PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT

Vol. 30, No. 4, April 2021, pp. 839–863
ISSN 1059-1478 | EISSN 1937-5956 | 21 | 3004 | 0839
DOI 10.1111/poms.13256

Examining the Impacts of Clinical Practice Variation on Operational Performance

Seokjun Youn*
Department of Management Information Systems, Eller College of Management, The University of Arizona, 1130 E Helen St., McClelland Hall 430CC, Tucson, Arizona 85721, USA, syoun@arizona.edu

Gregory R. Heim
Department of Information and Operations Management, Mays Business School, Texas A&M University, 4217 TAMU, Wehner 320U, College Station, Texas 77843, USA, gheim@mays.tamu.edu

Subodha Kumar
Department of Marketing and Supply Chain Management, Fox School of Business, Temple University, 1801 Liacouras Walk, Alter Hall 530, Philadelphia, Pennsylvania 19122, USA, subodha@temple.edu

Chelliah Sriskandarajah
Department of Information and Operations Management, Mays Business School, Texas A&M University, 4217 TAMU, Wehner 320U, College Station, Texas 77843, USA, chelliah@mays.tamu.edu

This study explores whether and how lower variations in clinical practice relate to hospital operational performance. This relation is critical to the overall search for pathways that will allow the healthcare industry to bend the cost curve, implying significant implications for practice and regulators. We define practice variation as all variation not resulting from patient mix and construct a novel measure using inpatient discharge data for each patient cohort having an identical medical condition. Hospitals in our dataset show a broad practice variation spectrum. Using statistical process control (SPC) as a theoretical lens, we hypothesize the negative impacts of practice variation on operational performance. We also consider intervening impacts of hospital quality evaluations on the relationship. Analyzing data of six years from hospitals in NY and FL states using a dynamic panel system Generalized Method of Moments estimator, we find that higher practice variation relates to longer average patient length-of-stay and higher total cost per capita. This phenomenon is even stronger when a hospital provides services with higher quality in patient experience because such a hospital tends to provide more responsive care to patients, which is often resource-intensive. By delving into granular dimensions of practice variation based on detailed charge data, we find that higher care-delivery practice variation (i.e., the provision of healthcare) is directly associated with poor operational performance. We also find that pursuing higher quality measures may be harmful to some hospital operational performance measures as they have combined effects with the test-ordering practice variation (i.e., detecting disease and monitoring its status). Taken together, these findings imply that careful attention to the two dimensions of practice variation and the nuanced joint relationship with quality measures may address the trade-off between high quality and low cost, and provide room for improvement in practice, ultimately reducing waste in the healthcare industry. Our measure of practice variation also contributes since it enables researchers and managers to rigorously measure and visualize the status of hospitals' practice variation linked to hospital operational performance.

Key words: healthcare; clinical practice variation; operational performance; process and experiential quality; empirical operations
History: Received: October 2017; Accepted: August 2020 Sergei Savin after three revisions.

1. Introduction

The U.S. healthcare system is experiencing excessive national expenditures and a history of often poor health outcomes (Mossialos et al. 2016, OECD 2016). The United States spent 17.8% of its gross domestic product on healthcare in 2016, dominating other high-spending countries such as the 12.4% in Switzerland, but without buying notably better quality or outcomes (Papanicolas et al. 2018). Based on several domains of potential waste in healthcare[1] (Berwick and Hackbarth 2012), recent studies estimate that a significant portion, about 25%, of total healthcare spending in the United States is waste, ranging from $690 billion to $935 billion every year (Lallemand et al. 2012, Shrank et al. 2019). Naturally, there has been a sustained call for efforts to reduce such waste of healthcare resources. Of the wasteful spending, unwanted clinical variation was found to be a major contributing factor, with \$265 billion attributed to unnecessary services or missed opportunities (IOM 2012), which motivates this study.

Although wide variation in payments for Medicare beneficiary services suggests room for practice-level improvement (Clough et al. 2015, 2016), working with healthcare providers to reduce variation is not easy. No two patients are alike. Differences in patient characteristics, disease course, and comorbidities make it challenging to drill down into the root causes of this variation and provide meaningful feedback to healthcare providers. Even worse, current methods to address variation have made little progress (QURE Healthcare 2018), thus we rarely observe systematic approaches to measure clinical variation, other than conceptual studies and related discussions. Given such challenges, this study aims to close the research gap by proposing a novel approach to measure clinical variation and by empirically examining links between the level of clinical variation inside a hospital and its operational performance (measured as patient length-of-stay and cost metrics per capita). Considering the high costs and inconsistent care quality of the U.S. healthcare system in an environment of limited resources (Shi and Singh 2012), it is societally meaningful to investigate whether and how lessening variations in clinical practice might relate to better operational performance.

1.1. Research Questions

A major research opportunity exists in precisely measuring clinical practice variation observed during the process of delivering healthcare and identifying its impacts on operational performance. Existing literature classifies practice variations into warranted and unwarranted (Appleby et al. 2011, Clancy and Cronin 2005). Case-mix-index (CMI)[2] is often used to reflect variation due to a hospital's patient mix, which we will call warranted variation. CMI controls for diversity, clinical complexity, and resource needs of the patient population in a hospital, enabling fair hospital-to-hospital comparisons of medical operations. However, certain care delivery variations exist even after considering a hospital-level patient mix; thus CMI cannot capture all aspects of variation. Wennberg (2002) called all clinical variation not resulting from the patient-mix by the term unwarranted variation. Unwarranted variation can result from variation in care providers' decisions, variation to customize care, or variation in medical procedure supplies. Because prior work focuses on CMI, the impact of these unwarranted variations on hospital performance is not clear.

Surprisingly, however, no CMI-like nationally standardized metric for unwarranted variations exists. We thus identify a research opportunity for measuring unwarranted-type variations in medical practice for patient cohorts having the identical condition (hereafter, we will use the term *practice variation* to refer to the overall contribution of these three unwarranted variation facets). Subsequently, to bridge this research gap, our first research question is: *(i) How can researchers measure practice variation for a patient cohort with reasonable precision?* Next, relying on SPC theory (Oakland 2007, Wheeler et al. 1992), we posit that high practice variation may harm hospital operational performance. We close the research gap by asking: *(ii) Does lower practice variation of a hospital (or within patient cohort care episode groups) necessarily relate to better operational performance?*

The U.S. healthcare systems have incorporated several quality initiatives to encourage better care quality. We believe these quality initiatives may change participating healthcare providers' behaviors. Thus, we also consider the possible effects of these quality initiatives on the link between practice variation and hospital performance, which is unexplored. Process quality and experiential quality are two key measures of care quality that are publicly reported, span salient aspects of operations, and are used by CMS to evaluate healthcare providers (Sadeghi et al. 2012). *Process Quality* concerns how well a hospital adheres to evidence-based medical guidelines to diagnose and treat patients (Andritsos and Tang 2014, Chandrasekaran et al. 2012, Nair et al. 2013, Theokary and Ren 2011). *Experiential Quality* aggregates the reports of patients about their observations of and participation in healthcare (Sadeghi et al. 2012), and thus relates to external perceptions of care quality from a patient's perspective (Donabedian 1980, Li and Benton 1996). Both process quality and experiential quality are relevant exogenously mandated measures that are closely related to the U.S. healthcare systems.

Given trade-offs between quality and cost, exploring the roles of these quality initiatives on the link between practice variation and operational performance is essential since collecting and reporting such quality measures are often unnecessarily time-consuming and costly tasks for healthcare providers (Casalino et al. 2016). Under the U.S. government's recent emphasis on simultaneous pursuit of "The Triple Aim: Care, Health, and Cost,"[3] (Berwick et al. 2008), we examine whether, and if so to what extent, the pursuit of high process quality or experiential quality amplifies the negative link between the practice variation and operational performance. We thus investigate our next research question: *(iii) Do quality initiatives affect the relationship between practice variation and operational performance?* Our study sheds light on potential detrimental effects of process and experiential quality on operational performance when considered jointly with practice variation. This issue merits investigation since the results may offer directions for quality improvement initiatives together with practice variation reduction efforts, a gap that our study seeks to address.

To better understand how the joint consideration of practice variation and quality measures (as an interaction) is associated with performance, we conduct post hoc analyses, looking into the granular dimensions of practice variation. Given the distinct objectives of test-ordering (i.e., detecting disease and monitoring its status) and care-delivery (i.e., the provision of healthcare) practices, we disaggregate practice variation into the test-ordering and care-delivery dimensions. We then investigate the interactions between these dimensions and quality measures with respect to operational performance such as patient length-of-stay, test-ordering cost, and care-delivery cost. Delving into granular practice variation measures helps to identify actionable plans for improvement, which is understudied in the healthcare industry, and will provide insights for designing policy incentives.

1.2. Key Findings and Contributions

We analyze six years of a comprehensive dataset from 387 hospitals in NY and FL states, focusing on Medicare patients with three conditions: acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN). To examine hypotheses, we use a dynamic panel model with System Generalized Method of Moments (GMM) estimation (Arellano and Bover 1995, Blundell and Bond 1998) to account for the dynamic endogeneity of practice variation and operational performance. Using several new metrics to measure practice variation, we observe a broad practice variation spectrum across hospitals even when they have the same level of CMI. We first find that higher overall hospital practice variation relates to significantly longer patient length-of-stay (LOS), and higher care-delivery cost per capita. These findings are even stronger if the hospital provides service having high experiential quality, as we proposed. This is because such hospitals pay attention to patient experience under their systems with high practice variation. For this reason, the hospitals may try to mitigate potential negative consequences by devoting more time and resources to the customized experiences.

Our results from post hoc analyses indicate that the dimension of care-delivery practice variation could be the first target for reduction, as it directly relates to the greater patient LOS and all the cost metrics considered without significant interactions with quality measures. Interestingly, the dimension of test-ordering practice variation substantially interacts with quality measures in relation to the total cost and care-delivery cost per capita. This finding suggests that hospital providers might mitigate the trade-off between quality and cost by investing efforts that can help reduce test-ordering practice variation when they concurrently aim to improve the two quality measures. Taken together, these findings imply that careful attention to the two dimensions of practice variation and the nuanced joint relationship with quality measures may provide room for improvement in practice, ultimately reducing waste in the healthcare industry. Our measure of practice variation also contributes since it enables researchers and managers to rigorously measure and visualize the status of hospitals' practice variation and its link to the hospital operational performance.

The remainder of the article is organized as follows. Section 2 provides background and develops hypotheses. Section 3 presents data, variables of interest, research methods, and econometric models, followed by estimation results in section 4. Section 5 provides a broader discussion of our findings and implications. Finally, section 6 concludes and provides directions for future research.

2. Background and Hypotheses Development

In this section, we first review the literature on clinical practice variation as well as process and experiential quality. We then develop hypotheses pertaining to practice variation in hospitals.

2.1. Clinical Practice Variation

Due to persistent healthcare variation (Clancy and Cronin 2005, Ham 1988), practitioners and researchers have tried to identify where the variation originates (Miller et al. 2011), to disentangle warranted variation from unwarranted variation (Wennberg 2002), and to investigate how unwarranted variation can be addressed (Appleby et al. 2011). Wennberg et al. (2002) divide unwarranted variation into three categories: effective care, preference-sensitive care, and supply-sensitive care. Effective care refers to care delivery services of proven effectiveness; thus variation will reflect a failure to deliver needed care. In this category, virtually every patient who is eligible for treatment should be treated. Preference-sensitive care involves care decisions based on patient preferences and values, where at least two valid alternative treatment strategies are available (e.g., surgery vs. alternative medicine without surgery). Supply-sensitive care refers to the frequency of resource usage (e.g., laboratory and radiology tests, physician visits, referrals to specialists, hospitalizations, and stays in intensive care units) that relies on the judgments of a care provider, in the absence of clinical theories of benefit governing the relative frequency of use. These three care categories reflect the varying decision-making processes under differing clinical theory, medical evidence, patient preferences, and supply of resources (Sipkoff 2003).

Measuring variation for the three separate care categories, as they are defined above, is challenging. For example, since every patient has different types and levels of conditions, the proper amount and frequency of needed care become subjective decisions made by a physician and other stakeholders in coordination with the patients, who have varying preferences. As such, without differentiating those three categories, we aim to capture the level of overall practice variation within a patient care episode cohort. Because every hospital has its own "chargemaster" and captures charges for all services and items provided (Ferenc 2013, Melnick and Fonkych 2008), we use detailed medical charge information to construct a measure of practice variation. Our study contributes by constructing a precise measure of practice variation applicable to both hospital and condition (i.e., care episode) level. In contrast to prior work that examines hospital-level geographic variations, our work tracking the extent of practice variation inside a hospital may facilitate the development of incentives that encourage actions to deal with practice variation.

2.2. Process Quality and Experiential Quality

To avoid preventable healthcare errors and to evaluate healthcare quality, the Hospital Inpatient Quality Reporting Program, launched in 2003, mandates the collection and disclosure of process quality measures, for example, measuring the percentage of patients who receive treatments known to lead to the best results (CMS 2014). Hospitals are motivated to report their process quality data to receive financial incentives (CMS 2010). Prior operations management literature sheds light on the impacts of higher process quality under various research settings. On the one hand, Nair et al. (2013) find that higher process quality and clinical flexibility reduce patient LOS in a cardiology unit. Andritsos and Tang (2014) subdivide process quality into clinical and administrative dimensions, and find that higher clinical process quality is associated with lower patient LOS. On the other hand, medical experts point out that documentation and monitoring of process quality are resource-intensive tasks (Fonarow and Peterson 2009). Improving process quality can incur substantial costs for hospitals (Senot et al. 2016). Collectively, prior literature identifies inconsistent associations between process quality and resource usage, for which LOS and cost efficiency are often used as proxy measures.

While clinical processes supported by strong evidence for efficacy are likely to result in better health outcomes (i.e., "Health" of the Triple Aim), they may not address the specific requirements of patients. Indeed, patients also can evaluate perceived quality attributes, such as experiential quality (Nair et al. 2013). The Hospital Consumer Assessment of Healthcare Providers and Systems (HCHAPS) survey measures include communication with caregivers and their responsiveness to patients' requests (i.e., "Care" of the Triple Aim). Consideration of experiential quality is important because caregivers (e.g., physicians and nurses) can use communication to learn more about patients, come up with an accurate diagnosis, and select the most appropriate care method depending on the individual patient's characteristics (Elwyn et al. 2000). Increased attention to each patient's characteristics has been shown to associate with patient satisfaction (Rubin et al. 2001), lower readmission rates (Senot et al. 2016), and decreased patient LOS (Nair et al. 2013). On the other hand, being responsive to patients or communicating more frequently with them can require substantial resources such as advanced information systems (Khunlertkit and Carayon 2013), rapid response teams (Kapu et al. 2014), and responsive registered nurses (Smolowitz et al. 2015), leading to a higher cost per patient discharge (Senot et al. 2016).

Taken together, previous studies view process quality and experiential quality as either complementary assets (Nair et al. 2013, Toussaint 2009) or trade-offs (Chandrasekaran et al. 2012) in achieving various operating outcomes or performance metrics. To the best of our knowledge, however, our work is the first study that examines the relationship between a hospital's practice variation and operational performance (i.e., Cost of the Triple Aim) and how the two quality dimensions can affect the strength of the relationship.

2.3. Hypotheses Development

We hypothesize a relationship between practice variation and resource usage (Hypothesis 1). We then motivate how process quality (Hypothesis 2a) and experiential quality (Hypothesis 2b) may influence the relationship.

2.3.1. Direct Effects of Practice Variation on Operational Performance. For reimbursement purposes, hospitals rely on their chargemaster that lists all the billable services and equipment that might be administered to a patient. Hence, patient-level detailed medical charge information captures the clinical pathway that each patient went through. Thus, variations in medical charges provide an opportunity to measure variability in the underlying clinical pathways (Ferenc 2013, Melnick and Fonkych 2008). If considerable practice variation exists in the total claim charges within a hospital for patients having the same medical condition, managers need to examine where the variation originates and how it can be addressed or alleviated (Appleby et al. 2011).

Quality management literature distinguishes between two types of variation that can cause quality problems (Cachon and Terwiesch 2008, Garvin 1988). One type is *common variation*, purely random variations in an output of a process. The other type is *assignable variation*, whose source can be identified and possibly managed. Key objectives of quality management are to ensure that the outputs of a process are consistent (i.e., the process is in control) and to meet the customer's expectations (i.e., the process is capable). If the process is out of control or not capable, quality management teams must identify assignable variation and reduce it. We conceptually link the assignable variation concept to practice variation.

Considering the potential adverse effect of assignable variation on process output, we posit that high practice variation may harm hospital-level care delivery efficiency. Using SPC to monitor and control a process enables it to reach its full potential (Oakland 2007, Wheeler et al. 1992); in that process, managers can minimize waste and produce as much conforming output as possible. Similarly, in the healthcare context, assembling all the necessary medical services for a common care episode (e.g., AMI, HF, or PN) is a recurring managerial decision process. After a patient is admitted to a hospital, the sequence of recurring decisions made by the administrative group, physicians, and nurses will be reflected in the charges for each patient.

Thus, highly variable charges for patients with the same disease and similar illness severity indicate unstable clinical pathways. This instability possibly increases variability in service times during a care episode, which is known to cause waiting times (Cachon and Terwiesch 2008, Gupta et al. 2016). Using a queueing framework, for example, Dai et al. (2016) claim that physicians' decisions about ordering tests influence patients' waiting times, which may drive overall service time or patient length-of-stay. Unstable hospital care processes might also increase the uncertainty of patient outcomes, possibly leading hospitals to devote additional efforts/resources to alleviate negative outcomes that may critically harm the hospital's reputation.

Collectively, variable service times may lead to longer patient LOS, whether due to value-added time when providers are delivering appropriate care or non-value-added time when a patient is staying without receiving care. Similarly, any chance of uncertain outcomes may lead hospitals to incur a greater amount of expenditure for patient treatment, whether due to meaningful efforts when providers are delivering appropriate care or due to wasteful practice when a patient is given excessive care. Lower practice variation may reduce resource usage for patient care by eliminating non-value-added waiting, lessening overtreatment, limiting excessive patient customization, and lowering material supply uncertainty. Hence, we hypothesize a positive relationship between practice variation and resource usage, proxied as patient LOS and cost metrics.

HYPOTHESIS 1. *Higher clinical practice variation is associated with greater resource usage, for which per patient LOS and cost metrics are used as proxy measures.*

2.3.2. Indirect Effects of Practice Variation on Operational Performance: Interaction with Process Quality and Experiential Quality. Process quality and experiential quality measures, which hospitals are required by CMS to self-report, can be seen as exogenous institutional influences (Westert and Groenewegen 1999). More precisely, process quality is driven by hospital capabilities along a particular clinical path (i.e., *did the hospital follow recommended guidelines at specific steps of a given path?*) and experiential quality is driven by hospital competences on responsive and customized care (i.e., *did the hospital satisfy patient needs by providing necessary care and instructions in a timely manner?*) As a whole, the levels of these two quality measures capture the eagerness of physicians and other hospital stakeholders to comply with the CMS goals in a manner to achieve the reimbursement for services provided.

Note that these quality measures are delicately different from practice variation, which is driven by clinical decision making (i.e., *what care pathway should the hospital put a patient on?*). Accordingly, each of the two quality measures may influence the options that physicians choose when providing services to patients, affecting the relationships described in Hypothesis 1. In principle, it is indeed possible that a hospital is not consistent in the clinical pathways it chooses for otherwise similar patients (i.e., high practice variation) but is capable at specific checkpoints along the way to execute the proper actions (i.e., high process quality).

While high process quality indicates a faithful adherence to evidence-based care standards (Donabedian 1988), improving process quality may result in considerable costs for hospitals. The healthcare industry often faces rapidly evolving knowledge of what is called best practice (Bohmer and Lee 2009, Senot et al. 2016), and indeed, the list of process quality metrics is updated regularly (Mitchell 2014). Although timely adjustments to the new standards may lead to better health outcomes, such activities are likely to incur additional time and cost to coordinate systems, enhance training, and provide corrective feedback. Several studies point out that monitoring, documenting, and reporting such quality measures are often resource intensive, unnecessarily time-consuming, and costly tasks for healthcare providers (Casalino et al. 2016, Fonarow and Peterson 2009).

As an illustration of how pursuing process quality may amplify the effect of practice variation on resource usage, consider CMS process quality guidelines. These guidelines dictate that, for example, an eligible pneumonia patient should get a vaccination to protect against pneumococcal infections caused by bacteria. Following the evidence-based guidelines is likely to improve the process quality "mean" level and its stability (like the X-bar chart in SPC) after going through the aforementioned resource intensive steps. Meanwhile, the frequency of vaccinations or even its eligibility may depend on the care provider's course of decision-making or patient preferences. Of note Wennberg et al. (2002) viewed the supply-sensitive care and preference-sensitive care as unwarranted variation, a level of dispersion in practice (like the R chart in SPC). The higher the practice variation, the more inconsistent and divergent are clinical pathways for a patient cohort having an identical medical condition, perhaps implying that the hospital's operations frequently rely on subjective judgments, and thus are somewhat haphazard. A hospital with haphazard operations is likely to end up devoting more unnecessary resources to care delivery, compared to other hospitals with less practice variation, particularly if the haphazard hospital simultaneously pursues higher achievements for the quality initiatives. Collectively, a hospital that has high scores on process quality will devote more resources, which may strengthen the relationship between practice variation and resource usage described in Hypothesis 1.

HYPOTHESIS 2A. *For hospitals with higher process quality* (*PQ*), *the positive relationships between overall practice variation and resource usage* (*i.e., H*1) *become stronger.*

Collecting and reporting experiential quality is also time-consuming and resource intensive (Boulding et al. 2011). Experiential quality is less controllable than process quality due to extensive patient engagement. Relying on organizational learning theory (Benner and Tushman 2003, March 1991), experiential quality is often referred to as patient-specific exploration (Chandrasekaran et al. 2012). Higher experiential quality reflects managerial and caregiver efforts within a hospital to provide more responsive treatment and individualized attention to every patient. Indeed, establishing such close relationships with patients demands significant added effort and time for care providers (Weick and Sutcliffe 2006), along with the burden of quality reporting.

To improve communications with patients and to enhance responsiveness, that is, operationalizing experiential quality, a hospital may hire more nurses and staff or run communication training programs, which is costly for hospitals (Bechel et al. 2000). We expect the importance of time and monetary resources on the labor and training to be even more for a hospital with higher practice variation, similar to Hypothesis 2a. That is, hospitals that pay attention to providing better patient experience may be likely to devote more time/effort/resources to their patient treatment services. In less efficient systems, for example, hospitals with higher practice variation, such efforts may lead to even greater resource usage to achieve a similar patient satisfaction goal. Thus, we posit that the higher level of responsive care may amplify the positive link between practice variation and resource usage described in Hypothesis 1.

HYPOTHESIS 2B. *For hospitals with higher experiential quality* (*EQ*), *the positive relationships between overall practice variation and resource usage* (*i.e., H*1) *become stronger.*

3. Data, Variables, and Model Development

We next describe data sources and discuss how we construct variables. We then develop econometric models to examine our hypotheses.

3.1. Description of Dataset

We use a comprehensive dataset built by merging four databases. First, we use *State Inpatient Discharges* data from the *Healthcare Cost and Utilization Project (HCUP)* from New York (NY) and Florida (FL) states to identify patient-level information. NY and FL states are the second- and fourth-largest healthcare markets in the United States, respectively, in terms of healthcare expenditures by state of residence (CMS 2011).[4] Both states provide information to track readmitted patients. We use six years of data, from 2008 to 2013, the period for which the two quality metrics are fully available. Each yearly HCUP data file contains the domain of the inpatient discharge record, such as patient demographics, comorbidities, diagnoses, procedures, LOS, physician identifiers, payer, and claim charge information. We extract data records that relate to Medicare patients with conditions AMI, HF, and PN, since the data for the two quality measures are Medicare claims. Using the HCUP data, we construct measures of practice variation (i.e., our main independent variable), LOS, and cost components (i.e., dependent variables).

Second, we use *CMS Timely and Effective Care* data maintained by CMS to obtain process quality metrics. The measures apply to Medicare patients. From among the small list of conditions having process quality measures, we focus on AMI, HF, and PN. We construct a hospital-level composite score of process quality that we discuss later. To incentivize hospitals to participate in quality data collection processes, HHS withholds 0.4% of Medicare fees from the hospitals that choose not to participate (Jha 2006). Thus, we expect these data to be reasonably comprehensive. Third, we use *CMS HCAHPS* survey data to construct experiential quality measures. Fourth, we merge these databases with the *Historical Inpatient Impact File for Acute Inpatient Prospective Payment System*, which is maintained by CMS on an annual basis. We obtain the hospitals' structural characteristics including bed size, CMI, ownership type, corporate goals, location, and teaching intensity, each of which are used as control variables.

In our analysis, we use the merged dataset for 387 healthcare providers from NY and FL. Hospital structural characteristics are summarized in Table 1.[5] Figure 1 shows the process of reducing patient-level HCUP SID data to the dataset pertaining to the 387 hospitals. Our final 1,094,111 patient-level records include patients with three medical conditions (i.e., AMI, HF, and PN) who are Medicare beneficiaries and are not within the top or bottom 1% outliers for each state, year, and MS-DRG code, in terms of total medical charges.[6] Since process quality and experiential quality measures are not reported for medical conditions having <10 patients in a year within a hospital, the actual number of providers in our analyses is slightly <387 per year.

3.2. Variables

The variables used in this study are summarized in Appendix B.

3.2.1. Hospital Operational Performance.

We operationalize the main dependent variables (i.e., *Total LOS* and *Total Cost*) by tracking patients' initial visits and revisits to hospitals within 30 days. *Total LOS* and *Total Cost* measures are often used to represent the level of resource usage. The total LOS (total cost) comprises (i) initial LOS (initial cost) during a patient's initial admission and treatment period and (ii) readmission LOS (readmission cost) of any unplanned readmissions within a 30-day post-discharge period.[7] Via this approach, we obtain more accurate estimates of resource usage during an entire care episode instead of only for an individual discharge (Andritsos and Tang 2014). The 30-day post-discharge window is commonly used in practice since hospitals consider the 30-day window as clinically meaningful and a long enough period for collaborations with care communities to reduce readmissions (Drye et al. 2012).

Cost measures are estimated from patient-level charge data in HCUP SID by following the approach suggested in Chen et al. (2010) and also promoted by the Agency for Healthcare Research and Quality. We first properly apply a national consumer price index for hospital services to the charge data to convert them into 2013 U.S. dollars. We then multiply the inflation-adjusted charges by the CMS annually announced cost-to-charge ratio to estimate each

Figure 1 Process of the Patient Sample Selection

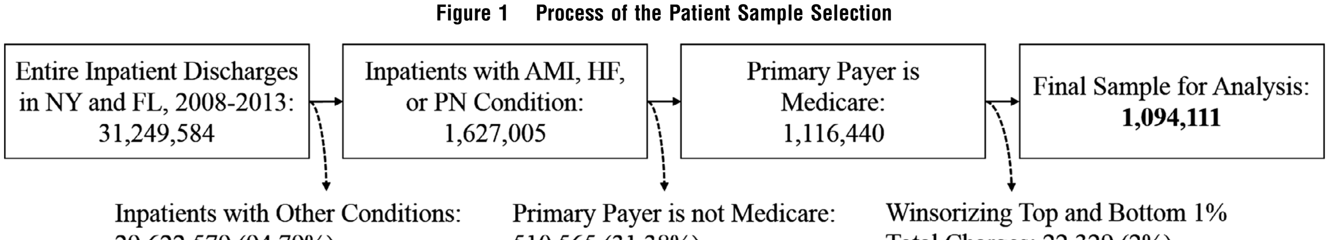


Table 1 Hospital Structural Characteristics

	Frequency	Percentage		Frequency	Percentage
Teaching status			*State*		
Non-teaching	217	56.07	FL	188	48.58
Teaching	170	43.93	NY	199	51.42
Ownership and corporate goals			*Geographic location*		
Governmental	56	14.47	Rural	52	13.44
For-profit	83	21.45	Urban	307	79.33
Non-profit	248	64.08	Missing	28	7.24

hospital's inpatient operating costs. We also obtain *test-ordering cost* by summing charges related to laboratory and radiology tests and multiplying the cost-to-charge ratio. Similarly, we sum charges related to medical and surgical supply, cardiology, respiratory, intensive care, coronary care, pharmacy care services, and room charge and apply the cost-to-charge ratio to obtain *care-delivery cost*.[8]

Further, we construct a risk-adjusted hospital-level mean value of each dependent variable. In other words, we adjust for the warranted variation, to compare hospitals in a fair manner (Andritsos and Tang 2014, CMS 2015). This process removes the influence of patient characteristics that possibly affect LOS and cost metrics for reasons not related to practice variation or care quality of a focal hospital. Specifically, we control for patient demographics (i.e., age, gender, and race), admission-type indicators (i.e., emergency, urgent, and elective), and 29 comorbidity indicators, consistent with recommendations from CMS, National Quality Forum, and the American Heart Association (Horwitz et al. 2011). This process requires estimation of the following model:

$$y_{ijkt} = \alpha + \beta \mathbf{P_k} + \gamma \mathbf{A_k} + \delta \mathbf{C_k} + \epsilon_{ijkt}, \quad (1)$$

where y_{ijkt} is either $ln(Total\ LOS)_{ijkt}$ or $ln(Total\ Cost)_{ijkt}$ for a patient k with condition j who visited hospital i in year t. P_k, A_k, and C_k are vectors of patient k's demographic factors, admission-type indicators, and comorbidity indicators, respectively.[9] Then, we calculate the predicted dependent variable, $\hat{y}_{ikt}$, for each i, k, and t using the estimated model of Equation (1). We then obtain hospital i's mean predicted dependent variable, that is, $\hat{y}_{it}$, by taking the average across patients in a focal hospital. Similarly, we also calculate hospital i's mean observed dependent variable, that is, y^o_{it}, by averaging the actual dependent variables of patients in a hospital. Lastly, we obtain the hospital-level risk-adjusted dependent variable as $y_{it} = \frac{y^o_{it}}{\hat{y}_{it}} \cdot \bar{y}$ where $\bar{y}$ is the mean of the actual dependent variable across all patients and all years used to estimate Equation (1).

3.2.2. Practice Variation. To measure the degree of practice variation in a hospital, we construct a metric called weighted average coefficient of variation (WACV) for each year and care episode (e.g., AMI, HF, and PN) treatment within a hospital.[10] Our main purpose of this measure is to track practice variation for a clinically coherent set of patients, which enables us to relate a hospital's observed outcomes for each condition to the resource demands and associated costs experienced by the hospital for that condition.

When any Medicare beneficiary is discharged from a hospital, a single MS-DRG[11] code is assigned to that patient, which is determined based on multi-dimensional information including principal diagnosis, secondary diagnoses, surgical procedures, age, gender, and discharge status of the patient. For each medical condition, in general, there are three MS-DRG codes to differentiate patients in terms of illness severity, prognosis, and treatment difficulty (CMS 2016). For example, an AMI patient can be assigned to MS-DRG 280 (i.e., with Major Comorbidities, wMCC), 281 (i.e., with Comorbidities, wCC), or 282 (i.e., without Comorbidities, woCC). As such, the patients assigned to the same MS-DRG code are expected to receive similar therapeutic and bed services used in the management of a particular disease, and hence become a homogeneous group in terms of payments. Therefore, if the amounts of total charges vary substantially for the patients within such a coded group, then we conclude that the degree of practice variation of the group is high.

To generate WACV, we calculate the coefficient of variation (CV) for patients within the same MS-DRG code, and then obtain a weighted-average value at the hospital-level, with the weight as the number of patients in each code (see Figure A1 in Appendix D). Let t denote year (i.e., $t \in T = \{2008,\ldots,2013\}$), i a hospital in the integrated dataset, j an element of the set of conditions (i.e., $j \in J = \{AMI, HF, PN\}$), and s a MS-DRG code that reflects illness severity (i.e., $s \in S_j = \{wMCC, wCC, woCC\}$). The CV is the ratio of the sample standard deviation sd_{ijst} to the sample mean $\bar{x}_{ijst}$ of total medical charges for each i, j, s, and t (i.e., $CV_{ijst} = sd_{ijst}/\bar{x}_{ijst}$ where $sd_{ijst} = \frac{1}{n_{ijst}-1} \sum_k^{n_{ijst}} (x^k_{ijst} - \bar{x}_{ijst})^2$ where k indicates patient.).[12]

The CV is useful since it is independent of the unit in which the measurement is taken, whereas the usual standard deviation measure must always be understood in the context of the mean of the data. For this reason, to compare multiple datasets (e.g., hospitals) having far different means or different units, CV is recommended. In the context of healthcare spending, various adjustment factors, such as wage index and CMI, are typically included in empirical models to rule out possible differences across regions and locations. When we derive practice variation via CV, which is a dimensionless number, the value itself is robust for comparisons between and within hospitals, even if we do not perform the wage index or CMI adjustments. WACV for hospital i in year t is computed as:

$$WACV_{it} = \frac{\sum_{j \in J} \sum_{s \in S_j} N_{ijst} \cdot CV_{ijst}}{\sum_{j \in J} \sum_{s \in S_j} N_{ijst}}, \quad (2)$$

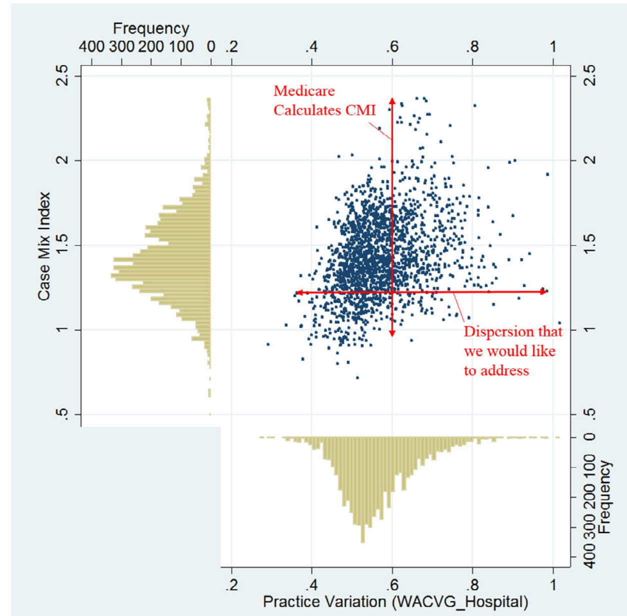
where N_{ijst} is the number of patients and CV_{tijs} is defined as $sd_{tijs}/\bar{x}_{tijs}$ for each i,j,s, and t.

The concept of a weighted average of CV across illness severities is adopted to capture variations after controlling for the medical condition, illness severity, and treatment complexity as much as possible.[13] Table A4 in Appendix E contains the number of patient cases, summary statistics for age, gender, comorbidities, diagnoses, procedures, LOS, and total charges across MS-DRG codes for each condition. From this table, we observe that the total charges might depend on the level of illness severity and treatment complexity. That is, the higher the severity level and complexity, the higher the total charges, because hospitals provide more care services to those patients. Hence, if we derive CV for each entire health condition (e.g., one group made up of all patients with heart failure, regardless of illness severity), then the magnitude of CV can be highly affected by the distribution of patient cases. In essence, both warranted and unwarranted variation in claim charges would be captured by such a metric, which makes it difficult to interpret the magnitude of CV. Hence, we use WACV instead, based on MS-DRG codes as the boundary between patient groups. We claim that higher WACV values indicate higher practice variation. To further rule out other sources of warranted variation (e.g., age, gender, and comorbidities as listed in Table A4, and admission type and race as listed in Table A5 in Appendix E) that are not directly related to the "level of practice," we go through a risk-adjustment process for the charge measures, as previously discussed in section 3.2.1.

Compared to CMI, the WACV metric contributes by enabling appropriate analysis of within-hospital procedural variability after risk-adjusting for patient diversity. Indeed, as Figure 2 shows, the level of practice variation can vary significantly (e.g., [0.36, 1.00]) even when hospitals have the same level of CMI (e.g., 1.25). The histogram of practice variation in Figure 2 is roughly bell-shaped. We note that the hospitals located toward the left side of the histogram (e.g., <0.5) have relatively lower practice variation compared to those located toward the right side.

We also carefully rule out possible limitations of a measure based on CV. If the sample size for each MS-DRG code is too small, then the reliability of WACV may be problematic (Kelley 2007). Therefore, we include hospital-level observations only if the number of patient discharges per condition per fiscal year is greater than or equal to 25. This approach is in line with CMS recommendations (Nair et al. 2013). For robustness purposes, we also test the same regression models by including all of the observations, which consistently shows the same results. We also calculated WACV using an alternative boundary between patient groups (i.e., principal diagnosis within each medical condition, instead of MS-DRG code), and again obtain consistent results. We discuss in detail these robustness checks in section 4.4.

Figure 2 Practice Variation vs. Case Mix Index [Color figure can be viewed at wileyonlinelibrary.com]



3.2.3. Process Quality and Experiential Quality. We follow an approach suggested by CMS to derive the composite process quality (PQ) score from individual measures via a weighted average approach. The size of the eligible patient population becomes a weight for each measure. Prior literature also uses this approach (e.g., Chandrasekaran et al. 2012). Specifically, we use 12 quality measure items for AMI, four items for HF, and seven items for PN to generate composite PQ for each hospital or for each condition, where the list of items slightly varies from year to year. The descriptions and summary statistics for the measure items are provided in Appendix F. We interpret the resulting value as a compliance score that reflects the extent to which medical guidelines are followed. Let m denote a process quality measure item and M_j denote the set of measures for condition j. The number of patients differs across measures since there may be some reasons that some patients do not need to receive a certain treatment. We derive the weighted average process quality score per hospital for each year as follows:[14]

$$PQ^o_{it} = \frac{\sum_{j\in J}\sum_{m\in M_j} N_{ijmt}\cdot q_{ijmt}}{\sum_{j\in J}\sum_{m\in M_j} N_{ijmt}}, \tag{3}$$

where q_{ijmt} is the associated process quality score and N_{ijmt} is the number of patients for each i, j, m, and t.[15] The distribution of each q_{ijmt} in the CMS database and the resulting PQ^o_{it} measures are left-skewed. Hence, we perform a logit transformation of the odds ratio of PQ^o_{it} to make this distribution less skewed (Cohen et al. 2003):

$$PQ_{it} = ln\left(\frac{PQ^o_{it}}{1 - PQ^o_{it}}\right). \quad (4)$$

In a similar vein, we derive the composite experiential quality (EQ) score (e.g., Nair et al. 2013). The experiential quality score is based on six items in the HCAHPS survey (see Appendix F). Within the survey, the responses for the first five items (i.e., Comp1-Comp5) are presented as "Never/Sometimes," "Usually," or "Always," and for the sixth item (i.e., Comp6) the response is reported as "Yes" or "No." To handle this difference in the data structure for the first five items, we use the percentage of patients who answered "Always" as the measure of each individual item (Senot et al. 2016). For the sixth item, we designate the percentage of patients who answered the question with "Yes" as a final score. Lastly, we compute the hospital-level overall score as the average of the percentage scores for the six items (i.e., EQ^o_{it}). Similar to the process quality score, the experiential quality score for each hospital is then calculated by taking the logit transformation of the percentage. Thus, EQ_{it} for a hospital i and year t is given by:

$$EQ_{it} = ln\left(\frac{EQ^o_{it}}{1 - EQ^o_{it}}\right). \quad (5)$$

To obtain easy-to-interpret results, we mean-center the practice variation measure and quality measures before computing their interaction terms (Hamilton 2012).

3.2.4. Interaction of Practice Variation with Quality. The interaction of two separate measures are often studied in prior literature to understand the effects of combined strategy (e.g., Cao et al. 2009, He and Wong 2004) or combined quality (e.g., Senot et al. 2016) that jointly captures how the simultaneous pursuit of both measures adds to or detracts from each other's value. Similarly, we define the interaction between practice variation and either process quality or experiential quality to reflect a joint measure of a hospital's status on both practice variation and the quality initiative.

3.2.5. Controls. We control for several hospital factors that are related to potential sources of heterogeneity in performance across hospitals. Hospital size is measured according to its total number of beds (*Bed Size*). We take the logarithm to account for heavy tails in this distribution. *CMI* captures the average DRG weight for different DRGs per hospital (Ding 2014). CMS derives CMI by calculating the ratio between the total DRG weights associated with Medicare discharges and the total discharges. Hospital *Teaching Intensity* is defined as the residents-to-bed ratio (Theokary and Ren 2011). *Wage Index* reflects the relative hospital wage level in the geographic area of the hospital compared to the national average hospital wage level (Shwartz et al. 2011). We also control for the CMS operating outlier adjustment factor (*Outlier Adjustment*), which reflects the extent of uncommonly costly patients treated by the hospital, and the CMS operating disproportionate share hospital payment adjustment factor (*OPDSH Adjustment*) for hospitals that serve a notably disproportionate number of low-income patients and receive supplemental payments from CMS to cover the costs of providing care to uninsured patients (Senot et al. 2016).

The control variables so far are all time-varying, which is appropriate for the dynamic panel model that we discuss later. We also use time-invariant controls for static OLS estimates. Each hospital is classified into three types: government-sponsored, private non-profit, and private for-profit. Setting for-profit hospitals as the base group, we use two binary variables, one for private non-profit hospitals (*Nonprofit*) and another for government-sponsored hospitals (*Governmental*). Hospital location (*Urban*) is classified as either urban (1) or rural (0). We include year dummies to control for unobservable factors that cause population change in hospital operational performance across time.

3.3. Methodologies and Econometric Models

We discuss the dynamic relationship between practice variation, quality, and operational performance, and develop econometric models to examine our hypotheses.

3.3.1. A Dynamic Model of Operational Performance. Endogeneity is pervasive across many aspects of healthcare operations. The specific effect of endogeneity may arise from the dynamic relationship between current hospital operations and a hospital's history. We examine how practice variation as well as quality measures relate to operational performance. The level of the two quality measures is dynamically endogenous with respect to operational performance because manager talent can affect quality measures that are closely linked with financial incentives. We expect a similar relationship between practice variation and operational performance.

As section 2 implies, practice variation and the two quality measures are choice-type variables in a broad sense, arising through a process of bargaining between decision makers inside a hospital (e.g., board members, physicians, and administrative staff). Although governmental policies might be the most critical driver, this process is also influenced by past performance, manager talent, and decision makers' beliefs about the cost and benefits of choosing reasonable clinical pathways for patients, leading to various levels of practice variation and quality across hospitals. Therefore, if practice variation and quality measures are dynamic, and hospital i (given its performance at time $t-1$ or earlier) chooses certain levels of practice variation and quality score $\mathbf{X_{it}}$ to achieve a particular level of expected operational performance at time t, then the dynamic model is:

$$\mathbf{X_{it}} = f(y_{it-1}, \ldots, y_{it-p}, \mathbf{Z_{it}}, \eta_i), \tag{6}$$

where $\mathbf{X_{it}}$ is a vector of endogenous predictors (i.e., practice variation, quality measures, and their interaction terms) that we call *practice-quality status*, $\mathbf{Z_{it}}$ is a vector of exogenous predictors (i.e., controls and time dummies), y_{i*} represents operational performance, and η_i is an unobserved hospital effect.

Equation (6) suggests that estimating the effect of *practice-quality status* on operational performance, conditional on hospital heterogeneity, requires estimating the following model:

$$y_{it} = \alpha + \sum_p \lambda_p \cdot y_{it-p} + \beta \cdot \mathbf{X_{it}} + \gamma \cdot \mathbf{Z_{it}} + \eta_i + \epsilon_{it} \tag{7}$$

where ϵ_{it} is an idiosyncratic error term and β is the coefficients of interest. To estimate Equation (7), we assume that current shocks are independent of historical realizations of performance and *practice-quality status*. In other words, past and current realizations of practice variation and quality scores are allowed to influence current performance. This assumption is not too strong and leaves open the possibility that hospitals may strategically choose their level of practice variation and target quality to affect current or future performance. If the level of *practice-quality status* that we observe today is the one that trades off the anticipated benefits and costs of doing so, then the unanticipated component of performance, many years in the future, will not be related to the *practice-quality status* that is realized today. Intuitively, we can write this in orthogonality form as $E(\epsilon_{it}|y_{it-s}, \mathbf{X_{it-s}}) = 0, \forall s > p$. If Equation (6) represents the true model for performance, that is, if we correctly identified every endogenous time-varying variable that affects performance, then ϵ_{it} is an expectational error and the orthogonality assumption is valid (Hansen and Singleton 1982). Equation (7) is simply a reduced-form model and thus the reduced-form error, ϵ_{it}, is at best a proxy for the pure expectational error (Wintoki et al. 2012).

Still, there are several challenges in empirically estimating the fixed-effect model (Pang et al. 2016). As the hospital operational performance is likely to be affected by hospital-specific unobserved heterogeneity (η_i) that may be correlated with explanatory variables, hospital time-invariant fixed-effects need to be accounted for. However, fixed-effects estimation does not completely control for the correlation between η_i and the lagged dependent variable[16] (Roodman 2006). Hence, we estimate a dynamic panel data model via System GMM estimation.

3.3.2. Estimation Strategy: Dynamic Panel System GMM. Under the assumption that unobserved heterogeneity is time-invariant, we obtain consistent and unbiased estimates of the relationship between *practice-quality status* and operational performance via a dynamic panel System GMM estimator (Arellano and Bover 1995, Blundell and Bond 1998). This estimator exploits the dynamic relationships inherent in our independent variables. The dynamic modeling approach has been widely used in areas such as economics and finance, where the structure of the problem contains a dynamic relationship between independent and dependent variables (e.g., Blundell and Bond 1998, Bond and Meghir 1994). Recently, operations management, information systems, and marketing fields also began examining salient problems using this approach (e.g., Bhargava and Mishra 2014, Narayan and Kadiyali 2015, Rego et al. 2013).

The estimation comprises two steps. First, we write the first-differenced form of Equation (7):

$$\Delta y_{it} = \alpha + \lambda_p \cdot \sum_p \Delta y_{i,t-p} + \beta \cdot \Delta\mathbf{X_{it}} + \gamma \cdot \Delta\mathbf{Z_{it}} + \Delta\epsilon_{it}, \tag{8}$$

where $p > 0$.

The first-difference effectively removes any bias that may arise from unobserved time-invariant heterogeneity. After first-differencing, we estimate Equation (8) via GMM using lagged values of performance, practice variation, quality scores, and other hospital-specific variables as instruments for current changes in these variables. An essential aspect of the dynamic panel estimator is its use of hospital-level history as instrument variables for our explanatory variables. Thus, in estimating Equation (7) or the first-difference as in Equation (8), our instruments will be drawn from the set of lagged dependent or independent variables (i.e., y_{it-k}, $\mathbf{X_{it-k}}$, $\mathbf{Z_{it-k}}$, where $k > p$.). For these instruments to be valid, they should satisfy two criteria. First, they must provide a source of variation for current *practice-quality status* (i.e., $\mathbf{X_t} = f(y_{t-k}, \mathbf{X_{t-k}}, \mathbf{Z_{t-k}})$). We show later that practice variation and quality scores are correlated to lagged operational performance and lagged values of other control variables (see Table A7 in Appendix G).

Second, the historical values of practice variation and quality scores must provide an exogenous source of variation that reflects current *practice-quality status*. As such, the lagged variables should not be correlated with the error term in Equation (7). Any information from the prior p periods is reflected in the current expected performance. Thus, p lags of performance are enough to address the impact of the hospital's past on its present. The hospital's history beyond period $t - p$ should be exogenous with respect to any shocks to performance in the current and future periods. Under the exogeneity assumption, the following orthogonality condition is valid (Wintoki et al. 2012):

$$E(\mathbf{X_{it-s}}\epsilon_{it}) = E(\mathbf{Z_{it-s}}\epsilon_{it}) = E(y_{it-s}\epsilon_{it}) = 0, \quad \forall s > p. \tag{9}$$

The number of lags included for each dependent variable in our analysis is revealed to be one according to how many lags are statistically significant in the corresponding regression.[17]

We then estimate the level and difference equations simultaneously, as Arellano and Bover (1995) and Blundell and Bond (1998) show that the GMM estimator can be improved compared to solely estimating a first-difference model. We use the first-differenced variables as instruments for the level equations in a system of equations as below (Roodman 2006):

$$\begin{bmatrix} y_{it} \\ \Delta y_{it} \end{bmatrix} = \alpha + \lambda \begin{bmatrix} \sum_p y_{it-p} \\ \sum_p \Delta y_{it-p} \end{bmatrix} + \beta \begin{bmatrix} \mathbf{X_{it}} \\ \Delta\mathbf{X_{it}} \end{bmatrix} + \gamma \begin{bmatrix} \mathbf{Z_{it}} \\ \Delta\mathbf{Z_{it}} \end{bmatrix} + \epsilon_{it}, \tag{10}$$

where $y_{it} = ln(Average\ Total\ LOS)$ or $ln(Average\ Total\ Cost)$ for a hospital i in year t, $\mathbf{X_{it}} =$ [WACV, PQ, EQ, WACV $*$ PQ, WACV $*$ EQ], and $\mathbf{Z_{it}} =$ a set of time-varying controls. Variable y_{it} and Δy_{it} denote the level and year-to-year change (from $t-1$ to t) in operational performance in hospital i. The level and change in practice variation, quality scores, and their interactions are captured by $\mathbf{X_{it}}$ and $\Delta\mathbf{X_{it}}$.

Note, however, that the level equations still have unobserved heterogeneity, η_i. As in Wintoki et al. (2012) and Kuhnen and Niessen (2012), we assume that the correlation between *practice-quality status* and control variables is constant over time. Relying on this assumption, we have another set of orthogonality conditions:

$$\begin{aligned} E[\Delta\mathbf{X_{it-s}}(\eta_i + \epsilon_{it})] &= E[\Delta\mathbf{Z_{it-s}}(\eta_i + \epsilon_{it})] \\ &= E[\Delta y_{it-s}(\eta_i + \epsilon_{it})] = 0, \\ \forall s > p. \end{aligned} \tag{11}$$

We check the validity of instruments $\mathbf{Z_{it}}$ with serial correlation tests and the Hansen test of over-identification (Arellano and Bond 1991) and show the test statistics in the results tables (i.e., Tables 3 and 4). According to the serial correlation tests, the assumptions of our specifications are valid, that is, the residuals in first-difference ($AR(1)$) are significantly correlated, but there is no serial correlation in second-differences ($AR(2)$). In addition, the Hansen test with insignificant p-values in all specifications indicates that the null hypothesis that our instruments are valid is not rejected. Lastly, the difference-in-Hansen test tells us that the subset of instruments used in the level equations is exogenous for all model specifications.

4. Model Estimation Results

Table 2 reports summary statistics and correlations of key variables.

4.1. Model-Free Evidence

Before we present the results of our proposed model, we provide model-free evidence of the practice variation effect on hospital operational performance. As part of these analyses, we present plots of practice variation and patient total LOS/total cost. For brevity, we compare the average per patient total LOS and total cost across two scenarios: hospitals with (i) practice variation less than the median during a year, and (ii) practice variation more than the median during a year. In Figure 3, we show that hospitals with a relatively higher level of practice variation exhibit longer patient LOS and higher total cost. We further shed light on the practice variation effect on test-ordering cost and care-delivery cost. In Figure 4, hospitals with relatively higher practice variation tend to spend less on test-ordering activities while spending more in care-delivery activities.

4.2. Estimation Results

Table 3 summarizes the estimates obtained for H1 together with H2a and H2b. With total LOS as a dependent variable, we first run a model with Process Quality and Experiential Quality (M1), and then include Practice Variation (M2). Lastly, we include interaction terms (M3). We similarly examine the model with the same explanatory variables but with total cost per capita as a dependent variable (M4 to M6). We also include the results when the dependent variables are test-ordering cost (M7) and care-delivery cost (M8), the two exclusive components of total cost per capita. All the instrument validity and identification tests

Figure 3 Model-Free Evidence: Main Effect of Practice Variation on Risk-Adjusted Average Length-of-Stay (left) and Total Cost (right) [Color figure can be viewed at wileyonlinelibrary.com]

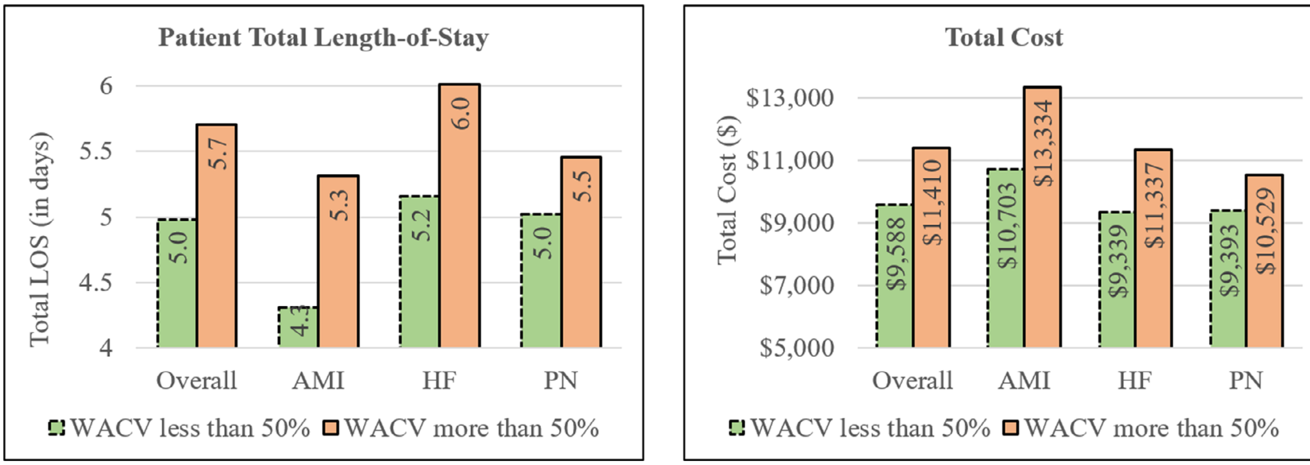
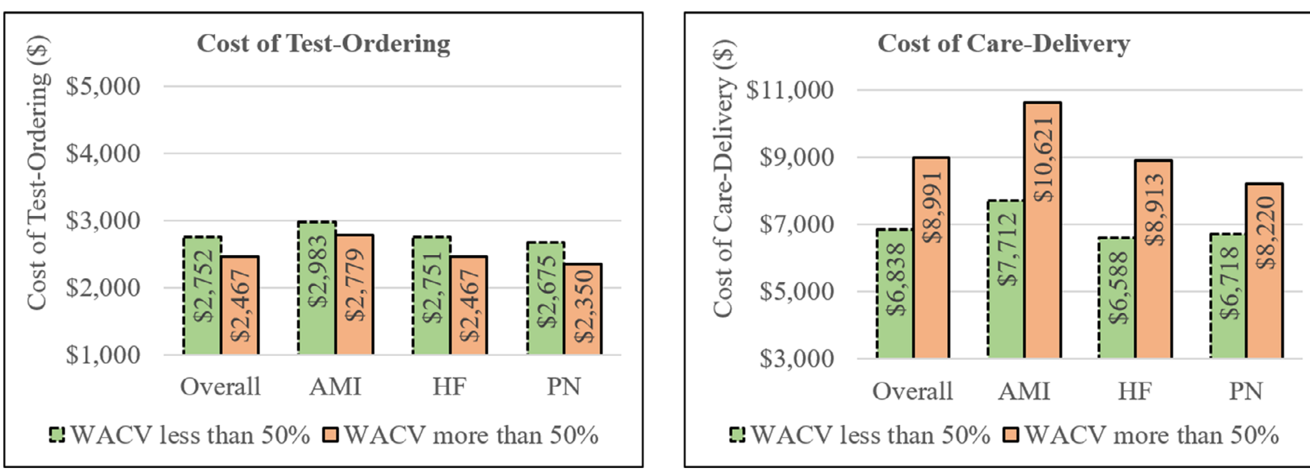


Figure 4 Model-Free Evidence: Main Effect of Practice Variation on Risk-Adjusted Average Cost of Test-Ordering (left) and Cost of Care-Delivery (right) [Color figure can be viewed at wileyonlinelibrary.com]



reported in Table 3 support the use of dynamic panel system GMM to estimate the models.

In the relationship of interest in H1, the main-effect results (M3) in Table 3 show a significant positive association between practice variation (i.e., WACVG) and total LOS ($\beta = 0.635$, $p < 0.01$). This result indicates that patients staying at hospitals with greater practice variation tend to stay longer during their entire episode of care.

Consistent with previous literature (e.g., Nair et al. 2013), patients at hospitals with higher experiential quality tend to stay a shorter period ($\beta = -0.549$, $p < 0.01$). However, we do not observe any significant relationship between process quality and patient total LOS ($\beta = -0.073$, $p > 0.10$) although the direction is aligned with prior findings (e.g., Andritsos and Tang 2014).

H2a and H2b posit that the relationship between practice variation and total LOS/cost is stronger when process quality and experiential quality are high, respectively. M3 in Table 3 shows the results on total LOS. The interaction between practice variation and process quality is not significant ($\beta = 0.130$, $p > 0.10$) with total patient LOS, indicating H2a is not supported. In contrast, the interaction between practice variation and the experiential quality shows a significant positive association ($\beta = 0.920$, $p < 0.05$) with total LOS, providing support to H2b. Thus, the benefit of higher experiential quality in reducing total patient LOS can diminish once we explicitly consider the level of practice variation for a focal hospital.

Similar to the mixed (or insignificant) results of previous research on the relationship between hospital quality measures and total cost (e.g., Nair et al. 2013), none of the quality measures are significant, as shown in M4, M5, and M6 of Table 3. This result is perhaps due to confounding data

Table 2 Descriptive Statistics and Correlation Table of Key Variables

	Variable	n	Mean	S.D.	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	ln(Total LOS)	2045	1.694	0.219	−0.667	2.933	1														
(2)	ln(Total Cost)	2032	9.092	0.424	6.146	10.322	0.452	1													
(3)	ln(Cost of Test-Ordering)	2027	7.74	0.545	2.32	9.232	0.024	0.639	1												
(4)	ln(Cost of Care-Delivery)	2032	8.872	0.443	6.015	10.33	0.518	0.92	0.352	1											
(5)	WACVG	2004	0.566	0.093	0.272	1.017	0.426	0.205	−0.124	0.309	1										
(6)	WACVG_Test	2017	0.602	0.138	0	2.94	0.143	−0.124	−0.433	0.075	0.371	1									
(7)	WaCVG_Care	2017	0.621	0.107	0	1.235	0.388	0.245	0.018	0.281	0.792	0.266	1								
(8)	ProcQual	2356	−0.017	1.239	−6.099	5.223	0.053	0.078	0.193	−0.022	0.116	−0.025	0.155	1							
(9)	ExpeQual	2265	0	0.282	−0.86	1.653	−0.346	−0.143	−0.026	−0.124	−0.263	−0.1	−0.33	−0.092	1						
(10)	ln(Bed Size)	2484	5.148	0.96	1.386	7.564	0.329	0.255	0.192	0.194	0.482	0.141	0.514	0.462	−0.391	1					
(11)	CMI	2329	1.42	0.246	0.488	2.366	0.113	0.193	0.164	0.16	0.317	0.101	0.293	0.433	−0.105	0.681	1				
(12)	Teacing intensity	2329	0.108	0.208	0	0.961	0.304	0.222	−0.19	0.345	0.569	0.313	0.517	0.028	−0.299	0.385	0.289	1			
(13)	Wage Index	2311	1.026	0.174	0.814	1.493	0.45	0.519	−0.006	0.639	0.423	0.212	0.388	0.019	−0.298	0.264	0.12	0.569	1		
(14)	Outlier adj factor	2329	0.04	0.048	0	0.658	0.099	0.272	0.18	0.258	0.316	0.049	0.276	0.12	0.022	0.373	0.408	0.272	0.112	1	
(15)	OPDSH adj factor	2329	0.141	0.173	0	0.86	0.222	0.049	−0.208	0.126	0.353	0.192	0.377	0	−0.324	0.283	−0.026	0.612	0.445	0.113	1

Notes: $p < 0.05$ if $|r| > 0.049$. n is the number of hospital-level observations for six years (2008 to 2013).

aggregation factors inherent in the construction of the explanatory measures and total cost, which possibly cancel out existing effects. We only find a weak association between the level of practice variation and total cost ($\beta = 0.233$, $p < 0.10$). However, by splitting the total cost into test-ordering cost and care-delivery cost, we find interesting results, especially for the case of care-delivery cost. The association between practice variation and care-delivery cost is positive and significant ($\beta = 0.379$, $p < 0.01$). The interaction between practice variation and experiential quality also shows a significant positive association ($\beta = 0.575$, $p < 0.05$) with respect to care-delivery cost. These results motivate us to investigate granular measures of practice variation that we discuss in our post hoc analyses. Among the control variables, the coefficient of *Wage Index* is significantly positive across the estimation models, implying that hospitals located in a geographic area with higher wage level tend to spend more resources compared to hospitals located in a national average wage level area.

The left plot in Figure 5 represents the interaction between overall practice variation and experiential quality with regard to total LOS. This plot reflects the importance of considering the practice variation and experiential quality together in reducing patient LOS. Consider the hospitals with high experiential quality (i.e., top 10th percentile). In this case, a 0.1 decrease in practice variation would correspond on average to 0.86 days decrease in patient LOS. In contrast, for hospitals with relatively low levels of experiential quality (i.e., 90th percentile), a 0.1 decrease in overall practice variation would result on average in only 0.24 days decrease in patient LOS. Similarly, the right plot in Figure 5 shows the interaction between overall practice variation and experiential quality with regard to care-delivery cost. Again, consider the hospitals with high experiential quality. A 0.1 decrease in practice variation would correspond on average to \$466.13 of reductions in care-delivery cost per capita. In contrast, for hospitals with low experiential quality, a 0.1 decrease in practice variation would result on average in only \$146.80 of reductions in care-delivery cost per capita. In sum, we highlight the importance of addressing the relationship between practice variation and operational performance, together with hospitals' experiential quality.

4.3. . Post Hoc: Granular Dimensions of Practice Variation

To further investigate how the joint consideration of practice variation and quality measures (i.e., as interactions) is associated with performance, we delve into granular practice variation measures.

Table 3 Hospital-Level Results of System Generalized Method of Moments (GMM) (WACVG, Dep Var: Total LOS, Total Cost, Test-Ordering Cost, and Care-Delivery Cost)

	Total LOS			Total Cost			Test Cost	Care Cost
Dep Var	(M1)	(M2)	(M3)	(M4)	(M5)	(M6)	(M7)	(M8)
ProcQuality (PQ)	−0.001	−0.001	−0.073	0.004	0.004	−0.015	0.038	−0.003
	(0.005)	(0.003)	(0.047)	(0.010)	(0.008)	(0.080)	(0.072)	(0.069)
ExpQuality (EQ)	−0.019	−0.050*	−0.549***	0.052	0.020	−0.134	0.147	−0.279**
	(0.050)	(0.028)	(0.205)	(0.134)	(0.071)	(0.171)	(0.204)	(0.141)
WACVG		0.553***	0.635***		0.318**	0.233*	0.116	0.379***
		(0.060)	(0.070)		(0.127)	(0.122)	(0.127)	(0.105)
PQ * WACVG			0.130			0.027	−0.034	−0.006
			(0.084)			(0.141)	(0.132)	(0.122)
EQ * WACVG			0.920**			0.291	−0.368	0.575**
			(0.369)			(0.338)	(0.368)	(0.289)
Teaching intensity	0.014	−0.056**	−0.054	0.021	−0.031	−0.010	−0.197**	0.036
	(0.023)	(0.025)	(0.038)	(0.063)	(0.048)	(0.055)	(0.095)	(0.042)
Bed size	0.010	−0.002	−0.001	0.006	−0.002	−0.007	−0.005	−0.018
	(0.007)	(0.006)	(0.008)	(0.022)	(0.016)	(0.015)	(0.017)	(0.013)
Case mix index	0.010	−0.002	−0.017	−0.029	−0.028	−0.012	0.004	0.007
	(0.021)	(0.020)	(0.027)	(0.057)	(0.042)	(0.039)	(0.042)	(0.037)
Wage index	0.064**	0.046*	0.113***	0.450*	0.352***	0.327**	0.099*	0.338***
	(0.031)	(0.027)	(0.034)	(0.232)	(0.120)	(0.139)	(0.050)	(0.104)
OPDSH adj factor	−0.011	−0.022	0.023	−0.248**	−0.208***	−0.175**	−0.164*	−0.155***
	(0.028)	(0.028)	(0.035)	(0.106)	(0.074)	(0.076)	(0.087)	(0.054)
Outlier Adj Factor	−0.060	−0.154*	−0.114	0.388	0.230	0.209	0.377*	0.031
	(0.089)	(0.086)	(0.105)	(0.349)	(0.224)	(0.216)	(0.198)	(0.132)
Dep Var$_{(t-1)}$	0.779***	0.705***	0.551***	0.673***	0.721***	0.739***	0.516***	0.777***
	(0.065)	(0.049)	(0.064)	(0.152)	(0.080)	(0.094)	(0.138)	(0.057)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1527	1527	1527	1527	1527	1527	1527	1527
Hospitals	324	324	324	324	324	324	324	324
Instruments	41	49	65	41	49	65	65	65
AR(1) (*p*-value)	(0.000)	(0.000)	(0.000)	(0.027)	(0.004)	(0.005)	(0.001)	(0.000)
AR(2) (*p*-value)	(0.039)	(0.094)	(0.205)	(0.453)	(0.459)	(0.461)	(0.325)	(0.867)
Hansen test of overid. (*p*-value)	(0.443)	(0.083)	(0.238)	(0.090)	(0.060)	(0.053)	(0.051)	(0.131)
Diff.-in-Hansen test of exogeneity (*p*-value)	(0.524)	(0.125)	(0.286)	(0.640)	(0.304)	(0.229)	(0.145)	(0.480)

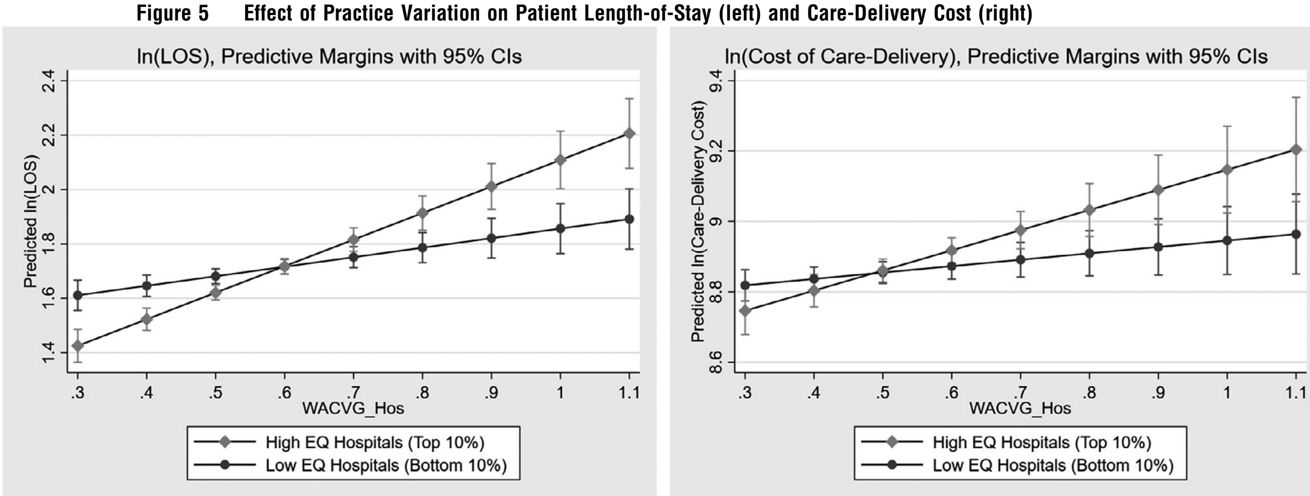
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Notes: The results are based on a system GMM model (Arellano and Bond 1991, Blundell and Bond 1998) estimated as in Equation (10). Standard errors are corrected for heteroskedasticity and are clustered at the hospital level. AR(1) and AR(2) are tests for first-order and second-order serial correlation in the first-differenced residuals, under the null of no serial correlation. The Hansen test of overidentification is under the null that all instruments are valid. The difference-in-Hansen test of exogeneity is under the null that instruments used for the equations in levels are exogenous.

4.3.1. Two Dimensions of Practice Variation: Test-Ordering and Care-Delivery. To this point, we focus on practice variation's impacts on LOS and cost metrics during a patient's care episode. Still, we recognize that each care episode comprises several stages, potentially including admission, diagnosis, treatment, recovery, and discharge. By classifying tasks during the entire care episode into test-ordering practice and care-delivery practice,[18] we may gain a deeper insight about whether and how practice variation originating from different tasks is associated with operational performance.

For each inpatient in our dataset, we have a detailed list of charges pertaining to a hospital revenue center. A few hundred revenue codes can be categorized as listed in Table A8 (in Appendix H). Almost all patients for the AMI, HF, and PN medical conditions receive laboratory and radiology tests. The amount of charges related to the tests accounts for a considerable portion of a patient's total charges. For example, 98.63% and 97.56% of patients with heart failure received at least one laboratory test and one radiology test, respectively, and on average, they account for 27.96% of total charges. Test-ordering practice is important to assign correct patient diagnoses in a timely manner and to monitor a patient's disease during a care episode (Grytten et al. 2016) while care-delivery practice aims to cure the identified disease. Given such distinct objectives of test-ordering and care-delivery practices, we disaggregate practice variation into the test-ordering and care-delivery dimensions and recalculate them following

Figure 5 Effect of Practice Variation on Patient Length-of-Stay (left) and Care-Delivery Cost (right)



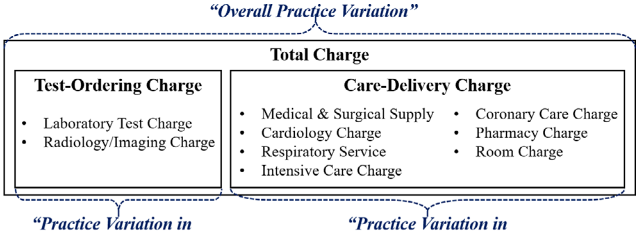
Note: The 10th–90th percentile ranges are displayed for Experiential Quality.

Equation (2). Figure 6 clarifies the boundary of these two granular dimensions of practice variation. Using data from patients with pneumonia in year 2011, Figure 7 illustrates how practice variations in test-ordering (left) and care-delivery (right) can vary across hospitals.

In test-ordering practice, reducing a physician's uncertainty would necessitate ordering more tests. Many medical malpractice lawsuits (about 35.2% of payouts) relate to diagnostic errors (NASEM 2016). A leading cause of such errors is a failure by a medical professional to order a proper set of medical tests. If needed medical tests are not performed, a patient's health condition can be overlooked, which possibly leads to serious patient harm (Berlin 2002, Gandhi et al. 2006). Thus, physicians may tend to take proactive defensive attitudes and make efforts to avoid such errors by ordering a permissive, rather than a restrictive, body of medical tests. Interestingly, however, Zhi et al. (2013) find that both overuse and underuse of tests are equivalently pervasive, implying considerable test-ordering practice variation.

In contrast, during care delivery, no common strategy (e.g., doing more) is known to reduce uncertainty. Instead, only a situational strategy exists: Do as your direct colleagues do (Eddy 1984), which is again subjective, perhaps leading to a significant amount of care-delivery practice variation.

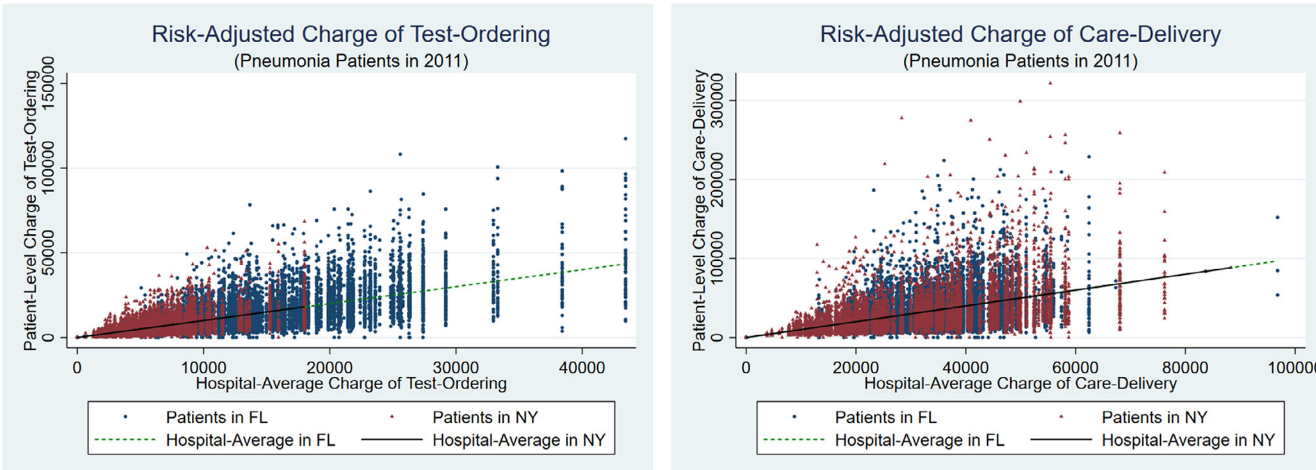
Figure 6 Boundary of Practice Variation Measures [Color figure can be viewed at wileyonlinelibrary.com]



Taken together, we are interested in how the two granular practice variation measures are related to operational performance and how the granular measures interact with the quality initiatives. Similar to our main hypotheses, we posit that hospitals' efforts on achieving high-quality initiatives would incur even higher resource usage if the two granular practice variations are high. Meanwhile, we expect the interaction effects with quality initiatives could be more prominent for high test-ordering practice variation. That is because the higher variation in test-ordering would imply less clear information about the disease, increasing the uncertainty about what to do next. A possible strategy to overcome would be simply trying more alternative efforts.

4.3.2. Post Hoc Analyses and Results. The results of post hoc analyses are listed in Table 4: from M1 to M8, the dependent variables are the same as in Table 3, but the two granular practice variation measures and their interactions with process quality and experiential quality are included accordingly. Note that, in these models, we include test-ordering practice variation (i.e., WACVG_Test) after controlling for that of care-delivery practice variation (i.e., WACVG_Care) to orthogonalize the two variables, because they are correlated with each other. M3 shows evidence that practice variations in both test-ordering ($\beta = 0.165$, $p < 0.01$) and care-delivery ($\beta = 0.395$, $p < 0.01$) are positively associated with Total LOS. Among the four interaction terms, the interaction between test-ordering practice variation and experiential quality is only positively significant ($\beta = 0.457$, $p < 0.05$) with patient

Figure 7 Scatter Plots of Test-Ordering Charges (left) and Care-Delivery Charges (right) [Color figure can be viewed at wileyonlinelibrary.com]



Total LOS.

For the cost metrics, we observe from M6 that higher care-delivery practice variation is associated with higher total cost per capita ($\beta = 0.410$, $p < 0.01$), but its interactions with the two quality measures are insignificant. Interestingly, test-ordering practice variation itself is not significantly associated with total cost ($\beta = -0.216$, $p > 0.10$), but its interactions with both PQ ($\beta = 0.384$, $p < 0.01$) and EQ ($\beta = 0.612$, $p < 0.05$) show positive and significant associations with total cost, supporting hypotheses H2a and H2b to some extent. These results are sharpened if the dependent variable becomes care-delivery cost in M8 ($\beta = 0.633$, $p < 0.01$ for WACV_Care, $\beta = 0.175$, $p < 0.01$ for PQ $*$ WACV_Test, and $\beta = 0.837$, $p < 0.01$ for EQ $*$ WACV_Test). Lastly, as in M7, care-delivery practice variation is positively associated with test-ordering cost ($\beta = 0.269$, $p < 0.01$), while test-ordering practice variation shows an opposite direction ($\beta = -0.461$, $p < 0.05$). Taken together, our findings suggest that the dimension of care-delivery practice variation directly relates to the greater patient LOS and greater cost metrics, but without significant interactions with quality measures. By contrast, the dimension of test-ordering practice variation has combined effects with both quality measures with respect to the total cost and care-delivery cost, while having negative direct effects on the test-ordering cost. Hence, hospitals may struggle with operational performance if they have a considerable amount of care-delivery practice variation or if they pursue high quality measures while having large test-ordering practice variation. We summarize the results of hypotheses tests in Table 5.

The left plot in Figure 8 represents the interaction between test-ordering practice variation and the two quality measures with regard to care-delivery cost. For the hospitals with high (i.e., top 10th percentile) experiential quality (resp. process quality), a 0.1 decrease in test-ordering practice variation would correspond on average to \$364.50 (resp. \$123.48) decrease in per capita care-delivery cost. In contrast, for hospitals with relatively low (i.e., 90th percentile) levels of experiential quality (resp. process quality), a 0.1 decrease in test-ordering practice variation would result on average in \$71.16 (resp. \$60.80) increase in per capita care-delivery cost. These results imply that if the hospital is competent in process quality and experiential quality, putting efforts to alleviate test-ordering practice variation would be beneficial for improving operational performance (or, in other words, reducing avoidable waste). Otherwise, rather comprehensive actions should be taken together on both reducing practice variation and improving quality measures.

4.4. Robustness Checks

Our results remain robust to several checks. First, we construct the measure of practice variation in a different way. In our main analysis, we rely on MS-DRG codes to define patient groups who require relatively homogeneous care. The MS-DRGs are generally assigned to patients based on multi-dimensional information such as their principal diagnosis and additional diagnoses, the principal procedure and additional procedures, gender, and discharge status. Since the DRG code is assigned around the time of discharge, the MS-DRG based practice variation is an *ex-post* type of practice variation (we name this variable WACVG; G stands for general). Alternatively, we consider an *ex-ante* type of practice variation. That is, we can use the principal diagnosis code, which is usually determined at an early stage of the diagnostic phase, to define a group of patients when calculating practice variation (we name this variable WACVD; D stands for diagnosis). WACVD, therefore, measures

Table 4 Hospital-Level Results of System Generalized Method of Moments (GMM) (WACVG_Test and WACVG_Care, Dep Var: Total LOS, Total Cost, Test-Ordering Cost, and Care-Delivery Cost)

	Total LOS			Total cost			Test cost	Care cost
Dep var	**(M1)**	**(M2)**	**(M3)**	**(M4)**	**(M5)**	**(M6)**	**(M7)**	**(M8)**
ProcQuality (PQ)	−0.000	−0.003	0.002	0.010	0.003	−0.242***	−0.022	−0.122***
	(0.005)	(0.003)	(0.029)	(0.012)	(0.011)	(0.081)	(0.066)	(0.037)
ExpQuality (EQ)	0.025	−0.040	−0.240*	0.180	0.114	−0.395*	0.551	−0.499***
	(0.042)	(0.026)	(0.127)	(0.137)	(0.118)	(0.210)	(0.456)	(0.167)
WACVG_Test		0.165***	0.153***		−0.375	−0.216	−0.461**	0.124
		(0.024)	(0.049)		(0.318)	(0.172)	(0.190)	(0.077)
WACVG_Care		0.395***	0.376***		0.335***	0.410***	0.269***	0.633***
		(0.050)	(0.065)		(0.116)	(0.121)	(0.085)	(0.095)
PQ * WACVG_Test			−0.002			0.384***	0.039	0.175***
			(0.049)			(0.129)	(0.110)	(0.059)
PQ * WACVG_Care			0.086			0.047	0.188	0.164
			(0.082)			(0.135)	(0.116)	(0.104)
EQ * WACVG_Test			0.457**			0.612**	−1.138	0.837***
			(0.233)			(0.310)	(0.788)	(0.283)
EQ * WACVG_Care			0.048			0.203	0.310	0.314
			(0.248)			(0.350)	(0.422)	(0.281)
Teaching Intensity	0.024	−0.042*	−0.017	0.043	0.078	−0.064	−0.075	−0.017
	(0.023)	(0.022)	(0.034)	(0.073)	(0.077)	(0.080)	(0.048)	(0.050)
Bed Size	0.013*	−0.000	0.007	0.019	0.002	0.002	−0.007	−0.021
	(0.008)	(0.006)	(0.008)	(0.027)	(0.025)	(0.018)	(0.011)	(0.013)
Case Mix Index	0.003	−0.001	−0.022	−0.043	−0.042	0.003	0.050	0.004
	(0.022)	(0.018)	(0.024)	(0.065)	(0.067)	(0.065)	(0.041)	(0.042)
Wage index	0.071**	0.070***	0.107***	0.554**	0.519*	0.583***	0.025	0.518***
	(0.033)	(0.023)	(0.030)	(0.226)	(0.273)	(0.186)	(0.044)	(0.127)
OPDSH adj factor	0.005	−0.051**	−0.011	−0.246**	−0.307**	−0.305***	−0.157*	−0.266***
	(0.028)	(0.025)	(0.032)	(0.112)	(0.146)	(0.107)	(0.081)	(0.067)
Outlier adj factor	−0.092	−0.114	−0.159*	0.372	0.416	0.453	0.206	0.104
	(0.094)	(0.072)	(0.097)	(0.387)	(0.462)	(0.332)	(0.140)	(0.199)
Dep Var$_{(t-1)}$	0.767***	0.723***	0.664***	0.609***	0.592***	0.566***	0.778***	0.678***
	(0.069)	(0.042)	(0.055)	(0.151)	(0.192)	(0.120)	(0.095)	(0.065)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1525	1525	1525	1525	1525	1525	1525	1525
Hospitals	324	324	324	324	324	324	324	324
Instruments	35	51	83	35	51	83	83	83
AR(1) (*p*-value)	(0.000)	(0.000)	(0.000)	(0.021)	(0.000)	(0.000)	(0.046)	(0.000)
AR(2) (*p*-value)	(0.063)	(0.182)	(0.289)	(0.421)	(0.577)	(0.626)	(0.321)	(0.862)
Hansen test of overid. (*p*-value)	(0.341)	(0.235)	(0.180)	(0.150)	(0.159)	(0.311)	(0.337)	(0.104)
Diff.-in-Hansen test of exogeneity (*p*-value)	(0.676)	(0.393)	(0.319)	(0.593)	(0.083)	(0.162)	(0.430)	(0.223)

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Notes: The results are based on a system GMM model (Arellano and Bond 1991, Blundell and Bond 1998) estimated as in Equation (10). Standard errors are corrected for heteroskedasticity and are clustered at the hospital level. AR(1) and AR(2) are tests for first-order and second-order serial correlation in the first-differenced residuals, under the null of no serial correlation. The Hansen test of overidentification is under the null that all instruments are valid. The difference-in-Hansen test of exogeneity is under the null that instruments used for the equations in levels are exogenous.

the practice variation for patients who initially got the same main diagnosis but possibly ended up with different clinical pathways. Our findings (Appendix I) are consistent and robust to this alternative practice variation measure in terms of the point of time for defining the patient cohort.

Second, we re-estimate our main models after restricting the sample to the DRGs corresponding to patients without comorbidities. This approach provides additional support to our conjectures about warranted and unwarranted practice variation. Cases that are uncomplicated by other comorbidities offer the highest-intra-cohort similarity, above and beyond controlling for other demographics. In Appendix J, we show largely consistent results. Although not directly comparable, we tend to find smaller magnitudes of coefficients for practice variation measures compared to our main analysis. This result makes sense because the degree of practice variation for DRGs without comorbidities is likely to be smaller than for those with comorbidities.

Third, we measure the practice variation metrics only for the first admission of a care episode, instead of during entire episodes of care that include test-ordering

Table 5 Summary of Hypotheses Testing Results

	Hypotheses	Empirical support
H1:	Higher overall clinical practice variation is associated with higher LOS and cost metrics.	Supported
H2a:	The positive relationships in H1 are stronger for hospitals with higher process quality (PQ).	Partially supported (in Post Hoc)
H2b:	The positive relationships in H1 are stronger For hospitals with higher experiential quality (EQ).	Supported

and care-delivery both during main admissions and during readmissions. We do so because the first admission is where the biggest risk for misdetection of a medical problem and for misdelivery of care presumably exists. As shown in Appendix K, the results are consistent with our main analysis. Fourth, we test our hypotheses condition-by-condition and consistently find supporting evidence (see Appendix L).[19]

5. Discussion and Implications

This study highlights practice variation, an important but understudied metric for research on healthcare operations and healthcare strategic planning. Using a high-frequency inpatient discharge dataset from 387 hospitals in NY and FL states with 1,094,111 patient encounters, our findings corroborate the impacts of practice variation on operational performance. Overall practice variation inside a hospital is associated with longer patient LOS and higher cost metrics, particularly in care-delivery, and this relationship is even stronger when experiential quality is high. By delving into granular dimensions of practice variation based on detailed charge data, we find that higher care-delivery practice variation is directly associated with worse operational performance. We also find that pursuing higher process quality and experiential quality may be harmful to operational performance as a result of its combined effects with the test-ordering practice variation. Therefore, it is worthwhile to carefully consider the level of practice variation in each dimension that helps to address the trade-off between high quality and low cost, which would ultimately guide hospitals to achieve the Triple Aim in healthcare (Berwick et al. 2008).

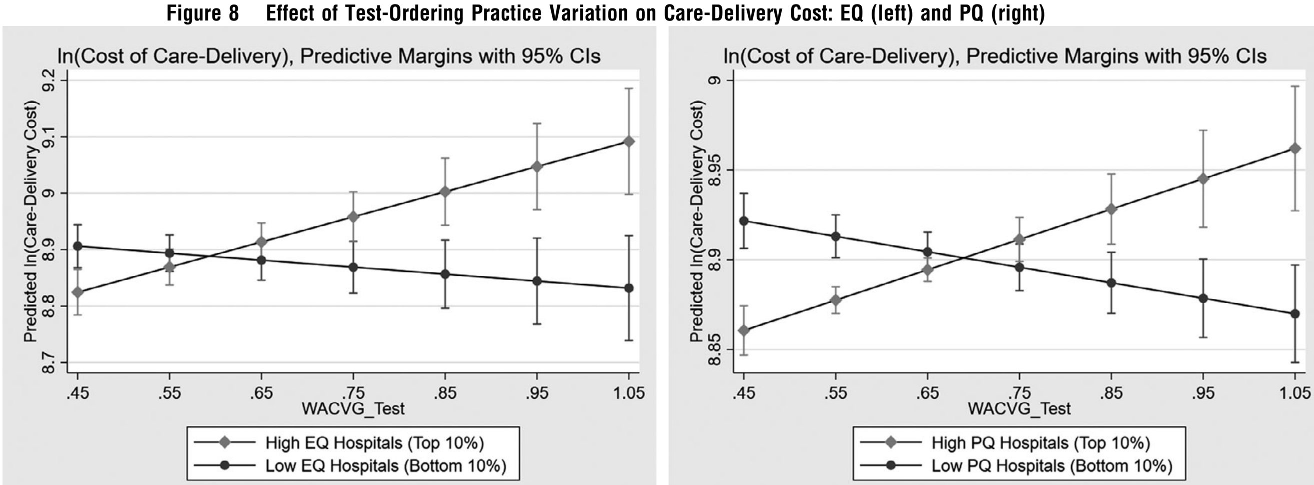
The findings are robust to alternative approaches for measuring practice variation. The strength of our empirical evidence provides assurance to researchers and healthcare managers that the level of practice variation in a hospital should be as salient a concern as the level of process and experiential quality measures, which is thus far the major focus of previous literature.

5.1. Theoretical Contributions

We suggest several implications for theory. First, our study exposes the salient role of practice variation in healthcare strategic planning, particularly focusing on operational performance. Our findings are essential because previous healthcare operations management research has neglected practice variation, even though practice variation can have a direct bearing on the financial stability of hospitals, payers, and even governments.

Our study of practice variation's role on operational performance advances the healthcare operations literature, because scholars have paid inadequate attention to how practice variation and quality initiatives can have competing direct effects on performance. We find more noticeable effects of quality measures on performance when the level of practice variation is considered together. This result provides a more complex perspective on the translation of quality improvement efforts into better performance and indicates the potential for opposing relationships among the performance drivers.

Figure 8 Effect of Test-Ordering Practice Variation on Care-Delivery Cost: EQ (left) and PQ (right)



Note: The 10th–90th percentile ranges are displayed for Experiential Quality (Left) and Process Quality (Right).

Furthermore, the interaction effects between practice variation and quality metrics render more nuanced evidence on the role of practice variation in generating better performance, or reducing waste. The operational performance metrics (e.g., LOS and care-delivery cost) in our study are more likely to improve for hospitals that reduce practice variation. Hence, there appears to be more dynamic and intricate relationships between quality initiatives (in their mean values) and hospital performance than those identified in previous research. Indeed, for research on the effect of healthcare quality attributes, our findings highlight the importance of including practice variation and its interplay with mandated quality metrics. Taken together, our study supports efforts to move beyond healthcare volume or a service mean and incorporate second-moment information like the variation.[20]

As an interesting aside, CMI, which researchers often use to measure hospital demand uncertainty, ends up insignificant in most of our analyses, unlike in previous studies that do not control for within-hospital practice variation (e.g., Ding 2014, Mishra et al. 2019, Senot et al. 2016). This finding possibly indicates that once within-hospital practice variation is taken into account, external demand uncertainty, as reflected by CMI, may not be such a salient driver of hospital performance.

5.2. Managerial Relevance and Implications

We suggest hospital managers and policymakers should pay more attention to practice variation and adopt it as an evaluation metric. A higher practice variation score indicates a greater opportunity for waste reduction and care standardization. Policymakers often desire to know if there are some efficiencies in low-spend hospitals that could be replicated in higher-spend hospitals, thus bending the curve of overall healthcare spending. In this regard, prudent policymakers may take advantage of practice variation information. Moreover, policymakers and healthcare system executives should pay attention to hospitals with higher practice variation to identify harmful practice styles and to promote improvement for a wider spectrum of healthcare providers contained within a hospital. Tracking of practice variation, especially high variation in test-ordering practice, may enhance the predictive validity of models linking quality scores to detailed cost components for patient care.

Indeed, the current paradigm of empirical research focusing solely on mean-level quality scores possibly derives inaccurate and biased assessments of hospitals or other types of healthcare providers. With low practice variation, experiential quality improvement leads to shorter LOS and lower care-delivery costs by providing timely and effective services, thus underestimating the power of quality evaluations if practice variation is ignored. Conversely, with high practice variation, experiential quality's impact on operational performance can be lower, thus overestimating the power of quality evaluations if practice variation is ignored. In this regard, without considering practice variation, hospitals may be over-(under-)rewarded for quality improvements if there is an increase (decrease) in practice variation. Thus, the design of incentives and penalties for better hospital operations should be adjusted to take such effects into account.

For the patient community, we highlight practice variation as a metric that reflects stability or instability of hospital operations. CMS releases several quality measures regularly to help consumers make informed healthcare decisions (CMS 2014). Similarly, better-informed patients with earlier or accurate practice variation information might take advantage of WACV-type measures because hospitals with higher practice variation tend to have longer LOS and higher care-delivery cost, which are not preferred by the patient community. Thus, policymakers should consider practice variation to be a vital healthcare operations metric and add it to the hospital performance dashboard.

The common attempts to mitigate unwarranted variation in practice might include equipping hospitals with business intelligence technologies, forming a task force team, designing alignment structures for physicians, and adjusting incentive models (QURE Healthcare 2018). As physicians influence 80% of healthcare spending (Crosson 2009), these approaches need active clinical collaborations with physicians, otherwise, lasting change would be difficult. Meanwhile, as our results on operational performance imply, reducing practice variation would be beneficial in both fee-for-service and value-based purchase schemes. Doing so would reduce hospital cost structures under the fee-for-service scheme and offer flexibility under the value-based purchase scheme, for example, to convey savings to healthcare providers participating in bundled payment models (Liao et al. 2018).

One shortcoming of the present uses of quality measures is that CMS provides process quality measures for only a few medical conditions. Reporting quality metrics for a limited range of care may lead to biased decision-making for patients who are not afflicted by one of the conditions on the limited condition list. Moreover, CMS annually modifies the list of care quality measures that must be reported by healthcare providers. Process quality measures with overall high performance are removed from the list when CMS considers the majority of healthcare providers across the United States to have met the quality goal (Mitchell 2014). CMS then adds new measures that have more opportunity to be improved. However, this topped-out measurement approach seems like a haphazard process of improvement. Also, it risks the existing processes going out of control again. While our study focuses on medical conditions for which process quality measures are currently available, our metric of practice variation contributes in that it can be applied to any medical condition, enabling future extensions of the quality analysis to other conditions as relevant data become available. Compared to CMI, our practice variation metric enables appropriate analyses of within-hospital variation. Having such a metric is important as it offers a venue to address the trade-off between quality and cost, as our results imply.

In addition, one should not overlook two important points. First, process standards used in performance management should be valid, in that they must either be self-evident measures of quality or be evidence-based (Lilford et al. 2007). Second, in addition to the validity of the measures, the process standards must also be beneficial to healthcare, since the opportunity cost of improving some processes may exceed the contingent gains (Demirezen et al. 2016, Hayward 2007).

Overall, our study suggests that healthcare policy should be mindful of the potential negative effects of practice variation and introduce provisions that might help to harness such variation. Meanwhile, because such practice variation also might be linked to innovation and continuous improvement in clinical practices, a careful understanding of any adjustments applied to clinical practices should be addressed together.[21]

6. Conclusions and Future Research Directions

This study contributes to the healthcare operations management literature by precisely measuring practice variation within a hospital and by examining the relationship between practice variation and hospital operational performance. From a theoretical lens of SPC, we empirically observe a positive association of practice variation on patient LOS and cost metrics per capita. From the granular investigation of practice variation, we find consistent detrimental impacts of care-delivery practice variation on the operational performance measures. We also find differential impacts of test-ordering practice variation on test-ordering cost itself and care-delivery cost, especially when the process quality and experiential quality measures are taken into account together. Such insights enable managers and policymakers to understand conditions on which better performance is achieved.

The potential limitations of our study motivate several directions for future research. First, a limited number of conditions are examined from a subset of hospitals in New York and Florida states focusing on Medicare populations. These results require confirmation on larger datasets for other regions and for a broader spectrum of patient populations. Second, practice variation in a hospital may occur for reasons personal (e.g., physician's preference (Wennberg and Gittelsohn 1973)) or organizational (e.g., changes in best practices, hospital characteristics (Westert and Groenewegen 1999)) to the hospital. Behavioral investigation of the causes of such practice variation is warranted. Third, in this study, the level of analysis of the quality measures is either at patient-cohort level or at hospital level. A more granular level (e.g., at patient level) of data could be beneficial to confirm our insights. Fourth, taking a step further, future research may extend our study by systematically measuring underuse and overuse of tests, and investigating their impacts on performance. Indeed, both underuse and overuse of tests are pervasive in the healthcare industry (Zhi et al. 2013), as perhaps also indicated by the broad spectrum of our test-ordering practice variation measure. Lastly, this study finds a relationship between practice variation and hospital operational performance. Meaningful extensions might involve testing on outcome performance to see whether mitigating practice variation can lower readmission and mortality rates,[22] and also on social efficiency to check the impact of unnecessary variation on social values interrelated with other hospitals (e.g., Greenberg and Campion 2006).

In conclusion, this study documents novel evidence for the role of practice variation in healthcare operations. Practice variation, which appears common and pervasive, may lead to poor operational performance. If not managed well, the practice variation can severely diminish the benefits of quality measurement initiatives. We hope this study promotes research to explicate this important practice variation metric further.

Acknowledgments

The authors are grateful to the department editor Sergei Savin, the senior editor, and the two anonymous reviewers for their constructive comments and valuable suggestions on this research. The authors also thank the participants at 2017–2018 Production and Operations Management Society Annual Meeting, 2017 Conference on Health IT and Analytics, 2017–2019 Institute for Operations Research and the Management Sciences Annual Meeting, 2018 Decision Sciences Institute Annual Meeting, as well as seminar participants at Texas A&M University and the University of Arizona. Their comments helped us to improve this article substantially.

Notes

[1]The six domains of healthcare waste include: failure of care delivery, failure of care coordination, overtreatment or low-value care, pricing failure, fraud and abuse, and administrative complexity.

[2]A list of acronyms in this study is provided in Appendix A.

[3]Improving the U.S. healthcare system needs simultaneous pursuit of three aims: improving the care experience, improving the population health, and reducing per capita healthcare costs (Berwick et al. 2008).

[4]1st: California \$230,090; 2nd: New York \$162,845; 3rd: Texas \$146,735; 4th: Florida \$132,463; 5th: Pennsylvania \$97,414; ⋯; Overall in the United States: \$2,089,862 (in 2009, in millions). Thus, New York and Florida states account for about 14.13% of total health spending in the United States.

[5]*Note*: A hospital identifier called *DSHOSPID* in the HCUP SID data is used to match hospitals in Florida state (using *CMS Provider Number*) and New York state (using *NY SPARCS PFI*).

[6]As a robustness check, we also considered top and bottom 3% outliers for each state, year, and MS-DRG code, resulting in consistent results.

[7]See Appendix C for different cases of patient revisit from HCUP SID Data.

[8]total cost = test-ordering cost + care-delivery cost.

[9]As previous literature pointed out significant geographic variation in cost and charges (e.g., Fisher et al. 2003, Miller et al. 2011), we estimate a separate model of Equation (1) for each condition and for each state.

[10]We can also derive this measure for each year, hospital, and condition (i.e., AMI, HF, and PN), and their histograms are shown in Figure A2 in Appendix L. We use this condition-level practice variation measure in the analysis by condition discussed in Appendix L.

[11]MS-DRG: Medicare Severity-Diagnosis Related Group.

[12]Depending on the "target value" that we discuss in Appendix I, $\bar{x}_{ijst}$ can be defined at either hospital-, county-, or state/CBSA-level.

[13]Constructing a weighted-average CV across conditions enables us to flexibly change the level of analysis. For example, WACV for hospital i and condition j in year t can be similarly computed as $WACV_{ijt} = (\sum_{s \in S_j} N_{ijst} \cdot CV_{ijst})/(\sum_{s \in S_j} N_{ijst})$, of which the estimation results are included in Appendix L.

[14]In calculation of a composite process quality measure, we drop AMI-4, HF-1, HF-4, and PN-4, which are more related to advice/counseling or instructions rather than clinical process (Andritsos and Tang 2014). However, our estimation results remain consistent even when including these measures.

[15]We can also construct a condition-level measure, which is used in robustness checks, for each hospital and year: $PQ^o_{ijt} = \frac{\sum_{m \in M_j} N_{ijmt} \cdot q_{ijmt}}{\sum_{m \in M_j} N_{ijmt}}$.

[16]No matter how we try to get close to the "true" model (e.g., by adding controls that determine *practice-quality status*), we cannot completely rule out the possibility that we have omitted an endogenous time-varying variable that has an empirically significant effect on both operational performance and *practice-quality status*.

[17]We tested the same model with different numbers of lags for each dependent variable to determine the number of lags to be included in our analysis.

[18]We acknowledge that they are not perfectly sequential and often go back and forth.

[19]We thank anonymous reviewers for their suggestions on the second, third, and fourth insights.

[20]We further discuss in Appendix N the value of using a dispersion measure in research.

[21]We acknowledge that the well-known negative association between process variation and product quality in a manufacturing environment may not always carry over to the complicated healthcare industry given the wider distribution of patient and ailment characteristics.

[22]In Appendix M, we test the impact of practice variation on a hospital's readmission rate and mortality rate, but more delicate examination on these relationships are warranted.

References

Andritsos, D. A., C. S. Tang. 2014. Linking process quality and resource usage: An empirical analysis. *Prod. Oper. Manag.* **23** (12): 2163–2177.

Appleby, J., V. Raleigh, F. Frosini, G. Bevan, G. HaiYan, T. Lyscom, H. Gao. 2011. *Variations in Health Care: The Good, the Bad and the Inexplicable*. The King's Fund, London.

Arellano, M., S. Bond. 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Rev. Econ. Stud.* **58**(2): 277–297.

Arellano, M., O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *J. Econom.* **68** (1): 29–51.

Bechel, D. L., W. A. Myers, D. G. Smith. 2000. Does patient-centered care pay off? *Jt Comm. J. Qual. Improv.* **26**(7): 400–409.

Benner, M. J., M. L. Tushman. 2003. Exploitation, exploration, and process management: The productivity dilemma revisited. *Acad. Manag. Rev.* **28**(2): 238–256.

Berlin, L. 2002. Liability for failure to order screening examinations. *Am. J. Roentgenol.* **179**(6): 1401–1405.

Berwick, D. M., A. D. Hackbarth. 2012. Eliminating waste in US health care. *J. Am. Med. Assoc.* **307**(14): 1513–1516.

Berwick, D. M., T. W. Nolan, J. Whittington. 2008. The triple aim: Care, health, and cost. *Health Aff.* **27**(3): 759–769.

Bhargava, H. K., A. N. Mishra. 2014. Electronic medical records and physician productivity: Evidence from panel data analysis. *Management Sci.* **60**(10): 2543–2562.

Blundell, R., S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *J. Econom.* **87**(1): 115–143.

Bohmer, R. M. J., T. H. Lee. 2009. The shifting mission of health care delivery organizations. *N. Engl. J. Med.* **361**(6): 551–553.

Bond, S., C. Meghir. 1994. Dynamic investment models and the firm's financial policy. *Rev. Econ. Stud.* **61**(2): 197–222.

Boulding, W., S. W. Glickman, M. P. Manary, K. A. Schulman, R. Staelin. 2011. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. *Am. J. Manag. Care.* **17**(1): 41–48.

Cachon, G., C. Terwiesch. 2008. *Matching Supply with Demand: An Introduction to Operations Management*. McGraw-Hill, Boston.

Cao, Q., E. Gedajlovic, H. Zhang. 2009. Unpacking organizational ambidexterity: Dimensions, contingencies, and synergistic effects. *Organ. Sci.* **20**(4): 781–796.

Casalino, L. P., D. Gans, R. Weber, M. Cea, A. Tuchovsky, T. F. Bishop, Y. Miranda, B. A. Frankel, K. B. Ziehler, M. M. Wong, et al. 2016. US physician practices spend more than $15.4 billion annually to report quality measures. *Health Aff.* **35**(3): 401–406.

Chandrasekaran, A., C. Senot, K. K. Boyer. 2012. Process management impact on clinical and experiential quality: Managing tensions between safe and patient-centered healthcare. *Manuf. Serv. Oper. Manag.* **14**(4): 548–566.

Chen, L. M., A. K. Jha, S. Guterman, A. B. Ridgway, E. J. Orav, A. M. Epstein. 2010. Hospital cost of care, quality of care, and readmission rates: Penny wise and pound foolish? *Arch. Intern. Med.* **170**(4): 340–346.

Clancy, C. M., K. Cronin. 2005. Evidence-based decision making: Global evidence, local decisions. *Health Aff.* **24**(1): 151–162.

Clough, J. D., K. Patel, G. F. Riley, R. Rajkumar, P. H. Conway, P. B. Bach. 2015. Wide variation in payments for medicare beneficiary oncology services suggests room for practice-level improvement. *Health Aff.* **34**(4): 601–608.

Clough, J. D., K. Patel, W. H. Shrank. 2016. Variation in specialty outpatient care patterns in the medicare population. *J. Gen. Intern. Med.* **31**(11): 1278–1286.

CMS. 2010. Fiscal Year 2009 Quality Measure Reporting for 2010 Payment Update, Centers for Medicare & Medicaid Services. Available at http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalRHQDAPU200808.pdf (accessed January 10, 2018).

CMS. 2011. National Health Expenditure (NHE) Fact Sheet, Centers for Medicare & Medicaid Services. Available at http://www.cms.gov/NationalHealthExpendData/25_NHE_Fact_Sheet.asp (accessed February 11, 2018).

CMS. 2014. Hospital Compare Downloadable Database Data Dictionary, Centers for Medicare & Medicaid Services. Available at https://data.medicare.gov/data/hospital-compare. (accessed May 15, 2017)

CMS. 2015. Risk Adjustment Fact Sheet by the Centers for Medicare and Medicaid Services. Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/Risk-Adjustment-Fact-Sheet.pdf (accessed May 15, 2017).

CMS. 2016. Design and Development of the Diagnosis Related Group (DRG) by the Centers for Medicare and Medicaid Services. Available at https://www.cms.gov/ICD10Manual/version34-fullcode-cms/fullcode_cms/Design_and_development_of_the_Diagnosis_Related_Group_(DRGs)_PBL-038.pdf. (accessed May 15, 2017).

Cohen, J., P. Cohen, S. G. West, L. S. Aiken. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum, Mahwah, NJ.

Crosson, F. J. 2009. Change the microenvironment. Delivery system reform essential to control costs. *Modern Healthc.* **39**(17): 20–21.

Dai, T., M. Akan, S. Tayur. 2016. Imaging room and beyond: The underlying economics behind physicians' test-ordering behavior in outpatient services. *Manuf. Serv. Oper. Manag.* **19**(1): 99–113.

Demirezen, E. M., S. Kumar, A. Sen. 2016. Sustainability of healthcare information exchanges: A game-theoretic approach. *Inform. Syst. Res.* **27**(2): 240–258.

Ding, D. X. 2014. The effect of experience, ownership and focus on productive efficiency: A longitudinal study of us hospitals. *J. Oper. Manag.* **32**(1): 1–14.

Donabedian, A. 1980. *The Definition of Quality and Approaches to its Assesment*, vol. 1. Health Administration Press, Chicago, IL.

Donabedian, A. 1988. The quality of care: How can it be assessed? *J. Am. Med. Assoc.* **260**(12): 1743–1748.

Drye, E. E., S.-L. T. Normand, Y. Wang, J. S. Ross, G. C. Schreiner, L. Han, M. Rapp, H. M. Krumholz. 2012. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: An observational study with implications for hospital profiling. *Ann. Intern. Med.* **156** (1_Part_1): 19–26.

Eddy, D. M. 1984. Variations in physician practice: The role of uncertainty. *Health Aff.* **3**(2): 74–89.

Elwyn, G., A. Edwards, P. Kinnersley, R. Grol. 2000. Shared decision making and the concept of equipoise: The competences of involving patients in healthcare choices. *Br. J. Gen. Pract.* **50** (460): 892–899.

Ferenc, D. P. 2013. *Understanding Hospital Billing and Coding*. Elsevier Health Sciences, Amsterdam.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, E. L. Pinder. 2003. The implications of regional variations in medicare spending. Part 1: The content, quality, and accessibility of care. *Ann. Intern. Med.* **138**(4): 273–287.

Fonarow, G. C., E. D. Peterson. 2009. Heart failure performance measures and outcomes: Real or illusory gains. *J. Am. Med. Assoc.* **302**(7): 792–794.

Gandhi, T. K., A. Kachalia, E. J. Thomas, A. L. Puopolo, C. Yoon, T. A. Brennan, D. M. Studdert. 2006. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. *Ann. Intern. Med.* **145**(7): 488–496.

Garvin, D. A. 1988. *Managing Quality: The Strategic and Competitive Edge*. Simon and Schuster, New York, NY.

Greenberg, L., D. Campion. 2006. *Efficiency in health care: What does it mean? How is it measured? How can it be used for value-based purchasing?*. 290-04-0001, Academy Health, AHRQ, 290-04-0001.

Grytten, J., L. Monkerud, R. S⊘rensen. 2016. Practice guidelines and practice variation: Diagnostic technology in maternity care. A. Johnson, T. A. Stukel, eds. *Medical Practice Variations. Medical Practice Variations*. Springer Reference, New York, NY, 505–517.

Gupta, D., S. J. Potthoff, et al. 2016. Matching supply and demand for hospital services. *Foundat. Trends® Tech. Inform. Oper. Manag.* **8**(3–4): 131–274.

Ham, C. 1988. *Health Care Variations: Assessing the Evidence*. King's Fund Institute, London.

Hamilton, L. C. 2012. *Statistics with Stata: Version 12*. Cengage Learning, Boston, MA.

Hansen, L. P., K. J. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econom. J. Econom. Soc.* **50**(5): 1269–1286.

Hayward, R. A. 2007. Performance measurement in search of a path. *N. Engl. J. Med.* **356**(9): 951–953.

He, Z.-L., P.-K. Wong. 2004. Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis. *Organ. Sci.* **15** (4): 481–494.

Horwitz, L., C. Partovian, Z. Lin, J. Herrin, J. Grady, M. Conover, J. Montague, C. Dillaway, K. Bartczak, J. Ross. 2011. Hospital-wide (all-condition) 30-day risk-standardized readmission measure. Technical report, Yale New Haven Health Services Corporation Center for Outcomes Research & Evaluation.

IOM. 2012. Iom: 30% of health spending was waste. Available at https://www.advisory.com/Daily-Briefing/2012/09/07/IOM-report (accessed August 21, 2017).

Jha, A. K. 2006. Measuring hospital quality: What physicians do? How patients fare? or both? *J. Am. Med. Assoc.* **296**(1): 95–97.

Kapu, A. N., A. P. Wheeler, B. Lee. 2014. Addition of acute care nurse practitioners to medical and surgical rapid response teams: A pilot project. *Crit. Care Nurse* **34**(1): 51–59.

Kelley, K. 2007. Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behav. Res. Methods* **39**(4): 755–766.

Khunlertkit, A., P. Carayon. 2013. Contributions of tele–intensive care unit (Tele-ICU) technology to quality of care and patient safety. *J. Crit. Care* **28**(3): 315–e1.

Kuhnen, C. M., A. Niessen. 2012. Public opinion and executive compensation. *Management Sci.* **58**(7): 1249–1272.

Lallemand, N. C., et al. 2012. Reducing waste in health care. *Health Aff.* **13**: 1–5.

Li, L. X., W. C. Benton. 1996. Performance measurement criteria in health care organizations: Review and future research directions. *Eur. J. Oper. Res.* **93**(3): 449–468.

Liao, J. M., E. J. Emanuel, G. L. Whittington, D. S. Small, A. B. Troxel, J. Zhu, W. Zhong, A. S. Navathe. 2018. Physician practice variation under orthopedic bundled payment. *Am. J. Manag. Care*. **24**(6): 287–293.

Lilford, R. J., C. A. Brown, J. Nicholl. 2007. Use of process measures to monitor the quality of clinical practice. *BMJ.* **335**(7621): 648.

March, J. G. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1): 71–87.

Melnick, G. A., K. Fonkych. 2008. Hospital pricing and the uninsured: Do the uninsured pay higher prices? *Health Aff.* **27**(2): w116–w122.

Miller, D. C., C. Gust, J. B. Dimick, N. Birkmeyer, J. Skinner, J. D. Birkmeyer. 2011. Large variations in Medicare payments for surgery highlight savings potential from bundled payment programs. *Health Aff.* **30**(11): 2107–2115.

Mishra, S., P. A. Salzarulo, S. B. Modi. 2019. Patient care effectiveness and financial outcomes of hospital physician contracting emphasis. *J. Oper. Manag.* Forthcoming.

Mitchell, K. 2014. Measuring What Matters: CMS Introduces a New Methodology for "topped out" Measures. *Morning Consult*, October 23 2014. Available at http://morningconsult.com/opinions/measuring-matters-cms-introduces-new-methodology-topped-measures/ (accessed June 2, 2017).

Mossialos, E., M. Wenzl, R. Osborn, C. Anderson. 2016. *2015 International Profiles of Health Care Systems*. The Commonwealth Fund, New York, NY.

Nair, A., M. Nicolae, R. Narasimhan. 2013. Examining the impact of clinical quality and clinical exibility on cardiology unit performance: Does experiential quality act as a specialized complementary asset? *J. Oper. Manag.* **31**(7): 505–522.

Narayan, V., V. Kadiyali. 2015. Repeated interactions and improved outcomes: An empirical analysis of movie production in the United States. *Management Sci.* **62**(2): 591–607.

NASEM. 2016. *Improving Diagnosis in Health Care, National Academies of Sciences, Engineering, and Medicine*. National Academies Press, Washington, DC.

Oakland, J. S. 2007. *Statistical Process Control*. Routledge, New York, NY.

OECD. 2016. OECD Health statistics 2016. Available at https://innovation.cms.gov/Files/x/ACE-Solicitation.pdf (accessed May 24, 2018).

Pang, M.-S., A. Tafti, M. S. Krishnan. 2016. Do CIO IT budgets explain bigger or smaller governments? Theory and evidence from US state governments. *Management Sci.* **62**(4): 1020–1041.

Papanicolas, I., L. R. Woskie, A. K. Jha. 2018. Health care spending in the United States and other high-income countries. *J. Am. Med. Assoc.* **319**(10): 1024–1039.

QURE Healthcare. 2018. Unwarranted Variation: A Preventable National Crisis. Available at https://www.qurehealthcare.com/news/2018/10/28/unwarranted-variation-a-preventable-national-crisis (accessed May 11, 2019).

Rego, L. L., N. A. Morgan, C. Fornell. 2013. Reexamining the market share–customer satisfaction relationship. *J. Mark.* **77**(5): 1–20.

Roodman, D. 2006. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata J.* **9**(1): 86–136.

Rubin, H. R., P. Pronovost, G. B. Diette. 2001. The advantages and disadvantages of process-based measures of health care quality. *Int. J. Qual. Health Care.* **13**(6): 469–474.

Sadeghi, S., A. Barzi, M. M. Shabot, O. Mikhail. 2012. *Integrating Quality and Strategy in Health Care Organizations*. Jones & Bartlett Publishers, Burlington, MA.

Senot, C., A. Chandrasekaran, P. T. Ward, A. L. Tucker, S. D. Moffatt-Bruce. 2016. The impact of combining conformance and experiential quality on hospitals' readmissions and cost performance. *Management Sci.* **62**(3): 829–848.

Shi, L., D. A. Singh. 2012. *Essentials of the US Health Care System*. Jones & Bartlett Publishers, Burlington, MA.

Shrank, W. H., T. L. Rogstad, N. Parekh. 2019. Waste in the us health care system: Estimated costs and potential for savings. *J. Am. Med. Assoc.* Forthcoming.

Shwartz, M., A. B. Cohen, J. D. Restuccia, Z. J. Ren, A. Labonte, C. Theokary, R. Kang, J. Horwitt. 2011. How well can we identify the high-performing hospital? *Med. Care Res. Rev.* **68** (3): 290–310.

Sipkoff, M. 2003. 9 ways to reduce unwarranted variation. *Manag. Care.* **12**(11): 20–4.

Smolowitz, J., E. Speakman, D. Wojnar, E.-M. Whelan, S. Ulrich, C. Hayes, L. Wood. 2015. Role of the registered nurse in primary health care: Meeting health care needs in the 21st century. *Nurs. Outlook.* **63**(2): 130–136.

Theokary, C., Z. J. Ren. 2011. An empirical study of the relations between hospital volume, teaching status, and service quality. *Prod. Oper. Manag.* **20**(3): 303–318.

Toussaint, J. 2009. Writing the new playbook for us health care: Lessons from Wisconsin. *Health Aff.* **28**(5): 1343–1350.

Weick, K. E., K. M. Sutcliffe. 2006. Mindfulness and the quality of organizational attention. *Organ. Sci.* **17**(4): 514–524.

Wennberg, J. E. 2002. Unwarranted variations in healthcare delivery: Implications for academic medical centres. *BMJ.* **325** (7370): 961.

Wennberg, J. E., A. M. Gittelsohn. 1973. Small area variations in health care delivery. American Association for the Advancement of Science, Washington, DC.

Wennberg, J. E., E. S. Fisher, J. S. Skinner. 2002. Geography and the debate over medicare reform. *Health Aff.* **21**(2): 10–10.

Westert, G. P., P. P. Groenewegen. 1999. Medical practice variations: Changing the theoretical approach. *Scand. J. Public Health.* **27**(3): 173–180.

Wheeler, D. J., D. S. Chambers, et al. 1992. *Understanding Statistical Process Control*. SPC Press. Knoxville, TN.

Wintoki, M. B., J. S. Linck, J. M. Netter. 2012. Endogeneity and the dynamics of internal corporate governance. *J. Financ. Econ.* **105**(3): 581–606.

Zhi, M., E. L. Ding, J. Theisen-Toupal, J. Whelan, R. Arnaout. 2013. The landscape of inappropriate laboratory testing: A 15-year meta-analysis. *PLoS ONE.* **8**(11): e78962.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Table of Acronyms
Appendix B: Variable Descriptions
Appendix C: Decision on Revisits from HCUP SID Data
Appendix D: Measurement of Practice Variation: An Illustration
Appendix E: Summary Statistics, Admission Types, and Race of Patient Sample by Condition
Appendix F: Measures and Summary Statistics for Process Quality & Experiential Quality (2007–2013)
Appendix G: Correlation Table for Key Variables
Appendix H: Charge Components by Condition
Appendix I: Robustness Check: Ex-ante Type of Practice Variation
Appendix J: Robustness Check: Practice Variation for DRGs w/o Comorbidities
Appendix K: Robustness Check: Practice Variation based on Initial Admissions
Appendix L: Robustness Check: Results by Condition
Appendix M: Post hoc Analysis: Patient Outcomes as Dependent Variables
Appendix N: Impact of Dispersion on Value