghtloss	−0.05	0.14	0.71***	0.24	1.04***	0.15	0.03	0.21
Others								
repair			−0.32*	0.19	−0.25***	0.09	0.11	0.11
σ_α	0.13	0.05	0.00	0.00	0.06	0.03	0.07	0.07
σ_β	0.11	0.04	0.00	0.00	0.00	0.00	0.03	0.04
constant	1.13	0.54	−4.14	1.39	−3.35	0.68	−2.31	0.76
log likelihood	−1543.01		−127.00		−696.78		−445.73	

Note. Robust standard errors are clustered by surgeon.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

measure.[14] Then we perform t-tests to determine whether a surgeon is statistically significantly different from the state average.

We summarize the performance of surgeons with high (above 55), medium (20-30) and low (below 15) volumes of elective mitral cases in Table 5. First, as we surmised earlier, because mortality is a rare event, it is very difficult to discern statistical differences among surgeons using quality metrics based on these events. As a result, no surgeon has a mortality rate statistically below the state average. This indiscernibility does not necessarily indicate surgeons have the same quality, especially since other quality metrics demonstrate sizable differences among surgeons. Second, we see that surgeons who perform well in mitral valve repair also perform well (or at least at the average) with regard to readmissions and complications. This means that choosing an elite surgeon based on mitral valve repair rate will not erode quality as measured by other metrics such as complication or readmission rate. Finally, as we showed in Table 4 and as indicated in the medical literature (LaPar et al., 2010), mitral valve repair is associated with lower complication and mortality rates during hospitalization, even after controlling for patient demographics and

[14] An alternative way to calculate the state average is to weight an individual surgeon's rate by his/her surgical volume. However, because surgeons are the focus of this analysis and surgical volume is endogenous and can change over time, we calculate the averages at the surgeon rather than the patient level.

comorbidities. Mitral valve repair rate focuses on both short-term and long-term quality of care, whereas the conventional measures of quality focus only on short-term quality of care. For these reasons, we argue that repair rate is a informative measure of surgeon quality for mitral valve surgery. Accordingly, we will focus on it for the rest of this section and in the policy analyses of the next section. Note that our methodology can be easily applied to any individual quality metric or a weighted average of multiple quality metrics.

Table 5 Quality Metrics of Surgeons with High, Medium and Low Volumes of Elective Mitral Cases

Volume	Repair Rate	Complication Rate	Readmission Rate	Mortality Rate
High	98.22%+++	0.65%+++	1.03%+	0.17%
	83.08%+++	1.49%	2.00%	0.04%
	85.61%+++	1.50%	0.71%+++	0.05%
	79.81%+++	2.42%	1.99%	0.08%
	70.09%++	3.51%	1.94%	0.05%
	64.26%	0.78%+++	2.31%	0.03%
	80.73%+++	1.75%	0.80%++	0.05%
	61.72%	1.71%	0.82%++	0.05%
	59.60%	3.88%	2.68%	0.05%
	87.68%+++	0.69%+++	4.11%	0.03%
Medium	67.52%+	4.37%	1.88%	0.04%
	37.58%	2.70%	1.12%	0.05%
	56.89%	4.49%−	1.05%	0.04%
	37.65%	2.29%	0.95%+	0.03%
	72.13%++	2.26%	1.97%	0.03%
	48.26%	3.00%	0.62%++	0.05%
	61.78%	3.07%	0.62%++	0.05%
	58.72%	1.73%	2.09%	0.03%
	69.71%+	1.17%+	3.47%	0.03%
	72.69%++	1.14%+	2.27%	0.03%
Low	76.92%++	2.13%	1.13%	0.04%
	68.95%+	1.22%+	3.53%	0.02%
	20.59%−−−	1.34%	1.58%	0.03%
	28.51%−	4.53%−	1.37%	0.04%
	46.31%	1.67%	0.86%+	0.04%
	56.12%	1.27%	2.24%	0.03%
	54.43%	2.29%	2.14%	0.03%
	44.27%	3.88%	4.58%	0.02%
	54.00%	1.38%	3.25%	0.03%
	53.83%	4.81%−	1.45%	0.04%

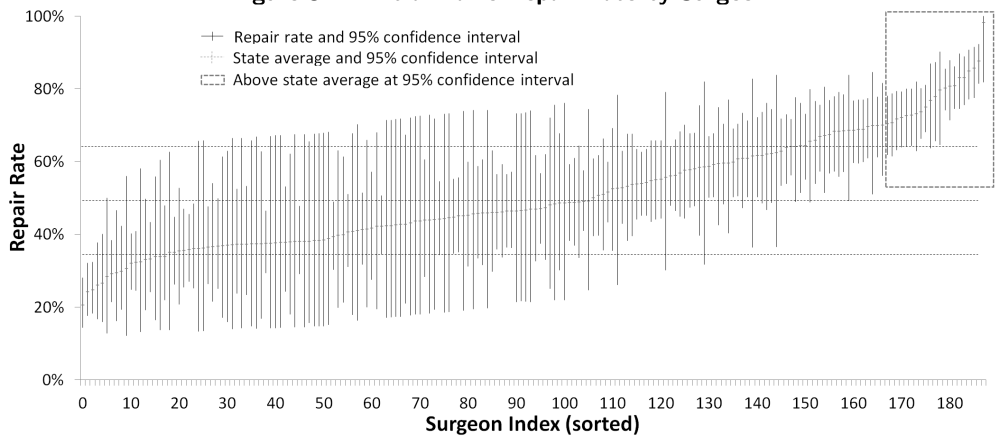
+++, ++, +: better than state average at 99%, 95% and 90% confidence level
−−−, −−, −: worse than state average at 99%, 95% and 90% confidence level

We illustrate the quality differences between surgeons visually in Figure 3 by plotting the population-median repair rates of individual surgeons along with their 95% confidence intervals.[15]

[15] The confidence intervals of individual surgeons are calculated based on standard errors of coefficients estimated by the quality model. The confidence interval of the state average is calculated numerically. We randomly simulated a population-median patient with error terms for all surgeons to calculate the state average. The simulation was performed 200 times to obtain the confidence interval.

The average repair rate across all surgeons is around 50%. The confidence intervals are heavily influenced by the number of cases. Surgeons with low volumes tend to have wide confidence intervals, and thus be either indistinguishable from the state average or below it. While most elite surgeons are high-volume surgeons, not all high-volume surgeons have high mitral valve repair rates.

Figure 3 Mitral Valve Repair Rate by Surgeon



5.4. Patient-centric Outcomes

Figure 3 clearly suggests that some surgeons are better than others. We informally label the top performers, who have repair rates above the state average (i.e., those in the dotted grey box of Figure 3), as "elite surgeons". What may not be so obvious is that the quality gap between elite surgeons and other surgeons is not uniform across patients of different demographics and levels of acuity. We illustrate this point by defining three groups of patients with different levels of acuity: "sick" (i.e., 90 years old with comorbidities) patients, "typical" (i.e., population-median) patients, and "healthy" (e.g., 30 years old with no comorbidities) patients. As shown in Figure 4, all three groups of patients benefit from visiting elite surgeons, but the magnitude of benefit is different, as indicated by the slopes of the curves. For the purpose of illustration, we calculate the gap in repair rate between a surgeon at the 90th-percentile and a surgeon at the median for these three groups of patients. The gap is 23.8% for typical (population-median) patients, 15.1% for the healthy patients and 16.7% for the sick patients.[16]

This indicates that neither the sick nor the healthy group of patients benefit as much as the population-median patients from visiting an elite surgeon. A plausible explanation for this is that many sick patients, with hard-to-repair valves, are likely to get a replacement regardless of which surgeon they visit, while many healthy patients, with easy-to-repair valves, are likely to get a

[16] The quality gap between the top (100th-percentile) surgeon and a surgeon at the median is also heterogeneous with the healthy patients benefit less than the sick and typical patients. Our optimization model to be discussed later captures the quality gap between any two surgeons across all patient types.

replacement from any above-median surgeon. The patients in between, however, tend to present "difficult but not impossible" repair challenges, and therefore are substantially more likely to receive a repair from an elite surgeon than from a median surgeon. Of course, these three sample groups are only illustrative. The heterogeneity in surgeon impact on patient outcomes will differ across all patient types.

Furthermore, in addition to benefitting differently from elite surgeons with respect to likelihood of getting a repair, patients with different demographics and levels of acuity also benefit differently from the same level of increase in repair rate. For example, younger and healthier patients have more years to live and, as a result, their lifetime benefit from a successful repair is greater than that of older and sicker patients. We can quantify the benefit to a given patient from a given surgeon by calculating the expected remaining life days, which depend on the patient's demographics and clinical conditions plus the likelihood of receiving a repair from that surgeon. To do this, we made the use of the results of Wang et al. (2015), who found that, depending on patients' age and comorbidities, patients' remaining life days increase by 4 to 15 days for each percentage point increase in mitral valve repair rate.

To illustrate the type of comparison this calculation makes possible, Figure 5 compares increases in expected life days for patients with diabetes and for patients with atrial fibrillation when they are treated by a 90th-percentile surgeon rather than a median surgeon. We see that the increase in expected life days from visiting the 90th-percentile surgeon is a concave function of patients' age for both comorbidities. In each case, patients of around 35 years old receive the most benefit from the elite surgeon, in part because their youth gives them many years to benefit from repaired valve.[17] There is also heterogeneity in surgeon impact across comorbidities. For instance, patients with diabetes benefit more than those with atrial fibrillation at the same age. Of course, these are only illustrative examples. The expected gain in life years will differ across all patient groups. Because it includes patient demographics and clinical descriptors as control variables, the model described in Section 4 can be used to estimate surgeon performance, and hence quality gaps, for all patient groups.

6. Managerial Implications

The American Heart Association/American College of Cardiology Valvular Heart Disease Guidelines state that mitral valve diseases should be repaired at a Center of Excellence (CoE).[18] To help patients identify these CoEs, existing studies and rating systems compare health care providers

[17] The quality gap between the 90th- and 50th-percentile surgeons is smaller for patients younger than 35 than it is for patients of 35 years old, so their lifetime benefit is also smaller.

[18] https://www.guideline.gov/content.aspx?id=48267

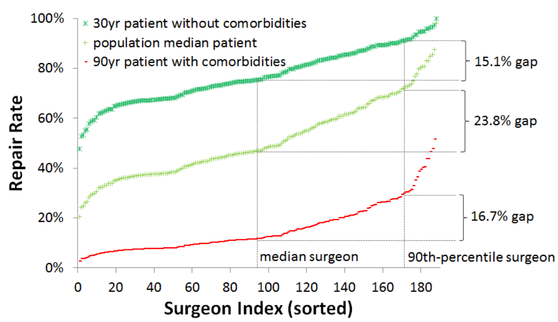


Figure 4 **Mitral Valve Repair Rate by Surgeon for Patients of Different Levels of Acuity**

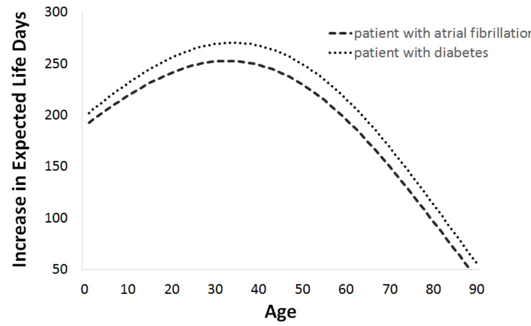


Figure 5 **Difference in Patients' Life Days between 90th- and 50th-percentile Surgeons**

based on various quality measures and provide an overall score or a ranking for each provider. However, these scores and rankings are reported only for the population average. This makes it appear that the benefit of visiting an elite surgeon is the same for all patients regardless of their demographics and levels of acuity. But our analysis above shows that this is not the case.

When only population-average quality information, which describes the benefits to an average patient from an elite surgeon, is provided, patients cannot accurately assess the benefit that they will get from visiting an elite surgeon. Primary and secondary care physicians (such as the cardiologists in our context) will also lack sufficiently detailed information to make patients appropriate referrals. Presumably, they will be inclined to refer all patients to elite surgeons unless non-medical (e.g., ability to travel) issues preclude this. This is neither good advice for patients nor an efficient use of limited health care resources. Indeed, it is very likely that elite surgeons will not have sufficient capacity to accommodate all patients, which will result in allocation of their capacity on the basis of which patients are willing/able to wait longest. The implication is that patient-centric quality information, which describes the benefits to patients from an elite surgeon as a function of their demographics and clinical characteristics, can help primary and secondary care physicians to make better referrals and patients to make more informed decisions regarding the choice of surgeons.

In this section, we examine the value of using population-average and patient-centric quality information to guide mitral valve patients to surgeons. Since patients choose surgeons on the basis of many factors, it is impossible to predict the precise impact of providing more accurate outcome information. Therefore, to evaluate the magnitude of the opportunity offered by the two types of information, we compute an upper bound on the improvement in patient outcomes by assigning patients to surgeons in a manner that maximizes total patient utility. Furthermore, since outcomes can be improved by having more patients treated by elite surgeons, we also examine the impact of increasing surgeon capacities. In practice, this could be achieved in many ways, including increasing operating room capacity, altering allocation of operating room time across services, and reducing administrative or teaching duties of surgeons. These calculations allow us to contrast the value of making outcome information more specific to individual patients with the value of enabling the best surgeons to treat more patients.

6.1. Problem Formulation

To state an optimization problem to maximize total patient utility, we define the utility to patient i from treatment by surgeon j as the expected number of remaining life days ($LifeDays_{ij}$) minus a linear function of the distance ($Dist_{ij}$) the patient must travel to be treated by the surgeon. Letting C_j represent the capacity of surgeon j and x_{ij} be a binary decision variable that equals one if patient i chooses surgeon j, we can express the problem of maximizing total patient utility as the following combinatorial optimization problem:

$$\begin{aligned} \text{maximize} \sum_i \sum_j (LifeDays_{ij} - \beta Dist_{ij}) X_{ij} & \\ \text{s.t.} \sum_i X_{ij} \le C_j & \\ \sum_j X_{ij} = 1 & \\ X_{ij} \in \{0,1\} & \end{aligned} \tag{P}$$

Obviously this model simplifies reality in many ways. Most importantly, it assumes the relative weights patients place on life days and distance are the same across all patients. In reality, patient preferences vary depending on age, medical condition and other individual characteristics. However, our model may still serve as a reasonable approximation. We explain why below.

First, note that the results of Figures 4 and 5 suggest that the medical benefits of traveling to an elite surgeon are smaller for the young and healthy or old and sick extremes than for patients in the middle. So, suppose we choose the coefficients on life days and distance to represent a “median patient” (i.e., one in the middle range of the age and acuity scale). As such, they will accurately characterize the patients who benefit most from traveling and hence are most likely to be influenced by accurate outcome information. However, these “median” coefficients will underestimate the willingness to travel by young, healthy patients, and will overestimate the willingness to travel by old, sick patients. Since the results in Figures 4 and 5 show that the health benefits from these two groups are similar, the two types of biases discussed above will likely offset each other, and therefore, the aggregate benefits will not be distorted significantly by this bias.

Having said this, there are very likely other patient characteristics that influence willingness to travel for better health that are not captured by our model. Wealth is one such characteristic. Wealthy people are likely to be less averse to travel than are poor people, because wealthy people can better bear the associated costs. Hence, we would expect more rich people who benefit only modestly, and fewer poor people who would benefit substantially, to travel to elite surgeons than our model predicts. This implies that the mix of patients who actually wind up travelling to elite surgeons will receive smaller total benefits than the mix of patients generated by our optimization model due to wealth constraints. Of course it may be possible to address such distortions in patient mix via policy measures (e.g., travel subsidies). Estimating an upper bound on the benefits from better use of information can inform the discussion of which policies make sense to pursue.

We begin our analysis by solving (P) under the assumption that patient utility is computed using population-average outcome information. To do this, we let $\overline{Rate}_j$ denote the population-average repair rate for provider j and compute $LifeDays_{ij} = q_i\overline{Rate}_j$, where q_i indicates expected increase in life days for patient i per percentage point increase in repair rate.[19] We then solve (P) under the assumption that patient utility is computed using patient-centric quality information. We do this by defining $Rate_{ij}$ to be the likelihood of a repair for patient i treated by surgeon j, and computing $LifeDays_{ij} = q_i Rate_{ij}$. Finally, to evaluate the value of enabling elite surgeons to treat more patients, we solve (P) under both population-average information and patient-centric information when surgeon capacity, C_j, is increased by various percentages from zero to 40%.

To parameterize these optimization problems, we first note that $LifeDays_{ij}$ and $Dist_{ij}$ have different units, so we scale them by their largest observed values such that both variables take values between zero and one. The expected increase in life days for patients at different ages and with different comorbidities is obtained from Wang et al. (2015). Because we cannot directly observe the capacity of each surgeon, we use as a proxy the maximum monthly volume multiplied by the number of months under this study. For surgeons who did not operate in every month from 2009-2012, we multiply the maximum by the number of months in which they did operate. The annual mitral volume of surgeons and the arrival rate of patients of different demographics and levels of acuity were quite stable over our study period. Because there is flexibility in scheduling treatment for elective patients, we focus on the assignment problem over the entire study period rather than over a year or a month. We conduct sensitivity analyses by varying the values of weight on distance (β ranging from 0.01 to 100).

6.2. Numerical Analyses

We consider twenty-five different variants of problem (P) (five capacity levels and five distance weights) under both population-average and patient-centric information. For each combination, we used AMPL to solve the assignment problem and obtain the optimal solution. The resulting

[19] Allowing q_i to vary by patient is a more conservative approach to defining population average quality information than using a constant coefficient q. In the former, patient benefits in terms of life days vary by patient types due to varying q_i, which is accessible through existing medical literature (see a survey of literature in Wang et al., 2015); in the latter, patient benefits in terms of life days do not vary across patients.

comparison of the relative impact of using patient-centric information and increasing surgical capacity is provided in Table 6. Columns (1) and (2) are input parameters, with column (1) showing surgeon capacity, where the baseline is actual capacity as described above, and column (2) is the coefficient β, which represents the increase in life days required for a patient to be willing to travel one additional mile. Columns (3) and (4) show the expected number of repairs and average travel distance per patient in the optimal solution with population-average information, and columns (5) and (6) show these two measures with patient-centric information.

Table 6 Comparison of the Effectiveness of Patient-Centric Information and Capacity Increase

		Population-average Information		**Patient-centric Information**	
% Increase of Surgeon Capacity (1)	Equivalent life days per mile (2)	Expected Number of Repairs (3)	Average Distance (miles) per Patient (4)	Expected Number of Repairs (5)	Average Distance (miles) per Patient (6)
baseline	0.04	2,132	46	2,222	62
	0.4	2,118	28	2,206	29
	4	2,084	17	2,156	18
	40	1,906	11	1,940	11
	400	1,539	10	1,544	10
10%	0.04	2,156	55	2,248	71
	0.4	2,142	26	2,228	28
	4	2,111	18	2,190	18
	40	1,912	11	1,957	11
	400	1,550	10	1,554	10
20%	0.04	2,172	60	2,270	76
	0.4	2,164	26	2,248	28
	4	2,131	18	2,211	18
	40	1,916	11	1,957	11
	400	1,559	10	1,562	10
30%	0.04	2,191	67	2,291	82
	0.4	2,181	26	2,268	28
	4	2,149	18	2,222	18
	40	1,920	11	1,960	11
	400	1,566	10	1,568	10
40%	0.04	2,211	68	2,311	86
	0.4	2,200	26	2,286	29
	4	2,166	19	2,246	18
	40	1,919	12	1,960	11
	400	1,572	10	1,572	10
Actual		1,557	19		

To interpret these results, we first recall that the 2,718 New York mitral valve patients we considered in our empirical analysis traveled an average distance of 19 miles to receive surgery and that 1,557 had their valves successfully repaired. In terms of clinical outcomes this is comparable to the case with population-average information and β equal to 400, which results in 1,539 repairs and an average travel distance of 10 miles. This might suggest that actual patients behaved as if they were strongly travel averse.

But it doesn't seem reasonable that an average patient would be willing to sacrifice 400 days of life to avoid traveling one extra mile. Indeed, many studies of patient choices of health care providers have found that patients are willing to travel for better care. For example, Finlayson et al. (1999) found that 80% of patients are willing to bypass a local hospital in favor of a more distant regional hospital if this will result a 20% reduction in the likelihood of mortality. Groux et al. (2013) found that 36-41% of cancer patients surveyed were willing to travel any distance to receive the best available treatment. With these in mind, we conclude that the behavior of New York patients is not primarily driven by travel aversion.

A second possible explanation for the failure of patients to travel to better care is that their choices are limited by their insurance providers. But our empirical analysis showed that patients with non-restrictive coverage (e.g., Medicare) are almost as likely to receive inferior local treatment as are patients with more restrictive (e.g., HMO) coverage.[20] So, while insurance may play a role, we don't think it is the dominant driver of surgeon choice.

A third explanation is that patients either fail to find outcome data, or, if they do, fail to understand or trust it. Without information with which to distinguish the performance of different surgeons, patients fall back on other criteria like convenience or familiarity when choosing a surgeon. But the studies cited above on patient willingness to travel imply that patients are willing to act on quality information if they have it. This observation has led many scholars (e.g., Emmer and Schlesinger, 2016, Sinaiko et al., 2012) to conjecture that better presentation of medical quality information would cause patients to make more use of it in their decisions. A feature that is often cited as desirable is personalized information tailored to individual patients (Paddock et al., 2015, Sinaiko et al., 2012). The implication is that more usable information could make the outcomes in Table 6 possible.

To get a sense of which outcomes in Table 6 are plausible, we note that Finlayson et al. (1999) reported that most patients would travel to a regional hospital if it reduced mortality rate by 20%. The baseline mortality rate and travel distance were not specified. But we can make a rough estimation of the β coefficient if we assume that the mortality rate at the local hospital is 1%, the extra travel distance to the regional hospital is 35 miles, and a patients remaining life expectancy is 20 years. Under these assumptions, the increase in life days from a 20% reduction of the 1% mortality rate is 20 yrs $\times$ 365 days/yr $\times$ 0.2% = 14.6 days. Because patients would choose to travel to the regional hospital under these conditions, this means that β is less than 14.6 life days $\div$ 35 miles = 0.42 life days per mile. So for purposes of interpretation, we will use $\beta = 0.4$.

[20] Based on NY data from 2009-2012, we calculated that 64% of Medicare and 63% of HMO patients were treated by non-elite surgeons.

Table 6 shows that the expected number of repairs under population-average information is 2,118, which represents a 36% increase over the actual number of 1,557. To achieve this, patients would have to travel an average of 28 miles as opposed to 19 miles. If patient-centric information is used, then the expected number of repairs increases to 2,206, which is a 42% increase over the actual number. The additional 6% increase relative to the population-average information case is the result of guiding more of the patients who benefit most to elite surgeons. Achieving this more efficient patient mix between elite and non-elite surgeons requires only one more mile in average patient travel distance, to 29 miles.

Another way to interpret the value of patient centric information is to look at the impact of increasing surgeon capacity under population-average information. For instance, when $\beta = 0.4$, a 40% increase in surgeon capacity results in 2,200 repairs. This is almost identical to the 2,206 repairs that can be achieved with patient-centric information and no increase in surgeon capacity. In the population-average case, the improvement is achieved by enabling the elite surgeons with the highest repair rates to treat more patients. In the patient-centric case, the improvement is achieved by having the elite surgeons treat the right patients. In a very practical sense, more targeted information is equivalent to a 40% increase in surgeon capacity.

7. Conclusion

The past decade has seen increasing efforts by the US government, payers and health care providers to improve health care quality and control health care costs. Significant energy has been devoted to improving information transparency in the hope of guiding patients to the right providers. In addition to being complex and difficult to use, currently available quality information about healthcare providers is based on population averages. Our results show that population-average information is valuable to patients, but that patient-centric quality information is even more valuable. Used properly, patient-centric information can guide patients to the providers that are best for them and for society as a whole.

This study addresses the challenges of measuring provider quality and offers insights into better matching of patients with care. Using mitral valve surgery as the clinical setting, we studied the quality of cardiac surgeons in NY based on different quality metrics. We used a multilevel probit model to capture hospital and surgeon volume effects, as well as their specific effects, on patient outcomes. We corrected for both observed differences in patient mix and potential selection bias using distance-based instruments. This analysis shows that twenty one New York surgeons achieved repair rates statistically significantly above the state average, but that patients of different demographics and levels of acuity benefit differently from these elite surgeons.

We compared the effectiveness of providing patient-centric quality information and that of increasing surgeon capacity when patients are optimally assigned to surgeons with capacity constraints. We estimate that providing patient-centric quality information offers societal benefits comparable to those achievable with a 40% increase in surgeon capacity.

Beyond the obvious benefits to patients, patient-centric information offers potential benefits to physicians and payers. Cardiologists in particular can make use of patient-centric information to make better referrals. With population-average information, cardiologists are inclined to refer all patients to elite surgeons. However, armed with patient-centric information, they will appropriately refer some patients to non-elite surgeons, because these surgeons' quality is comparable to that of elite surgeons for some patients. This will not only help spread the workload across surgeons but also help non-elite surgeons improve their skills by giving them some patient volume.

Finally, Wang et al. (2015) have shown that, although hospitals with elite surgeons tend to charge a premium on the surgical procedure itself, their lifetime treatment costs are likely to be lower due to avoidance of complications and readmissions. Thus, patient-centric information can help payers to save money by guiding patients (via reduced co-pays for patients and/or value-based compensation of hospitals and surgeons) to providers that offer both better clinical outcomes and lower lifetime costs.

References

Amemiya, Y. (1990). Two-stage instrumental variable estimators for the nonlinear errors-invariables model. Journal of Econometrics. 44(3):311-332.

Barr, J. K., Giannotti, T. E., Sofaer, S., Duquette, C. E., Waters, W. J., Petrillo, M. K. (2006). Using public reports of patient satisfaction for hospital quality improvement. Health Services Research, 41(3p1), 663-682.

Bastani, H., Goh, J., Bayati, M. (2015). Evidence of strategic behavior in Medicare claims reporting. Working Paper.

Batt, R. J., Terwiesch, C. (2015). Waiting patiently: An empirical study of queue abandonment in an emergency department. Management Science, 61(1), 39-59.

Bavafa, H., Hitt, L. M., Terwiesch, C. (2013). Patient portals in primary care: Impacts on patient health and physician productivity. Working Paper.

Birkmeyer, J. D., Siewers, A. E., Finlayson, E. V., Stukel, T. A., Lucas, F. L., Batista, I., ..., Wennberg, D. E. (2002). Hospital volume and surgical mortality in the United States. New England Journal of Medicine, 346(15), 1128-1137.

Birkmeyer, J. D., Stukel, T. A., Siewers, A. E., Goodney, P. P., Wennberg, D. E., Lucas, F. L. (2003). Surgeon volume and operative mortality in the United States. New England Journal of Medicine, 349(22), 2117-2127.

Bock, R. D., Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46(4), 443-459.

Bolling, S. F., Li, S., O'Brien, S. M., Brennan, J. M., Prager, R. L., Gammie, J. S. (2010). Predictors of mitral valve repair: clinical and surgeon factors. The Annals of Thoracic Surgery, 90(6), 1904-1912.

Brooks, J. M., Irwin, C. P., Hunsicker, L. G., Flanigan, M. J., Chrischilles, E. A., Pendergast, J. F. (2006). Effect of dialysis center profit status on patient survival: A comparison of risk - adjustment and instrumental variable approaches. Health Services Research, 41(6), 2267-2289.

Carroll, R. J., Horn, S. D., Soderfeldt, B., James, B. C., Malmberg, L. (1995). International comparison of waiting times for selected cardiovascular procedures. Journal of the American College of Cardiology, 25(3):557-563.

Chassin, M. R., Kosecoff, J., Park, R. E., Winslow, C. M., Kahn, K. L., Merrick, N. J., ..., Brook, R. H. (1987). Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. Jama, 258(18), 2533-2537.

Daneshmand, M. a, Milano, C. a, Rankin, J.S., et al. (2009). Mitral valve repair for degenerative disease: A 20-year experience. The Annals of Thoracic Surgery, 88(6):1828-37.

Dimick, J. B., Welch, H. G., Birkmeyer, J. D. (2004). Surgical mortality as an indicator of hospital quality: The problem with small sample size. Jama, 292(7), 847-851.

Dranove, D., Kessler, D., McClellan, M., Satterthwaite, M. (2003). Is more information better? The effects of "Report Cards" on health care providers. Journal of Polictical Economics, 3(3), 555-588.

Emmert, M., Schlesinger, M. (2016). Hospital Quality Reporting in the United States: Does Report Card Design and Incorporation of Patient Narrative Comments Affect Hospital Choice?. Health Services Research.

Fasken, L. L., Wipke-Tevis, D. D., Sagehorn, K. K. (2001). Factors associated with unplanned readmissions following cardiac surgery. Progress in Cardiovascular Nursing, 16(3), 107-115.

Freeman, M., Savva, N., Scholtes, S. (2015). Gatekeepers at work: An empirical analysis of a maternity unit. Working Paper.

Gammie, J. S., Sheng, S., Griffith, B. P., Peterson, E. D., Rankin, J. S., O'Brien, S. M., Brown, J. M. (2009). Trends in mitral valve surgery in the United States: Results from the Society of Thoracic Surgeons Adult Cardiac Surgery Database. The Annals of Thoracic Surgery, 87(5):1431-7.

Gerteis, M. (1993). Through the patient's eyes: Understanding and promoting patient-centered care.

Gibbons, R. D., Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. Biometrics, 1527-1537.

Gilman, M., Adams, E. K., Hockenberry, J. M., Wilson, I. B., Milstein, A. S., Becker, E. R. (2014). California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Affairs, 33(8), 1314-1322.

Gupta, P. K., Fernandes-Taylor, S., Ramanan, B., Engelbert, T. L., Kent, K. C. (2014). Unplanned readmissions after vascular surgery. Journal of Vascular Surgery, 59(2), 473-482.

Healthgrades (2015). Healthgrades 2015 report to the nation: Making smart choices.

Huckman, R. S., Pisano, G. P. (2006). The firm specificity of individual performance: Evidence from cardiac surgery. Management Science, 52(4), 473-488.

Institute of Medicine (US). Committee on Quality of Health Care in America. (2001). Crossing the quality chasm: A new health system for the 21st century. National Academy Press.

Iribarne, A., Chang, H., Alexander, J. H., et al. (2014). Readmissions after cardiac surgery: Experience of the NIH/CIHR cardiothoracic surgical trials network. The Annuals of Thoracic Surgery, 98(4), 1274-1280.

Berry Jaeker, J., Tucker, A. (2016). Past the point of speeding up: The negative effects of workload saturation on efficiency and quality. Management Science.

Kattan, M. W., Vickers, A. J. (2004, August). Incorporating predictions of individual patient risk in clinical trials. In Urologic Oncology: Seminars and Original Investigations (Vol. 22, No. 4, pp. 348-352). Elsevier.

KC, D., Staats, B. R. (2012). Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. Manufacturing & Service Operations Management, 14(4), 618-633.

KC, D., Staats, B. R., Gino, F. (2013). Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. Management Science, 59(11), 2435-2449.

KC, D., Terwiesch, C. (2011). The effects of focus on performance: Evidence from California hospitals. Management Science, 57(11), 1897-1912.

Keeler, E. B., Rubenstein, L. V., Kahn, K. L., Draper, D., Harrison, E. R., McGinty, M. J., ... , Brook, R. H. (1992). Hospital characteristics and quality of care. Jama, 268(13):1709-1714.

Kent, D. M., Hayward, R. A. (2007). Limitations of applying summary results of clinical trials to individual patients: The need for risk stratification. Jama, 298(10), 1209-1212.

Kilic, A., Shah, A. S., Conte, J. V., Baumgartner, W. A., Yuh, D. D. (2013). Operative outcomes in mitral valve surgery: Combined effect of surgeon and hospital volume in a population-based analysis. The Journal of Thoracic and Cardiovascular surgery, 146(3), 638-646.

Kim, S. H., Chan, C. W., Olivares, M., Escobar, G. (2014). ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. Management Science, 61(1), 19-38.

Kolstad, J. T., 2013. Information and quality when motivation is intrinsic: Evidence from surgeon report cards. The American Economic Review, 103(7), pp.2875-2910.

Kravitz, R. L., Duan, N., Braslow, J. (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Quarterly, 82(4), 661-687.

LaPar, D. J., Hennessy, S., Fonner, E., Kern, J. a, Kron, I. L., Ailawadi, G. (2010). Does urgent or emergent status influence choice in mitral valve operations? An analysis of outcomes from the Virginia Cardiac Surgery Quality Initiative. The Annals of Thoracic Surgery, 90(1):153-60.

Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., Bratzler, D. W. (2007). Public reporting and pay for performance in hospital quality improvement. New England Journal of Medicine, 356(5), 486-496.

Lorch, S. A., Passarella, M., Zeigler, A. (2014). Challenges to measuring variation in readmission rates of neonatal intensive care patients. Academic Pediatrics, 14(5), S47-S53.

McClellan, M., McNeil, B. J., Newhouse, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality: analysis using instrumental variables. Jama, 272(11), 859-866.

McConnell, K. J., Newgard, C. D., Mullins, R. J., Arthur, M., Hedges, J. R. (2005). Mortality benefit of transfer to level I versus level II trauma centers for head injured patients. Health Services Research, 40(2), 435-458.

Merkow, R. P., Ju, M. H., Chung, J. W., Hall, B. L., Cohen, M. E., Williams, M. V., ..., Bilimoria, K. Y. (2015). Underlying reasons associated with hospital readmission following surgery in the United States. Jama, 313(5), 483-495.

Morgan, A., Khan, A., Amin, T. (2013). Challenges in evaluating all-cause hospital readmission measures for use as national consensus standards. The Permanente Journal, 17(4), 14.

Paddock, S. M., Adams, J. L., de la Guardia, F. H. (2015). Better-than-average and worse-than-average hospitals may not significantly differ from average hospitals: an analysis of Medicare Hospital Compare ratings. BMJ Quality and Safety, 24(2), 128-134.

Pracht, E. E., Tepas, J. J., Celso, B. G., Langland-Orban, B., Flint, L. (2007). Survival advantage associated with treatment of injury at designated trauma centers: A bivariate probit model with instrumental variables. Medical Care Research and Review, 64(1), 83-97.

Ramdas, K., Saleh, K., Stern, S., Liu, H. (2014). Variety and experience: Learning and forgetting in the use of surgical devices. Working Paper.

Savage, E. B., Ferguson, T. B., DiSesa, V. J. (2003). Use of mitral valve repair: Analysis of contemporary United States experience reported to the Society of Thoracic Surgeons National Cardiac Database. The Annals of Thoracic Surgery, 75(3):820-5.

Sinaiko, A. D., Eastman, D., Rosenthal, M. B. (2012). How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. Health Affairs, 31(3), 602-611.

Society of Thoracic Surgeons (2016). Online STS risk calculator.

Song, H., Tucker, A. L., Murrell, K. L. (2015). The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. Management Science.

US Food and Drug Administration (2013). Paving the way for personlized medicine: FDA's role in the new era of medical product development.

Vassileva, C. M., Shabosky, J., Boley, T., Markwell, S., Hazelrigg, S. (2012). Cost analysis of isolated mitral valve surgery in the United States. The Annals of Thoracic Surgery, 94:1429-1436.

Vassileva, C. M., Mishkel, G., McNeely, C., Boley, T., Markwell, S., Scaife, S., Hazelrigg, S. (2013). Long-term survival of patients undergoing mitral valve repair and replacement: A longitudinal analysis of Medicare fee-for-service beneficiaries. Circulation, 127(18):1870-6.

Vassileva, C. M., Boley, T., Standard, J., Markwell, S., Hazelrigg, S. (2013). Relationship between patient income level and mitral valve repair utilization. Heart Surgery Forum, 16(2):E89-95.

Wang, G., Li, J., Hopp, W., Fazzalari, F., Bolling, S. (2015). Cost-effectiveness of referring patients to centers of excellence for mitral valve surgery. Working Paper. Ross School of Business, University of Michigan, Ann Arbor, MI.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

Appendix. Evaluation of the Instrument

To test if distance-based instruments correlate with the sickness of patients, we analyze if patients living closer to high-volume providers or having higher expected volumes are sicker or healthier. The results are summarized below. We do not see evidence of such correlations.

Table A1 Relationship Between Distance to High-volume Hospitals and Patient Characteristics

Distance (in mile)	Number of Patients	Patients' Mean Age	Number of Chronic Conditions	Number of Comorbidities
below 5	499	65.6 (12.8)	6.4 (2.7)	2.2 (1.5)
5 to 30	1274	64.7 (12.9)	6.5 (2.5)	2.2 (1.4)
above 30	945	64.9 (12.2)	6.7 (2.6)	2.3 (1.5)
Total	2718	64.9 (12.7)	6.5 (2.6)	2.2 (1.5)

Note: High-volume hospitals are those with more than 70 surgeries per year. Distance is defined as the Euclidean distance between the centroid of a patient's zip code and that of the nearest high-volume hospital. Standard deviations are displayed in parentheses.

Table A2 Relationship Between Expected Hospital Volume and Patient Characteristics

Expected Volume	Number of Patients	Patients' Mean Age	Number of Chronic Conditions	Number of Comorbidities
Low	906	64.8 (12.2)	6.7 (2.6)	2.3 (1.5)
Medium	898	64.6 (12.8)	6.6 (2.7)	2.3 (1.5)
High	914	65.4 (13.0)	6.4 (2.5)	2.4 (1.4)
Total	2718	64.9 (12.7)	6.5 (2.6)	2.2 (1.5)

Note: Expected volumes are categorized into low, medium and high, with roughly one third of patients in each category. Standard deviations are displayed in parentheses.