Received: 30 January 2019 | Accepted: 11 February 2019
DOI: 10.1002/joom.1028

RESEARCH ARTICLE

WILEY

Operational efficiency and patient-centered health care: A view from online physician reviews

Dong-Gil Ko[1] | **Feng Mai**[2] | **Zhe Shan**[3] | **Dawei Zhang**[4]

[1]Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, Ohio
[2]School of Business, Stevens Institute of Technology, Hoboken, New Jersey
[3]Farmer School of Business, Miami University, Oxford, Ohio
[4]College of Business & Economics, Lehigh University, Bethlehem, Pennsylvania

Correspondence
Dong-Gil Ko, Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221.
Email: donggil.ko@uc.edu

Handling Editor: Anita Tucker

Abstract
Online reviews are playing an increasingly important role in how patients select and evaluate health-care providers. Physician rating websites not only act as open platforms for patients to share their experiences, but can also offer valuable feedback for physicians to improve the quality of care. In this study, we analyze over one million physician reviews across 17 medical specialties and investigate the relationship between operational efficiency and patient satisfaction. We combine econometrics models with text analytics techniques to quantify the effect using both patients' ratings of physicians and their qualitative review narratives. The results provide strong empirical evidence that operational inefficiency negatively influences patient satisfaction. Specifically, a waiting time longer than 17 min will, on average, reduce the odds of getting a high rating status by 14%. Though many health care ratings examined in this study do not mitigate the negative effects brought on by long waiting time, patient narratives reflecting the importance of technical and interpersonal qualities to patients suggest a more complex set of relationships between waiting time and patient satisfaction. Our study showcases the usefulness of online physician reviews and reveals unique insights for improving the delivery of patient-centered health care.

KEYWORDS
experiential quality, operational efficiency, patient satisfaction, text analytics, waiting time

1 | INTRODUCTION

Health care today demands more patient-centered health care. Patient-centeredness, or experiential quality, refers to the quality of health care as perceived by a patient. As patients today are taking a more active role in selecting physicians (Salzarulo, Bretthauer, Côté, & Schultz, 2011) and new reimbursement policies are incentivizing the delivery of patient-centered medical care (Senot, Chandrasekaran, & Ward, 2016), patient-centered health care has become a major concern for physicians as well as for patients. The general decline in patients' reports of primary care experiences (Murphy, Chang, Montgomery, Rogers, & Gelb Safran, 2001) lack of transparency about health-care providers (Agency for Healthcare Research and Quality, 2011) and a greater dedication to improving health-care quality transparency at the institutional level (Harris & Buntin, 2008; Jha, Premarajan, Nagesh, Khanal, & Thapa, 2005) underscore the importance of patients' access to health-care information. Given that the information asymmetry patients experience is particularly acute, and little information about the quality of a physician makes its way into the public domain (Gao, Greenwood, McCullough, & Agarwal, 2015), patients are shifting toward online health information. As a result, reviews of health experiences and ratings of medical providers through social media platforms are assuming an increasingly important role as patients consume and share at a

All authors contributed equally to this study.

startling rate (Fox & Jones, 2012; Hay, Strathmann, Lieber, Wick, & Giesser, 2008).

A national survey shows that 59% of U.S. respondents are relying on social media rating sites when choosing a physician (Hanauer, Zheng, Singer, Gebremariam, & Davis, 2014). Unlike physician-centered surveys and research, such as proliferation of patient-generated content not only represents patients' public and independent perspectives about health-care quality (The Economist, 2014), but also serves as a “missing link both for consumers seeking to understand the experience of other patients and for providers seeking to learn from patients to improve quality” (Schlesinger, Grob, Shaller, et al. 2015, p. 675). Their ratings and reviews point to the desire for a patient-centered health-care system and provide a new and promising venue for studying health care from a patient's perspective. Moreover, a large-scale distribution of information spanning multiple centers, physicians, and specializations offered through physician review websites can accommodate more comprehensive peer groups than has been traditionally possible (Xu, Armony, & Ghose, 2016). For these reasons, online physician reviews can be a significant driver of patient-centeredness in the health-care delivery system (Lee, 2017).

Although patient-centered health care has received considerable research attention (Chandrasekaran, Senot, & Boyer, 2012; Nair, Nicolae, & Narasimhan, 2013; Senot et al., 2016; Sharma, Chandrasekaran, Boyer, & McDermott, 2016), there is a limited empirical research examining the factors that facilitate patient satisfaction (Ancarani, Di Mauro, & Giammanco, 2011; Douglas & Fredendall, 2004, p. 864; Queenan, Angst, & Devaraj, 2011). Few recent studies (Dobrzykowski, Callaway, & Vonderembse, 2015; Senot et al., 2016) focus on the organizational antecedents of experiential quality at the hospital level. However, the voice of the patients as to what constitutes satisfactory medical care is largely ignored. In particular, there is a dearth of empirical research examining the role of operational efficiency from a patient's perspective, its interplay with other factors, and their collective impact on patient satisfaction. And, patients' values voluntarily expressed through textual reviews and their influence on patient satisfaction have not previously been explored. Accordingly, this study examines patient satisfaction using a comprehensive data set of online physician ratings and textual reviews and focuses on the impact of operational efficiency on patient-centered health care.

Specifically, we analyze operational performance embedded in the online reviews by examining the impact of waiting time from a patient's perspective. Our unique data set contains over one million physician ratings and textual reviews that span 17 medical specialties and patients from all 50 U.S. states. Our research framework considers the effects of waiting time, physician ratings, and specific topics that patients find important as reflected in textual reviews. Our models show that waiting time negatively impacts patient satisfaction. This relationship is both statistically and economically significant. Moreover, the importance of waiting time differs across specialties. Our analyses also reveal positive and significant interaction effects between waiting time with Ease of Appointment and Postvisit Follow-up, suggesting that ensuring appointments can easily be made and reaching out to a patient after a visit, respectively, are effective strategies for managing the perceived negative patient satisfaction resulting from waiting time. We combine theory with machine learning methods to compute the importance of quality dimensions to patients in the textual reviews and show that timeliness-related issues account for more than 18 % of the narratives. Using the importance of quality dimensions as moderating variables, we show that the negative impact on patient satisfaction brought on by waiting time is not as prominent for those patients who value technical quality. On the other hand, the negative impact on patient satisfaction brought on by waiting time is more significant for those patients who value interpersonal quality.

Our study offers two contributions to health-care operation management. First, to the best of our knowledge, our study is one of the first to examine operational efficiency from a social media perspective using patients' voluntary reviews—a response to a call for operations management (OM) research incorporating online user-generated content and social media data (Abrahams, Fan, Wang, Zhang, & Jiao, 2015; Chan, Wang, Lacka, & Zhang, 2016; Chen, Zheng, & Ceran, 2016; Pedraza-Martinez & Van Wassenhove, 2016; Tang, 2015). Through the lens of online patient-generated physician reviews, we analyze patients' narratives, while considering how important operational efficiency may mean for them. This effort is valuable for patient-centered health care given the (a) information asymmetry issues and (b) availability of “big-data” techniques for analyzing large volumes of physician reviews. Online reviews overcome information transparency and systematic feedback challenges experienced in physicians' attempts to deliver an effective, efficient, and patient-centered health-care system. Traditional methods such as surveys or focus groups have inherent limitations such as social desirability bias, time lag before measurement, and small sample size. Information on social media offers a unique opportunity to learn about patients' values, needs, and preferences directly, while overcoming these limitations (Verhoef, Van de Belt, Engelen, Schoonhoven, & Kool, 2014). In addition, a large number of online physician reviews combined with “big-data” techniques can promote data-driven approaches to investigate the patient-physician relationship. Our study demonstrates how online patient-generated physician reviews can help health-care providers understand their patients and develop strategies to address underperforming areas. In sum, integrating online physician reviews with ongoing quality improvement efforts can offer powerful insights in improving the delivery of health care, patient satisfaction, and ultimately patient health outcome.

Second, our results provide a more comprehensive understanding of the impact of operational efficiency on patient satisfaction. Quantitatively, the extant literature provides relatively simple and inconclusive accounts for the impact of waiting time on patient satisfaction, usually based on correlation measures. Our study uses rigorous econometric models to show that longer waiting time impacts overall patient satisfaction, and the effect is likely to be causal. Qualitatively, while prior research has looked at the textual reviews using content analysis methods (López, Detz, Ratanawongsa, & Sarkar, 2012), sample sizes have been confined to a few hundred due to the inherent limitation of human raters. We leverage a machine learning method called topic modeling (Blei, 2012) to extract meaningful information from over one million textual reviews. More importantly, we introduce a guided topic modeling approach to connect theory with a data-driven method. The new approach allows us to measure the importance of two theoretical dimensions—technical and interpersonal quality—that moderate the effects of waiting time on patient satisfaction. Overall, we augment prior findings by integrating patient-generated quantitative and qualitative information for understanding the role of operational efficiency—a step closer to realizing patient-centered health care. Our study showcases that as a new frontier for OM research, social media data combined with data-driven methods (Lam, Yeung, & Cheng, 2016; Simchi-Levi, 2013) enable us to explore novel approaches for improving patient satisfaction.

In the next section, we review prior literature and develop our hypotheses. Methods and analytical techniques are then presented along with our analyses and results. Finally, we draw implications for research and practice and end with limitations and concluding remarks.

2 | LITERATURE AND HYPOTHESES

2.1 | Online physician reviews

Providing the best health care possible continues to be a challenge given the limited access to health-care resources and rising costs. Though studies to date have suggested the importance of information transparency and greater access to physician information will improve overall health-care quality and cost, failures abound (Agency for Healthcare Research and Quality, 2011). Several transparency initiatives led by the centers for medicare & medicaid services (CMS), including physician quality reporting system (PQRS), have been unsuccessful (Centers for Medicare and Medicaid Services, 2011). Lack of transparent, explicit, systematic, data-driven performance measurement and feedback mechanisms concerning health-care providers make it difficult to transform health care to patient-centered health care (Luxford, 2012). Thus, limited access to information about health-care providers has necessitated a call for patients to assume an increasingly important role in facilitating the design and management of health-care systems.

Meanwhile, patients are turning to the internet-based, patient-generated physician reviews to help guide their decision-making process. Nearly 60% of U.S. adults rely on online health information resources to guide their decisions, and this number is expected to continue to grow by 90% annually (Fox & Jones, 2012). Although some physicians have reservations about the quality of online feedback, research has shown online ratings are highly correlated with survey measures of patient experience in family practices and hospitals (Greaves et al., 2012; Kadry, Chu, Kadry, Gammas, & Macario, 2011). Gao et al. (2015) compare online physician ratings with patient-perceived quality measures using a questionnaire developed by the Agency for Healthcare Research & Quality (AHRQ) and confirmed the positive correlation between online ratings and patient-perceived physician quality. There is also no evidence that online reviews are dominated by disgruntled patients (Gao et al., 2015). Together, the evidence suggests that patient-generated online reviews can be an effective driver of patient-centered health care and can be used to track health-care delivery performance (Lee, 2017).

As a result, online physician reviews and their implications have received considerable research attention recently. Broadly speaking, the existing literature falls into three groups. One group explores health-care social media sites as a new data source; they focus on the availability of information and the distribution of ratings (see Verhoef et al., 2014 for a review). Another group examines the connection between social media ratings and "true" quality of care (Gao et al., 2015; Greaves et al., 2012; Segal et al., 2011). The third group offers perspectives on how online reviews can be integrated into the health-care system and discusses related policy issues (Lee, 2017; Schlesinger, Grob, and Shaller 2015; Schlesinger, Grob, Shaller, et al. 2015).

This article differs from earlier studies in two ways as depicted in our conceptual framework (Figure 1). First, we focus on the significance of operational efficiency and its impact on patient satisfaction (H1). Prior studies on physician reviews have not examined the implications of operations-related issues. Since operations management is directly involved with delivery policies and allocation of resources, understanding its impact from a patient's perspective is imperative.

Second, most prior studies have not examined textual reviews—a distinctive data source that is voluntary, without

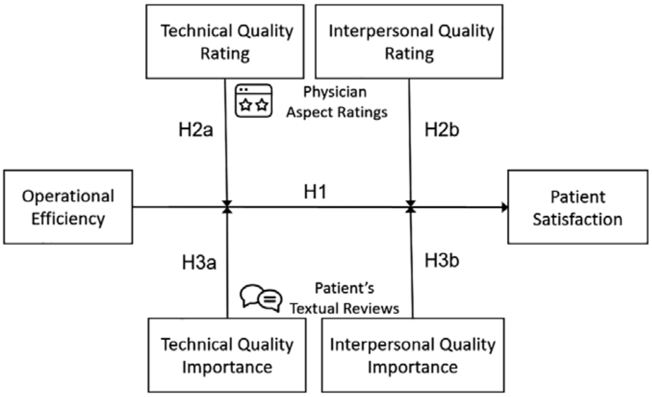

FIGURE 1 Conceptual framework

predetermined structure, and unguided for what information should be entered. These narratives often describe physicians' operations-related matters that are not captured by waiting time, thereby highlighting untapped, yet important, factors that are traditionally missing from predetermined surveys. Thus, physician textual reviews convey fine-grained evidence that cannot be fully captured with numerical ratings (Pavlou & Dimoka, 2006). Unlike ratings, which are designed to capture physicians' performance along predetermined attributes, textual reviews are unstructured narratives aimed at capturing patients' experiences that they find important. Therefore, a textual review is an expression of values and importance made by a patient who conveys attributes based on an encounter with a physician. Though there are many advantages to textual reviews, analyzing a large volume of unstructured data might be a challenging task for health-care organizations. In this study, we apply machine learning techniques to extract the importance of health-care quality dimensions from textual reviews. Since textual reviews act as an interpretive lens offering complementary information to physician ratings (Schlesinger, Grob, Shaller, et al. 2015), the insights derived should enhance our understanding of patients' perceptions of health care.

As a result, we delineate two different perspectives held by a patient: (a) an outward-looking perspective about his or her physician's competence, which is the basis for our H2 arguments (e.g., an outcome-based performance evaluation of a physician regarding a physician's expertise, professionalism, and competence), and (b) an inward-looking perspective regarding a patient's own disposition about the importance of certain attributes, which serves as the basis for our H3 arguments.

2.2 | Operational efficiency and patient satisfaction

At the heart of the patient-centered approach to evaluating and improving health care lies patient satisfaction (Grondahl, Wilde-Larsson, Karlsson, & Hall-Lord, 2013). Patient satisfaction is defined as a patient's judgment of the overall experience after receiving a medical service (Ancarani et al., 2011; Marley, Collier, & Meyer Goldstein, 2004). Patient satisfaction is not only becoming increasingly important as patients assume a more active role in selecting physicians, it is also serving as a de facto indicator of health-care quality (Salzarulo et al., 2011). It is often linked to performance measures, such as growth (Goldstein, 2003) and profitability (Ancarani et al., 2011). Despite its growing importance, patient satisfaction remains an under-theorized concept (Gill & White, 2009) and an investigation of its antecedents to date has been limited (Ancarani et al., 2011; Douglas & Fredendall, 2004).

In contrast to clinical quality, which aims to evaluate physicians based on their technical performance focusing on a "cure" system, or "what" health-care services are rendered, equally important is experiential quality that emphasizes a custom-tailored approach to a "care" system, or "how" health care is delivered for addressing the unique needs of each patient (Bensing, 1991; Boyer, Gardner, & Schweikhart, 2012; Donabedian, 1988; Harvey, 1998). The interpersonal dynamics between a physician and a patient relate to an external capability that emphasizes customer satisfaction (Sousa, 2003; Thirumalai & Sinha, 2011) as well as the level of interactions with a physician as experienced by a patient (Chandrasekaran et al., 2012; Nair et al., 2013). With the growing influence of experiential quality on hospital readmissions, costs, and other performance metrics (Nair et al., 2013; Senot, Chandrasekaran, Ward, Tucker, & Moffatt-Bruce, 2015), there is a growing interest among operation management researchers to investigate the drivers of experiential quality as a means to improve patient-centered health care (Chandrasekaran et al., 2012; Green, 2012; Senot et al., 2016).

Operational efficiency is an important element of experiential quality in the context of patient-centered health care and is commonly measured using *waiting time*—waiting time is considered as a "patient-centered metric" (Froehle & Magazine, 2013). Waiting time in the health-care industry is typically associated with time waiting in the office and "exam room"—that is, elapsed time between checking in at the front desk to a meeting with a physician. Despite its widespread adoption, limited existing studies offer mixed results on whether operational efficiency matters for clinical or experiential quality. For example, in inpatient settings or emergency departments, increased patient waiting time resulted in low patient satisfaction (Haraden & Resar, 2004); in outpatient settings (both primary and specialty care); however, many studies found no or weak association between patient waiting time and satisfaction (McCarthy, McGee, & O'Boyle, 2000; Zandbelt, Smets, Oort, Godfried, & De Haes, 2004). Huang (1994) suggests that patients can wait for as long as 37 min without negatively influencing their satisfaction. Chandrasekaran et al. (2012) find that hospitals' emphasis on process management is associated with decreases in experiential quality. These conflicting findings prompt further investigation, because health-care operations management needs to consider the fundamental trade-off between physician utilization and patient waiting time (Robinson & Chen, 2011). Also, physicians face the dual objectives of conformance quality and experiential quality that compete for limited resources (Senot et al., 2015).

Waiting a long time for a physician is generally a frustrating and unpleasant experience for patients. A shift in average work week from 40.6 hr in 1973 to 47 hr in 2014 and over 60 hr as a rule in demanding, competitive industries means that working professionals are placing greater emphasis and value in their time (Larson, Larson, & Katz, 1991; Saad, 2014). More than 30% of patients experiencing long waiting times leave before seeing a physician, while 20% will change physicians (Heath, 2018). In fact, patients today use waiting time as one deciding factor when choosing a new physician. It is becoming clear that waiting time can be detrimental to a patient experience, especially when waiting too long may worsen their conditions thereby potentially compounding the complexity of health care.

Operation management literature and anecdotal evidence highlight the important relationship that long waiting time negatively impacts patient satisfaction (Cayirli & Veral, 2003; Haraden & Resar, 2004). We argue that online reviews provide a promising venue for gaining an understanding of this relationship, and additional insights can help physicians effectively manage their utilization of time and promote efficient use of clinical resources. We formulate our main hypothesis as:

H1 As the waiting time to be seen by a physician increases, patient satisfaction decreases.

2.3 | Moderating effects of physician's quality dimensions

Health-care quality is a multidimensional construct and, therefore, the relationship between waiting time and patient satisfaction could also depend on how well physicians perform in other aspects of an encounter (Dagger, Sweeney, & Johnson, 2007)—that is, patients' perspectives about physicians. In fact, prior studies have shown that patients evaluate their physician's performance based on technical qualities ("what" service is delivered), interpersonal qualities ("how" the service is delivered), or both (Fung et al., 2005; Sharma & Patterson, 1999). These quality assessments refer to a patient's evaluation of an appointment journey analogous to a consumer purchase journey (Lemon & Verhoef, 2016) and revolves around a patient's point of view involving cognitive, emotional, behavioral, sensorial, and social components (Schmitt 2011, Verhoef et al., 2009). Patients are becoming more cognizant and taking greater control of the overall appointment journey recognizing their potential influence through social media outlets, such as Vitals, HealthGrades, RateMDs, and UCompareHealthcare. Both social media outlets and established patient satisfaction surveys (e.g., Press Ganey) routinely evaluate quality metrics (e.g., technical and interpersonal qualities) commonly related to patient satisfaction and those that may offset the dissatisfaction experienced while waiting. Therefore, we posit that the relationship between patients' waiting time and patient satisfaction is moderated by physician ratings on two primary dimensions of health-care service quality: technical quality and interpersonal quality (Dagger et al., 2007).

Technical quality: Technical quality is defined as the expertise, professionalism, and competency of a service provider (Dagger et al., 2007) and refers to physicians' competencies associated with an analysis and identification of the cause or nature of a condition or disease from its signs and symptoms to ensure the most effective treatment. A recent study showed that a total of 118 physicians correctly diagnosed 55% of easier and 6% of difficult medical cases (Meyer, Payne, Meeks, Rao, & Singh, 2013). This indicates that all medical facilities—private office physicians to specialists in hospitals—spanning from life-threatening to non-life-threatening cases misdiagnose a significant number of patients' medical needs, which ultimately lead to poor treatment plans. Clearly, patients are seeking higher levels of expertise, professionalism, and competency from their physicians (e.g., Mühlbacher, Johnson, Yang, Happich, & Belger, 2016). Though negative emotions presented at the waiting stage may affect attitudes toward subsequent interactions and manifest into online reviews (Rozin & Royzman, 2001), we speculate that the positive effects brought by high levels of technical quality leading to effective treatment plans will outweigh the negativity associated with waiting time. Similar observations are also witnessed in other settings.

Marketing research suggests that customers are willing to tolerate longer waiting time, if they perceive their experience will be of high quality (Lu, Musalem, Olivares, & Schilkrut, 2013)—for example, though waiting in a line at an amusement park ride is not pleasant, customers tolerate long waiting time with an expectation that they enjoy their experiential outcome. Operations management literature points out that quality signaling inferred from quality assessment of experience and waiting time positively influence customers' evaluation of service providers (Veeraraghavan & Debo, 2011). These suggest that the negative consequences of waiting time may not be as pronounced depending on the expectations held about the experiential outcome. In other words, a positive experience perceived by a patient based on a physician's expertise, professionalism, and competency regarding her health conditions could erode the negative experience brought on by waiting time. We therefore hypothesize:

H2a A physician's technical quality positively moderates the relationship between waiting time and patient satisfaction.

Interpersonal quality: Interpersonal quality reflects the relationship and interplay between a physician and a patient involving attributes such as manners and communication (Dagger et al., 2007). Interpersonal quality attempts to distinguish attitudes and behaviors that go beyond an appointment itself; rather, the focus is on a psychological state that occurs by virtue of interaction and relationship development with a physician (Brodie, Hollebeek, Jurić, & Ilić, 2011). For example, a positive interaction with a friendly, courteous, and respectful staff (Stewart, Nápoles-Springer, Gregorich, & Santoyo-Olsson, 2007) creates a welcoming environment when visiting the office—the first point of contact in an appointment journey. Another important theme in interpersonal quality, bedside manner, refers to a physician's approach or attitude toward providing health-care services to a patient and is commonly referred to as “mannerism,” “etiquette,” or “sociability” (e.g., Spake & Megehee, 2010). Bedside manner is important because it affects how patients feel about the appointment journey; it encompasses every aspect of an interaction with a physician including “what” is said and “how” it is expressed. Spending time with a patient is another important aspect of interpersonal quality (Haas-Wilson, 1994); it is concerned with the extent to which a physician listens, observes, and explains to patients.

Prior research suggests that as more time is given to a patient (e.g., consultation, education, and listening), the greater the quality of health care and patient satisfaction. Time allotted to “get to know” patients attributed to improved care and satisfaction (Attree, 2001); unhurried approach and taking the time to talk, listen, and be with a patient resulted in positive responses (Larrabee & Bolden, 2001; Milburn, Baker, Gardner, Hornsby, & Rogers, 1994). Other studies found that reduced time spent with a physician combined with increased waiting time coincided with notable drops in patient satisfaction (Camacho, Anderson, Safrit, Jones, & Hoffmann, 2006). More recently, several studies have highlighted the increasingly important roles of spending more time, not feeling rushed, and offering genuine care—that is, displays of concern and care characteristic of interpersonal quality—and that waiting time was not as important (Long et al., 2016; Merlino, 2016; Moore, Hamilton, Krusel, Moore, & Pierre-Louis, 2016). Putting patients first through improved interpersonal quality of care was also echoed as the most important driver for improving patient satisfaction, even by the President and Chief Transformation Officer of Press Ganey (Merlino & Raman, 2013).

As the need to address the emotional and psychological sides of patients are growing in importance (Dixon, Freeman, & Toman, 2010), the emotional and psychological support provided by courteous staff members or an attentive physician whose approach to health care addresses patient's individual concerns and care may be enough to overcome the dissatisfaction of waiting time. In other words, physicians who display high interpersonal quality will likely engage in activities associated with managing waiting time to alleviate anxiety and fear, minimize uncertainty, explain, and ensure equity while patients are waiting all of which are known to effectively manage waiting time for improving service provider satisfaction (Maister, 1985). Therefore, we believe the availability and accessibility of physicians and other care services could generate an improvement in satisfying the needs of the patients especially given that waiting time is becoming relatively less important in influencing patient satisfaction. We hypothesize:

H2b A physician's interpersonal quality positively moderates the relationship between waiting time and patient satisfaction.

2.4 | Moderating effects of quality importance to patients

The relationship between waiting time and patient satisfaction could also depend on patients' perspectives about what is important to them. In keeping with prior health-care studies (Dagger et al., 2007), we posit that the relationship is moderated by the importance of technical and interpersonal qualities, herein referred to as *technical quality importance* and *interpersonal quality importance*, respectively (see Figure 1).

Operations and service management literature have shown that the value of time differs according to individual dispositions. In a study examining queues arising in remote service systems (e.g., call centers), Zohar, Mandelbaum, and Shimkin (2002) observe patience, defined as value of time, varies significantly depending on customer priorities. Fung et al. (2005) find that patients place different priorities between technical and interpersonal qualities—when forced to make choices or trade-offs, some patients show a preference for physicians with high technical quality, while others preferred high interpersonal quality. These different preferences are akin to heterogeneous customer needs found in marketing research and suggest that patients' heterogeneous priorities and values can affect patient satisfaction (Cleary & McNeil, 1988). In particular, goal orientation, anticipatory, and information economic theories suggest that the importance of technical and interpersonal qualities to patients can moderate the effects of waiting time on patient satisfaction.

The premise underlying goal orientation theory is that waiting time experience is a function of the subjective importance of a goal (Meyer, 1994). In other words, the time spent waiting can be viewed as a “positive investment” (a reward) to attain the goal or as a “wasteful expense” (a sunk cost) depending on a subject's level of interest in the goal. Therefore, subjects with a high level of interest in the goal (high goal orientation) view waiting as a positive investment to attain the goal suggesting that they are less concerned with the passing of time itself. In fact, Meyer's (1994) experiments have shown that subjects with low level of interest in the goal (low goal orientation) were negatively impacted by waiting time. Since goal orientation is an individual disposition and technical quality is evaluated based on the outcome of a service, or the goal of the encounter (i.e., "what" service is delivered) (Sharma & Patterson, 1999), it is plausible that high goal-oriented patients will more likely perceive technical quality as highly important. Accordingly, it is likely that high technical quality importance positively moderates the effect of waiting time on patient satisfaction, such that the negative effect brought on by waiting time is less pronounced for patients with high technical quality importance.

Anticipatory and information economic theories offer complementary theoretical lenses for understanding waiting time. Anticipatory theory suggests that a subject's attention to the passing of time is amplified when awaiting an outcome that appears imminent (Hui, Thakor, & Gill, 1998). In other words, when a subject who is close to the natural ending of a process experiences a delay, it draws more anticipatory attention (and thus a negative effect of waiting) since a delay is seen as an interruption to the process. Information economic theory recognizes technical quality and interpersonal quality as different types of attributes (Dagger & Sweeney, 2007). Technical quality is classified as a credence attribute (difficult to evaluate), while interpersonal quality is classified as an experience attribute (easy to evaluate). It is well documented that experience attributes can be evaluated immediately after consumption, unlike credence attributes, which can only be evaluated after an extensive product or service usage, if at all (Maute & Forrester Jr, 1991). Therefore, for patients who value interpersonal quality, the interpersonal process and interaction with a physician are akin to the natural ending of an encounter. Since the evaluation of the interpersonal quality is immediate and imminent following a wait, it is likely that the waiting time will have a stronger effect on patient satisfaction. For patients who value technical quality, the natural ending of a process (e.g., healed) is farther apart from the actual waiting (i.e., distance with respect to time), suggesting that the effect of waiting time on patient satisfaction is likely weaker. Therefore, we hypothesize:

H3a The importance of technical quality to patients positively moderates the relationship between waiting time and patient satisfaction.

H3b The importance of interpersonal quality to patients negatively moderates the relationship between waiting time and patient satisfaction.

3 | DATA AND METHODOLOGY

3.1 | Context and data

We collected physician review data from Vitals (http://www.vitals.com) in August 2015, using a web crawler. Vitals is a leading online physician review website in the United States. Its popularity stems from its management practices that ensure its data is reliable and meaningful, including a minimum requirement that a physician must be board-certified, a review process for "posting" or "removing" content based on authenticity and appropriateness, and limiting to one review submission over a 30-day period. Another practice, anonymous reviews, not only protects the privacy of its reviewers but also allows more meaningful and candid reviews.

Our data set represents a more comprehensive sample compared with prior studies and allows us to carry out a large-scale inference on the impact of operational efficiency. It contains 1,560,639 reviews, covering all 50 states of the United States and spanning 60 medical specialties. To reduce the number of fixed effects, we map the medical specialties to 17 major ones as defined by U.S. News & World Report after consultation with physicians.

Each physician review consists of information on the overall rating of the physician, physician ratings, a textual comment, and reported waiting time. Physician information includes a physician's name, address, specialty, and years of experience. Rating information includes the overall star rating and six aspect ratings about a physician based on a 7-point Likert-scale from 1 to 4, in 0.5 step increments.[1] We use the overall star rating to measure patient satisfaction. The patient-reported average *Waiting Time* is a unique measure on Vitals that enables us to directly assess the impact of operational efficiency on patient satisfaction (Froehle & Magazine, 2013). The six aspect ratings we collected are *Accurate Diagnosis, Courteous Staff, Bedside Manner, Spends Time with Me, Ease of Appointment,* and *Postvisit Follow-Up.* We map these aspect ratings to Dagger et al.'s (2007) health-care quality model. Specifically, we use accurate diagnosis to measure technical quality (Rao, Clarke, Sanderson, & Hammersley, 2006). We use courteous staff (Stewart et al., 2007), bedside manner (Dagger et al., 2007), and spends time with me (Haas-Wilson, 1994) to measure interpersonal quality. Finally, we use ease of appointments (Thomas, Glynne-Jones, & Chait, 1997) and postvisit follow-up (Beadles et al., 2015) to measure administrative quality, which is defined as "elements (that) facilitate the production of a core service, while adding value to a customer's use of the service." These variables and their corresponding concepts are also described in Table 3 in Section 3.2.

We aggregate the data to physician level by taking the average of all numeric information. This step is necessary

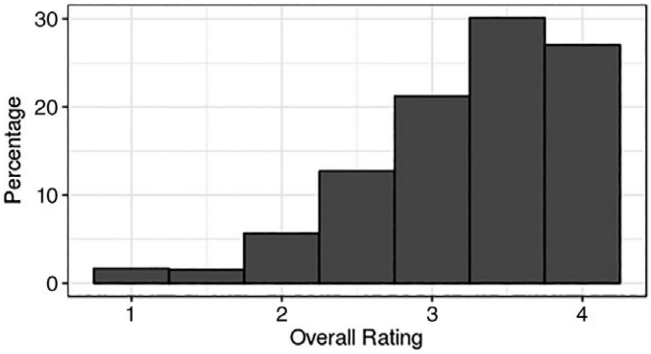

FIGURE 2 Histogram of physician's overall ratings

since *Waiting Time* is available only at the physician level as an averaged number. The average aspect ratings and overall rating are also displayed on a physician's profile page. After aggregation, 242,319 physicians have non-missing values on patient satisfaction, six aspect ratings, waiting time, and textual reviews. We proceed with our analyses with this subset of data and later demonstrate that selection bias is an unlikely issue.

Figure 2 depicts the patient satisfaction frequency chart revealing a J-shaped distribution commonly reported in previous studies examining online reviews. It further shows that about 25% of the patients gave a rating of 4 (the highest rating in Vitals' scale). Tables 1 and 2 present the summary statistics and the correlation matrix, respectively.

3.2 | Method

3.2.1 | Overview

To quantify the effect of operational efficiency on patient satisfaction, we analyze the physician reviews data set by estimating a set of regression models. We first assess the impact of waiting time on patient satisfaction (H1). We then test the moderating role of physicians' quality ratings on the effect of waiting time (H2). In our post hoc analyses, we also examine disparities of the negative impacts of high waiting time across medical specialties—a question for which the answer is unclear ex ante and which will lead to valuable managerial insights.

For H3, we focus on the qualitative part of the physician reviews: the textual comments. While the aspect ratings and the average waiting time are measures supplied by the platform for evaluating patient satisfaction, they do not convey what a patient believes to be important. Textual reviews, on the other hand, provide a less restrictive channel for patients to address the quality dimensions that they are most concerned about. We employ a novel computational linguistic method known as topic modeling, or latent Dirichlet allocation (LDA), to automatically extract the main topics in the reviews. While the numerical ratings reflect the performance of the physicians, the results from the topic model allow us to measure the importance of each quality dimension to the patients using the intensity of the conversation on each topic. We test whether the effect of waiting time on patient satisfaction depends on the importance of technical and interpersonal quality dimensions (H3). We next describe our regression specifications and details on topic modeling.

3.2.2 | Regression models

The dependent variable in our regression is the average overall rating for a physician. Since the distribution of overall ratings follows a J-shaped distribution, we use a log transformation to make the distribution more symmetric. Our main specification for the baseline model is a censored (tobit) regression model. We chose tobit because the standard OLS regression is inconsistent and biased when

TABLE 1 Summary statistics

	Median	Mean	SD	Minimum	Maximum
Years of experience	22	22.72	10.33	1	50
Overall rating	3.5	3.24	0.69	1	4
Number of reviews	3	4.45	6.20	1	351
Average waiting time (min)	17	20.36	12.68	5	60
Ease of appointment	3.5	3.32	0.66	1	4
Courteous staff	3.5	3.30	0.70	1	4
Accurate diagnosis	3.5	3.30	0.75	1	4
Bedside manner	3.5	3.27	0.79	1	4
Spends time with me	3.5	3.26	0.78	1	4
Postvisit follow-up	3.5	3.20	0.82	1	4
Population (zip code-level)	13,700	14,459	7,600	110	51,970
Median household income ($1,000, zip code-level)	60.41	78.55	62.96	17.40	1,149.29

TABLE 2 Correlation matrix

		VIF	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Years of experience	(1)	1.02													
Overall rating	(2)	-	−0.10*												
Number of reviews	(3)	1.03	−0.03*	−0.06*											
Average waiting time (min)	(4)	1.13	0.09*	−0.40*	0.05*										
Ease of appointment	(5)	2.14	−0.07*	0.60*	−0.03*	−0.45*									
Courteous staff	(6)	2.85	−0.08*	0.68*	−0.06*	−0.47*	0.71*								
Accurate diagnosis	(7)	6.62	−0.09*	0.82*	−0.04*	−0.42*	0.66*	0.74*							
Bedside manner	(8)	9.38	−0.11*	0.83*	−0.06*	−0.41*	0.65*	0.73*	0.90*						
Spends time with me	(9)	9.52	−0.10*	0.82*	−0.06*	−0.44*	0.66*	0.74*	0.90*	0.94*					
Postvisit follow-up	(10)	5.59	−0.09*	0.79*	−0.05*	−0.44*	0.67*	0.75*	0.87*	0.88*	0.89*				
Income	(11)	1.02	0.01**	0.02*	0.08*	−0.07*	0.03*	0.02*	0.03*	0.02*	0.03*	0.02*			
Population	(12)	1.01	0.01*	−0.02*	0.05*	0.02*	−0.01*	−0.03*	−0.01*	−0.02*	−0.02*	−0.02*	−0.04*		
Technical quality topic proportion	(13)	1.50	−0.00	0.04*	0.11*	−0.04*	0.08*	0.09*	0.06*	0.04*	0.03*	0.06*	0.06*	−0.01*	
Interpersonal quality topic proportion	(14)	1.91	−0.09*	0.31*	−0.08*	−0.19*	0.23*	0.28*	0.29*	0.30*	0.31*	0.30*	0.01**	−0.01*	−0.28*

Note. ***$p < .1$; **$p < .05$; *$p < .01$.

TABLE 3 Summary of concepts and measures

Conceptualization	Operationalization and literature	Perspective
Patient satisfaction	Overall rating (Ancarani et al., 2011)	Patient's judgment of the overall experience
Operational efficiency	Waiting time (Froehle & Magazine, 2013)	Reported waiting time between check-in and seeing a physician
Technical quality	Accurate diagnosis (Rao et al., 2006)	Evaluation of physicians' competence or outcome
Interpersonal quality	Courteous staff (Stewart et al., 2007) Bedside manners (Dagger et al., 2007) Spends time with me (Haas-Wilson, 1994)	Evaluation of physicians' competence or outcome
Administrative quality	Ease of appointments (Thomas et al., 1997) Postvisit follow-up (Beadles et al., 2015)	Evaluation of physicians' competence or outcome
Technical quality importance	Topic proportion of technical quality in reviews (Tirunillai & Tellis, 2014)	Importance of quality to patients
Interpersonal quality importance	Topic proportion of interpersonal quality in reviews (Tirunillai & Tellis, 2014)	Importance of quality to patients

a large proportion of the observed dependent variable equals to the lower (one-star rating) or upper (four-star rating) bound (Greene, 2012). The tobit model is commonly used in other studies with a censored response such as online ratings (Gao et al., 2015; Johnson, Klassen, Leenders, & Awaysheh, 2007).

We also consider two alternative specifications. The first is a logistic regression, in which the dependent variable is binary, denoting whether an average overall rating for a physician is high (greater than or equal to 3.5) or low (smaller than 3.5). The reason we use a dichotomous dependent variable is that individuals oftentimes simplify complex information into discrete categories (Gutman, 1982). For example, a customer tends to judge that the product with an above average rating to be “good” versus one with a below average rating to be “bad”. In fact, Fisher, Newman, and Dhar (2018) propose and find that customers tend to have a binary thinking (i.e., treating continuous data as dichotomous) regardless of certain graphical displays, modes of displays, and purchase decisions. The logistic regression model allows the interpretation of coefficients using the odds ratio—an interpretation that is robust to such “binary bias” when individuals evaluate online ratings (Fisher et al., 2018). That is, people tend to make a dichotomous distinction between positive ratings (e.g., 4 stars) and negative ratings (e.g., 1 and 2 stars), but they do not sufficiently distinguish between the most extreme (e.g., 4 stars) and less extreme values (e.g., 3.5 stars). To further mitigate the concern that the logit method may lose information that is in the original dependent variable, we adopt an ordered logit regression as our second alternative specification. Specifically, we place the physicians into six ordered bins based on their overall ratings (1–4 stars in 0.5 step increments). The ordered logit regression is more robust because the overall rating is an ordinal variable, and the transformation of the dependent variable loses much less information compared with the logit model.

Our key independent variable of interest is the average waiting time reported by patients. Since waiting time is a right-skewed duration variable, we perform a log transformation on it in our main model. In alternative specifications, we use a dichotomous indicator to show whether the average waiting time for a physician is greater than or less than the median (17 min). Such dichotomization provides an intuitive interpretation; it is also robust to the right censoring caused by the design of Vitals website that the maximum waiting time one can report is 60 min.[2]

Our baseline model includes six physician-level aspect ratings reported by patients: Accurate Diagnosis, Courteous Staff, Bedside Manner, Spends Time with Me, Ease of Appointment, and Postvisit Follow-up. We also use several control variables including the physician's years of experience (*Years_of_exp*) and the number of reviews (*N_reviews*) for each physician. Lastly, we control the population and mean income of each physician's zip code from the Statistics of Income data provided by the Internal Revenue Service.

To summarize, we run the following regression model to test H1:

$$\begin{aligned} Log\,(Overall\_Rating_i) = \beta_0 + \beta_1 Log\,(Waiting\_Time_i) + \gamma\, Technical\_Quality\_Ratings_i \\ + \delta\, Interpersonal\_Quality\_Ratings_i + \eta\, Administrative\_Quality\_Ratings_i + \rho\Gamma_i + \mu_i, \end{aligned} \tag{1}$$

where *Log(Overall_Rating$_i$)* is the natural log of the overall rating score for physician *i*; *Technical_Quality_Ratings$_i$* is measured by the aspect rating Accurate Diagnosis; *Interpersonal_Quality_Ratings$_i$* is a vector of aspect ratings including Courteous Staff, Bedside Manner, and Spends Time with Me; *Administrative_Quality_Ratings $_i$* is a vector of two aspect ratings: Ease of Appointment and Postvisit Follow-up; and *Γ* represents the vector of control variables, including number of reviews, years of experience, income, population, and medical specialties (fixed effects). In our alternative specifications using a logit model, we replace the dependent variable with *Logit[P(PositiveRating$_i$| X$_i$)]*, where PositiveRating$_i$ denotes the average overall rating for physician *i* greater than or equal to 3.5. In the ordered logit model, the dependent variable is *Logit[P(Overall_Rating$_i$ s |X$_i$)]*, where $s = 1, 1.5, \ldots, 4$.

To test H2, we form interaction terms between Waiting Time and each of *Technical_Quality_Ratings*, *Interpersonal_Quality_Ratings*, and *Administrative_Quality_Ratings*. We then add them as additional independent variables to access the moderation effects of these quality dimensions on the impact of waiting time on patient satisfaction. Model (2) below is our estimation form:

$$\begin{aligned} Log\left(Overall\_Rating_i\right) = {} & \beta_0 + \beta_1 Log\left(Waiting\_Time_i\right) + \gamma\, Technical\_Quality\_Ratings_i \\ & + \delta\, Interpersonal\_Quality\_Ratings_i + \eta\, Administrative\_Quality\_Ratings_i \\ & + \beta_2 Log\left(Waiting\_Time_i\right) \times Technical\_Quality\_Ratings_i \\ & + \beta_3 Log\left(Waiting\_Time_i\right) \times Interpersonal\_Quality\_Ratings_i \\ & + \beta_4 Log\left(Waiting\_Time_i\right) \times Administrative\_Quality\_Ratings_i + \rho\Gamma_i + \mu_i. \end{aligned} \tag{2}$$

To test H3, we first use topic modeling (Blei, 2012) to analyze the textual reviews and measure the importance of technical/interpersonal quality to patients using the topic proportion of technical quality/interpersonal quality in text reviews (Tirunillai & Tellis, 2014). The topic proportion of, for example, technical quality for a physician is the proportion of reviews of this physician devoted to the discussion of technical quality-related issues. We offer a detailed explanation of topic modeling in the following subsection (Section 3.2.3). To test the moderating effect of technical quality importance on the relationship between waiting time and patient satisfaction, we include the interaction terms between the topic proportion of technical/interpersonal quality and waiting time in the regression. The identification comes from the fact that each physician has a unique mixture of patients that vary collectively in their needs or dispositional characteristics. Whether such patient heterogeneity is due to random assignments or individual choices has little bearing on the hypothesis since our main concern is not the causal impact of quality importance.

There may be two reasons for patients to discuss more about a certain topic, for example, technical quality, in their reviews. First, it is possible that the physician is particularly good or bad at diagnosing conditions accurately, which elicit more comments on technical quality. Second, patients' own characteristics such as high goal-orientation would cause them to value more about technical quality and thus make them more likely to comment on it in their reviews. In order to tease out the physicians' impact on this measure and just focus on patients' heterogeneity, we included the physicians' quality ratings as control variables in our regression analysis. Therefore, we have the following model to test H3a and H3b:

$$\begin{aligned} Log\left(Overall\_Rating_i\right) = {} & \beta_0 + \beta_1 Log\left(Waiting\_Time_i\right) + \gamma\, Technical\_Quality\_Ratings_i \\ & + \delta\, Interpersonal\_Quality\_Ratings_i + \eta\, Administrative\_Quality\_Ratings_i \\ & + \beta_2 Log\left(Waiting\_Time_i\right) \times Technical\_Quality\_Importance_i \\ & + \beta_3 Log\left(Waiting\_Time_i\right) \times Interpersonal\_Quality\_Importance_i + \rho\Gamma_i + \mu_i. \end{aligned} \tag{3}$$

The coefficient β_2 (β_3) measures the moderating effect of technical (interpersonal) quality importance on the impact of waiting time on patient satisfaction, independent of physicians' technical, interpersonal, and administrative qualities ratings.

Given that the variation in physician ratings is influenced by heterogeneity across physicians, patients, treatments, locations, and so forth, our data set could exhibit heteroscedasticity, which is a common challenge to cross-sectional analyses. Adopting a Breusch-Pagan/Cook-Weisberg test, we reject the null hypothesis thatthere is no linear heteroscedasticity in our data (Chi-square = 17,568.26, p-value<0.0001). In addition, it is likely that physician ratings are correlated for physicians within the same geographic location (e.g., certain areas can attract high-performing physicians). We therefore control for heteroscedasticity and adjust for zip code level clustering when estimating the SEs. Finally, we construct the variance inflation factors (VIF) for our independent variables to detect whether there is a multicollinearity issue. The VIFs are reported in Table 2. The two variables with the highest VIFs are Bedside Manner (9.38) and Spends Time with Me (9.52). The VIF of these two variables should not be a concern because they are smaller than the commonly used VIF < 10 rule of thumb. In addition, multicollinearity does not bias the estimates, it only makes them less efficient. For example, a VIF = 9 means that the SE for the coefficient would be three times as large as it would be if its VIF was 1. The dummy variables for physicians' medical specialties have VIFs ranging from 1.12 to 12.99 with a mean 2.91; our results are robust without the medical specialty fixed effects. Given our main variable waiting time has a small VIF and is statistically significant, multicollinearity should not be an issue for our estimation.

3.2.3 | Topic model

Recall that in H3, we hypothesize that the effect of waiting time on satisfaction depends on the importance of technical and interpersonal qualities to the patient. To test these hypotheses, we use a topic model to extract the importance of these dimensions to patients from the textual comments (Tirunillai & Tellis, 2014). Topic models are machine learning algorithms for discovering the main themes that pervade a large collection of documents (Blei, 2012). In our analysis, the topic model helps us achieve three goals: (a) confirming that both technical quality and interpersonal quality, among other topics, are patients' major concerns expressed in the textual reviews; (b) computing the topic proportion (i.e., the percentage weight of each topic) in the reviews and using them to measure the relative importance of each dimension (i.e., how much patients value each dimension); and (c) exploring other operations-related factors that patients are either not able to express their views through aspect ratings or feel the need to elaborate on the ratings.

Intuitively, fitting topic models on textual reviews can be thought of as reversing the process by which a review is written. The process of a patient leaving a review begins with one or more topics (quality dimensions) that the patient is concerned about, then deciding how much to focus on each of the topics, and finally selecting words appropriate to discussing those topics. Topic model reverses that process by picking out clusters of words used frequently together in a review and determining the frequency with which those words occur. We can then use that data to suggest which topics those word clusters would be best suited to discuss. Stated differently, the model infers a probabilistic distribution of word combinations that is most likely to generate the observed reviews. The parameters of the fitted probabilistic distribution inform us the importance of each topic and the keywords associated with each topic. The specific method of topic modeling used here is known as latent Dirichlet allocation (LDA).

We make two adjustments to the LDA algorithm. First, because we are interested in measuring the importance of each dimension rather than their polarity, we remove all the sentiment-related words in the MPQA Subjectivity Lexicon from the textural reviews.[3] The lexicon contains 7,652 common words that express positive or negative opinions, emotions, and evaluations. This ensures that the algorithm converges to topics of objective quality dimensions rather than polarized opinions. Second, based on our theoretical framework, we guide the algorithm so that the first three topics measure the importance of Technical Quality, Interpersonal Quality, and Timeliness. Because LDA is essentially a Bayesian probabilistic model, we can guide the algorithm by assigning a higher prior probability for a set of seed words to appear together in the same topic (Jagarlamudi, Daumé III, & Udupa, 2012). We generate a set of seed words using Dagger et al.'s (2007) definitions and scales (see Section 4.3). One can think of the guided LDA as a confirmatory factor analysis, whereas a traditional LDA resembles an exploratory factor analysis. In the next section, we present multiple validation checks on the outputs of the algorithm. We refer interested readers to Appendix 1 for a more technical description of topic modeling.

It is worth noting that there are several advantages provided by LDA compared with two other text analytic methods—Latent Semantic Analysis (LSA) and supervised classification—both of which have been recently introduced in OM literature. The first alternative, LSA, has been used by Kulkarni, Apte, and Evangelopoulos (2014) to identify the latent topics in OM journals. LDA is more appropriate in our study because it allows each review to be a multi-membership mixture of different topics, while LSA restricts one review to be about only one topic. A single physician review often addresses different aspects of an encounter. LDA provides "soft" classification of topics for reviews and hence produces more realistic outputs to address this possibility. In addition, LDA is built upon a foundation of Bayesian statistical inference and therefore has a principled model fitting and selection procedure. It also allows us to incorporate prior knowledge, such as dimensions from the literature, in the model fitting process. Second, LDA is more suitable compared with supervised text classification methods such as naïve Bayes or support vector machine, which Chan et al. (2016) used to identify product defects from social media data. In particular, LDA does not rely on human-labeled data to train the classifiers. If a quality dimension from prior literature is a relevant concern raised by patients, the topic and its associated keywords would emerge from the data. On the other hand, if an important topic was overlooked by prior models, LDA will detect it by learning from the massive volume of textual reviews. To conclude this section, Table 3 provides a summary of the concepts and measures we used to test the hypotheses.

4 | ANALYSES AND RESULTS

4.1 | Baseline results

Our baseline results on patient satisfaction are presented in Table 4. Model 1 presents the results from tobit regression. Models 2 and 3 present the results from the logit regression model and the ordered logit model, respectively.

Overall, our results from the three models are highly consistent: longer waiting time is negatively associated with lower physician ratings and the relationship is statistically significant. The effect is also economically significant. For example, the estimate from the tobit model indicates that, if waiting time increases from 5 min to 30 min (a waiting time that should not cause dissatisfaction according to Huang (1994)), a physician's overall rating will drop by 0.26.[4] Estimate from the logit model indicates that waiting time longer than 17 min will, on average, reduce the odds of getting a high rating status by 14%.[5] Using the ordered logit model, Model 3 shows that, when the waiting time is longer than 17 min, the odds of getting a 0.5 increase in patient satisfaction are 0.84 times smaller, holding all other variables constant. In sum, we find consistent support for H1. We also find that technical quality, interpersonal quality, and administrative quality are highly significant contributors to patient satisfaction. Among all aspect ratings, Accurate Diagnosis

TABLE 4 Analysis of operational efficiency on patient satisfaction

	Model 1 (tobit) Log (overall rating)		Model 2 (logit) Overall rating ≥ 3.5		Model 3 (ordered logit) Overall rating	
Variables	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Operational efficiency						
Log (waiting time)	−0.013***	0.001	-	-	-	-
Waiting time > 17 min	-	-	−0.153***	0.013	−0.169***	0.009
Technical quality						
Accurate diagnosis	0.103***	0.002	1.023***	0.021	1.165***	0.017
Interpersonal quality						
Courteous staff	0.021***	0.001	0.312***	0.017	0.281***	0.011
Bedside manner	0.124***	0.002	1.315***	0.026	1.517***	0.020
Spends time with me	0.053***	0.002	0.669***	0.027	0.700***	0.020
Administrative quality						
Ease of appointment	0.012***	0.001	0.126***	0.016	0.096***	0.010
Postvisit follow-up	0.035***	0.001	0.377***	0.018	0.483***	0.013
Controls						
Number of reviews	−0.000***	0.000	0.001	0.001	−0.011***	0.001
Years of experience	−0.000***	0.000	−0.006***	0.001	−0.003***	0.000
Log (income)	−0.002*	0.001	0.018	0.012	−0.019*	0.008
Log (population)	−0.001	0.001	−0.004	0.011	−0.023***	0.007
Medical specialty FE	Yes		Yes		Yes	
Constant	0.177***	0.001	−12.190***	0.150	-	-
N	242,319		242,319		242,319	
(pseudo) R^2	0.428		0.483		0.371	

Note. Robust SEs adjusting for heteroskedasticity and zipcode-level clustering are used. ***$p < .001$; **$p < .01$; *$p < .05$.

and Bedside Manner have the strongest influence on patient satisfaction.

4.2 | Moderating effects of physician's quality dimensions

To test the moderating effects of a physician's technical and interpersonal quality dimensions (H2), we extend our baseline model by introducing the interaction terms (Model 4 in Table 5). Contrary to what we hypothesize in H2a and H2b, we find that technical quality and interpersonal quality are not significant moderators on the impact of waiting time. The results suggest technical and interpersonal qualities neither alleviate nor worsen the negative impact of waiting. In other words, how patients feel about having to wait longer is independent of how well the physicians treat them during their visit, both technically and interpersonally.

In addition, we find that the negative impact of waiting time is mitigated by two ratings related to administrative quality: Ease of Appointment (β = .004, $p < .01$) and Postvisit Follow-up (β = .004, $p < .001$). This means that making it easier for patients to make appointments as well as following up with them after their visits could mitigate their frustration with waiting time and thus alleviate their overall dissatisfaction with their health-care service providers. This particular result suggests that, in order to alleviate patients' negative feelings about longer waiting time, the key is to improve the administrative quality—especially the previsit administrative quality (e.g., Ease of Appointment) and postvisit administrative quality (e.g., Postvisit Follow-up).

4.3 | Moderating effects of quality importance to patients

H3 states that the effect of waiting time depends on the importance of quality dimensions to the patients. We use topic modeling to extract the main quality dimensions from the textual reviews and measure the importance of those dimensions to patients. To test whether technical quality importance and interpersonal quality importance lead to heterogeneous effects of waiting time, we guide the LDA algorithm using the following seed words from Dagger et al.'s (2007) qualitative study and scales.

- *Technical Quality*: outcome, expertise, treatment, result, qualified, competence, diagnose, skill, knowledge
- *Interpersonal Quality*: interaction, listen, communication, understand, explain, answer, relationship, empathetic, caring

We also include a *Timeliness* dimension in Dagger et al.'s (2007) model, which describes patients' perception of timeliness and operational efficiency.[6] The seed words are:

- *Timeliness*: waiting, appointment, hours, organized, efficient, reschedule

The guidance provided by seed words is quite flexible because they can help gather other words that are related to these words into a topic. Further, the seed words are not hard constraints. The model still respects the data by converging

TABLE 5 Moderating effects of physicians' quality ratings

	Model 4 (tobit) Log (overall rating)	
Variables	**Coefficient**	**SE**
Operational efficiency		
log (waiting time)	−0.028***	0.003
Technical quality		
Accurate diagnosis	0.105***	0.002
Interpersonal quality		
Courteous staff	0.023***	0.002
Bedside manner	0.123***	0.003
Spends time with me	0.053***	0.003
Administrative quality		
Ease of appointment	0.008***	0.001
Postvisit follow-up	0.023***	0.002
Moderating effects of physicians' quality ratings		
Accurate diagnosis × log (waiting time)	−0.003	0.002
Courteous staff × log (waiting time)	0.002	0.001
Bedside manner × log (waiting time)	0.000	0.002
Spends time with me × log (waiting time)	0.000	0.002
Ease of appointment × log (waiting time)	0.004**	0.001
Postvisit follow-up × log (waiting time)	0.004***	0.001
Controls		
Number of reviews	0.000***	0.000
Years of experience	−0.000***	0.000
Log (income)	0.001*	0.000
Log (population)	0.001*	0.000
Medical specialty FE	Yes	
Constant	0.165***	0.012
N	242,319	
(pseudo) R^2	0.488	

Note. Robust SEs adjusting for heteroskedasticity and zipcode-level clustering are used. ***$p < .001$; **$p < .01$; *$p < .05$.

TABLE 6 Topics from textual reviews

Topics	Topic proportion	Keywords	Examples reviews
Technical quality	12.1%	Perform, treatment, recovery, explain, fix, hospital, diagnose, option	(1 Star) I would not trust this physician's ability to diagnose properly...Please seek a more qualified and knowledgeable physician...I do not recommend this physician (2 Star) Dr. Busconi did scopes on both my hips. The right hip was better afterwards but the left hip was not. He had me go to pt and get a cortisone injection. Both made the pain my left hip worse. After that, the disinterested Dr. Busconi said there was nothing he could do for me and nothing any other ortho could do for me either. His policy is not to do more than one scope per hip. I went to New England Baptist to find out I have dysplasia in my left hip, along with a giant labral tear. And there is something that can be done about both conditions. (3 Star) Dr. Lasner impressed me with his total knowledge of the disk herniation operation. Although not the most personable man, his expertise was what I wanted. I can get a joke from my PCP any time! I have high hopes of a full recovery and have total confidence in Dr. Lasner. (4 Star) In my opinion, Dr. Richardson is the best spine doctor there is. My spine was full of arthritis and had stenosed my nerve so bad that I could hardly walk. He did a spinal fusion with a pin and screws. I feel I have been transformed now. My back, hips, and legs feel as good as they did when I was 20. I can say this doctor knows his stuff and I recommend him highly for back surgery.
Interpersonal quality	34.4%	Caring, listen, explain, bedside manner, attitude, spend, answer question, talk	(1 Star) Dr. Guerra is repulsive doctor with no bedside manner He is rude and refuses to talk to patients and becomes belligerent when asked questions. DO NOT USE THIS DOCTOR!!!!!!!!!! (2 Star) This doctor is ok but not the best. Needs to listen better and be more compassionate. (3 Star) Dr Shah always spends an adequate amount of time with us, she listens intently and answers all of our questions, we never feel intimidated by her, we feel we can tell her our issues. The problems we have experienced have been with the facility and other staff, not Dr Shah. (4 Star) Amazing doctor, very knowledgeable, understanding, that can listen to patient's problems without rushing you.
Timeliness	18.2%	Wait, appointment, minute, hour, leave, nurse, schedule, room	(1 Star) I cannot stand calling the doctor's office. Everytime I call I get put on hold for at least 30 min. Everytime I go in for an appointment I end up waiting over a hour from the time I was supposed to be seen. (2 Star) If you have the option of taking your child somewhere else, do it! I have waited up to 4 hr to be seen. They have given me appointments at 7 a.m. but their office opens at 8 a.m. Everyone I have spoken to that takes their kids there hate it, staff is rude, waiting time is too long, waiting room is crowded with only one restroom that you need a key for. (3 Star) Dr Swann is a great dr but the only complaint that I have is the wait time. There have been times that we have sat in his waiting room for over an hour and then when you get back to the room waiting almost another hour. It is one thing that makes you put off going to the dr. who has that kind of time to wait. (4 Star) Dr. Curtis is very efficient. She takes her time with you, she cares, listens & answers all your??'s She sympathizes with you, and will do everything she can to get you feeling better. She has

(Continues)

TABLE 6 (Continued)

Topics	Topic proportion	Keywords	Examples reviews
			flexible appts to work around your schedule. Sometimes the wait can be a little lengthy, but that's only because she's so good & you will notice all her patients are patient knowing you have the most caring & best Rhumetoidologist.
Cost and billing	15.9%	Test, pay, bill, money, practice, insurance, med, charge	(1 Star) Hi, I been to this doc and he will rip you off by ordering tests that are absurd and sends the blood work to out of network labs and rips you off with co-insurance. My sincere advice is STAY AWAY from this doc. In my opinion, he is a big scamster. (2 Star) My husband had him as his anesthesiologist. Bedsides manners great, but he did not take our insurance and we had to pay $1,000 out of our pocket. Thought when we paid up front that was what our charge was and it was not. Received a bill for over 900.00. (3 Star) I had no problem with the dr. It was supposed to be a routine colonoscopy but they coded it as a diagnostic. My insurance would have paid 100%. Be careful that you get everything in writing. They did not properly code this and I have a big bill to pay. (4 star) He was great. Not only was he caring and smart, he also worked out a great payment plan so that my expenses were lower. HIGHLY RECOMMEND.
Family members' experience	19.4%	Son, hospital, daughter, husband, caring, baby, mother, save life	(1 Star) No happy, because doctor no help me kid with ear. Still with ear troubles. (2 Star) She was very nice, however I felt very uncomfortable when every time she talked to me she stared at my husband. Then after my colonoscopy, she never came and talked to me about results. (3 Star) Dr. Pollack did surgery on my mother in March and he was great to her and made sure all went well with her after she got out of the hospital. We (the family), really think he's the best!!! (4 Star) I just heard Dr. Kadry may retire. Both my daughter and grand daughter were delivered by him and he did my hysterectomy. America needs more doctors like you! Health care has sure changed. Thank you!!

to topics that are semantically close to, but does not necessarily equate to, a simple mixture of the seed words.

Apart from the above three seeded topics, we allow the algorithm to find other dimensions that are not covered by prior health service quality research. Specifically, we experiment with topic models using different number of topics and provide the above set of seed words to the first three topics. We can then judge the quality of topic models using measures for model fit. We vary the number of topics from 3 to 20, which allows us to inspect the model performance with the number of "free" topics ranging from 0 to 17. We decide that the optimal number of topics is five based on the UMass coherence measures (Mimno, Wallach, Talley, Leenders, & McCallum, 2011). The UMass coherence measure is based on the insight that a good topic should be represented using a set of coherent words that are more likely to co-occur in the same review.[7] Benchmarking using a medical literature data set shows that the UMass coherence measure correlates strongly with National Institutes of Health (NIH) experts' judgment of topic quality (Mimno et al., 2011).

Table 6 displays the five quality dimensions emerged from the textual reviews using the LDA model. We assign labels to the dimensions according to the high probable keywords. The five quality dimensions are *Technical Quality*, *Interpersonal Quality*, *Timeliness*, *Cost and Billing*, and *Family Members' Experience*. We measure the importance of these extracted quality dimensions by their proportional weight of discussion in reviews. For example, an average physician has 34.4% of the reviews on *Interpersonal Quality*, and 15.9% of the reviews commenting on *Cost and Billing*-related issues. Notably, comments related to *Timeliness* comprise 18.2% of reviews. The last column of Table 6 also lists example reviews under each topic; our model finds these reviews have the corresponding topic's proportion

TABLE 7 Moderating effects of quality importance to patients

Variables	Model 5 (tobit) Log (overall rating) Coefficient	SE
Operational efficiency		
Log (waiting time)	−0.012***	0.001
Technical quality		
Accurate diagnosis	0.099***	0.002
Interpersonal quality		
Courteous staff	0.017***	0.001
Bedside manner	0.121***	0.002
Spends time with me	0.052***	0.002
Administrative quality		
Ease of appointment	0.010***	0.001
Postvisit follow-up	0.032***	0.001
Moderating effects of quality importance to patients		
Topic proportion (technical quality) × log (waiting time)	0.001*	0.001
Topic proportion (interpersonal quality) × log (waiting time)	−0.001**	0.001
Controls		
Number of reviews	0.000	0.000
Years of experience	−0.000***	0.000
Log (income)	−0.001	0.001
Log (population)	−0.000	0.001
Medical specialty FE	Yes	
Topic proportions	Yes	
Constant	0.166***	0.010
N	235,426	
(pseudo) R^2	0.481	

Note. Robust SEs adjusting for heteroskedasticity and zipcode-level clustering are used. ***$p < .001$; **$p < .01$; *$p < .05$.

greater than 90%. The example reviews demonstrate that our model can measure the importance of topics across the full spectrum of overall ratings.

We also validate the quality of the topic model using two methods other than the UMass coherence measure. First, since the topic model was fitted using an iterative Gibbs sampling algorithm, we train the model using 1,000 iterations and track the convergence of the algorithm by plotting the iteration count and the model likelihood. We conclude that the five-topic model properly converged after about 120 iterations. Second, we validate the results generated by the five-topic model using evaluations of independent human raters. In the evaluation task, two graduate students are given the definition of the dimensions from Dagger et al. (2007) and the seed words (for the seeded topics) or the high probable keywords (for the non-seeded topics). They are then shown 250 reviews, each with two topic labels. One of the topic labels is decided by our fitted model as the highest probability topic; the other label is chosen randomly from the other four topics. The raters are asked to pick the topic that is most relevant to the review. We find that the inter-rater reliability, measured using Cohen's Kappa, is 0.75. On average, human raters agree with the topic model results 82.6% of the time. This shows that our topic model aligns well with human judgment.

Recall that we use topic proportions to measure the importance of the quality dimensions to patients (Tirunillai & Tellis, 2014). To test H3, we estimate regression Equation (3) and present the results in Table 7. We find that technical quality importance and interpersonal quality importance both have a significant impact on the effect of waiting time on patient satisfaction: when technical quality is important to patients, they tend to be tolerant of longer waiting time ($\beta = .001$, $p < .05$); when interpersonal quality is important, patients are sensitive about having to wait longer and thus the negative impact of waiting time on satisfaction is stronger ($\beta = -.001$, $p < .01$). In sum, we find support for both H3a and H3b.

4.4 | The impact of waiting time across specialties

We now conduct a post hoc study to examine if the effect of waiting time is consistent across medical specialties. After all, if patients of different specialties do not value waiting time the same way, then physicians can adjust operational decisions such as overbooking policies and scheduling lead time

Figure 3 odds ratio of high overall rating when waiting time is high for 17 specialties accordingly. We apply regression model 3 for each of the 17 specialties. Figure 3 summarizes the odds ratio associated with waiting time for different specialties. The vertical dashed line indicates the baseline of odds ratio = 1, which means longer waiting time does not impact the conditional probability of a physician having a high overall rating. Two observations emerge from Figure 3. First, all specialties have a mean of odds ratio less than 1, and most have the 95% CI on the left of 1. This result suggests that longer waiting time is associated with a worse overall rating for all specialties, and the effect is statistically significant for most of the specialties. Second, as we expected, the magnitude of the effect fluctuates across specialties; overall ratings in rehabilitation, rheumatology, urology, pulmonology, and ENT suffer the most from longer waiting time. Therefore, to improve the overall patient satisfaction, these specialties should have a stronger motivation in devoting more resources to improving their operational efficiency. For example, those specialties can establish a sound business process to follow-up with patients after their visit.

4.5 | Endogeneity and robustness checks

4.5.1 | Endogeneity

One potential issue with our estimation is the endogeneity of our main variable of interest—waiting time. It is possible that patient satisfaction and waiting time are both driven by unobserved, individual-specific effects of physicians. For example, one may argue that physicians with better organization skills will have lower average waiting time and higher patient satisfaction rating. Although we have included control variables that are related to patient satisfaction in many ways and obtained consistent results, the threat of endogeneity remains due to unobserved omitted variables and measurement errors.

Moreover, there may be a reverse causality problem that may distort our results. Conceptually, patients may intentionally choose physicians with high online ratings, which may affect their average waiting time and ease of scheduling appointments with these physicians. In other words, overall ratings may reversely affect waiting time and ease of appointment.

We adopt two approaches to address these endogeneity concerns. The first is an instrumental variable approach. Specifically, we adopt four instrumental variables (IVs) that are correlated with the endogenous variable (i.e., waiting time or ease of appointment) and uncorrelated with the dependent variable (i.e., overall ratings). The first IV is constructed as the average waiting time for all physicians that belong to the same medical specialty within the same zip code. Physicians from the same medical specialty in the same zip code are highly likely to have similar management styles, customer profiles, and IT infrastructure. Therefore, their individual waiting times are expected to be correlated with the average waiting time. On the other hand, the average waiting time should only relate to a particular physician's overall rating through his own waiting time. Following the same logic, our second IV is the average score on ease of appointment for physicians in the same specialty within the same zip code, and we use it to instrument the easy of appointment variable.

The third IV represents the scarcity of physicians, which is the ratio of the number of physicians in the same medical specialty in the same zip code to the population in that zip code. We expect that the scarcity of physicians is related with waiting time for a particular geographic area, and it should not be directly related to patient satisfaction. The fourth IV is the road density (i.e., the total length of road/area) at the zip code level, assuming that the higher the road density, the more potential traffic delays for the patients and thus longer waiting times for other patients. Following the practice in Correia, Peters, Levy, Melly, and Dominici (2013), major roads are defined as limited access highways, primary roads without limited access, and secondary and connecting roads (Census Feature Class Code A1, A2 or A3). We calculate the total length of major roads within 200 m of census block centroids, and then integrate it at

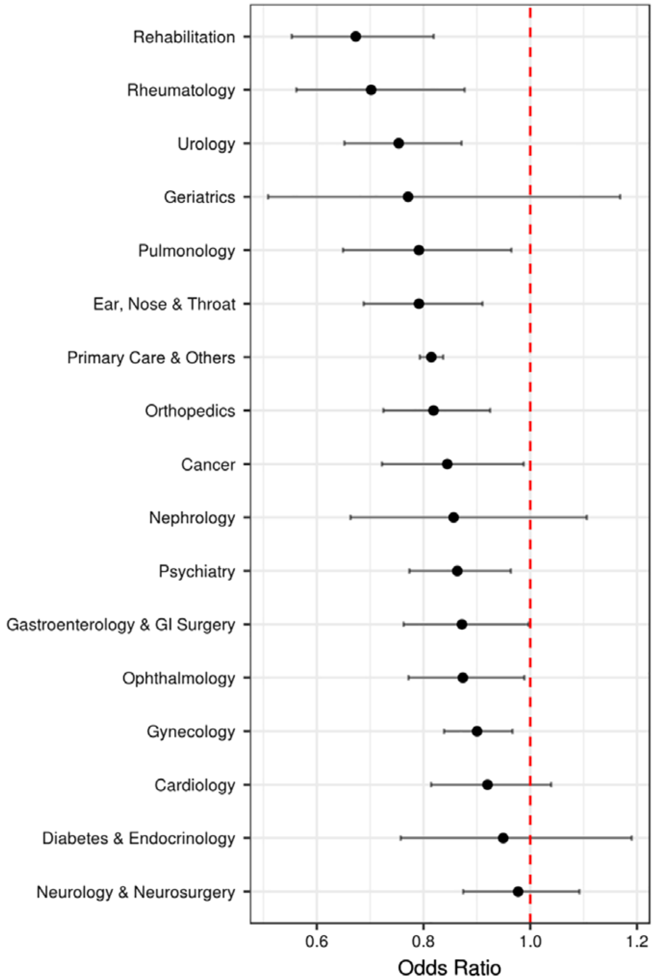

FIGURE 3 Odds ratio of high overall rating when waiting time is high for 17 specialties [Color figure can be viewed at wileyonlinelibrary.com]

the zip code level as the road density data. Both U.S. major road and census block data are obtained from ESRI ArcGIS U.S. Data set 2010.

We run a 2SLS regression with the four IVs and present the results in Table 8. As an additional robustness check, we run two probit regressions with the same four IVs—one with a dichotomous waiting time variable and the other with a continuous, log-transformed waiting time. Results using IVs are highly consistent with our main results.

Our instrumental variables need to be strong such that they account for enough variance in the endogenous variable. We conduct a first-stage F test, which can be used as a diagnostic for whether a particular endogenous regressor is "weakly identified" (Sanderson & Windmeijer, 2016). The F-statistic for the two endogenous variables, respectively, from the first stage of the 2SLS is 33,662 ($p < .001$) and 17,891 ($p < .001$), rejecting the null hypothesis that the excluded instruments are weakly associated with the endogenous variables.

Moreover, a precondition for the IV regressions is that the instrumental variables are indeed exogenous. That is,

TABLE 8 Results of instrumental variable regressions

Variables	2SLS Log (overall rating) Coefficient	SE	Probit with IV Overall rating ≥ 3.5 Coefficient	SE	Probit with IV Overall rating ≥ 3.5 Coefficient	SE
Operational efficiency						
Waiting time > 17 min	-	-	−0.124***	0.016	-	-
Log (waiting time)	−0.006***	0.001	-	-	−0.084***	0.011
Technical quality						
Accurate diagnosis	0.090***	0.001	0.551***	0.012	0.552***	0.012
Interpersonal quality						
Courteous staff	0.009***	0.001	0.152***	0.011	0.153***	0.011
Bedside manner	0.111***	0.001	0.726***	0.013	0.725***	0.013
Spends time with me	0.042***	0.001	0.357***	0.013	0.359***	0.013
Administrative quality						
Ease of appointment	0.021***	0.001	0.088***	0.017	0.086***	0.017
Postvisit follow-up	0.023***	0.001	0.202***	0.010	0.202***	0.010
Controls						
Number of reviews	0.001***	0.000	0.002**	0.000	0.002**	0.000
Years of experience	−0.000***	0.000	−0.003	0.000	−0.003***	0.000
Log (income)	0.001**	0.000	0.014*	0.006	0.014*	0.006
Log (population)	0.001****	0.000	0.000	0.005	0.000	0.005
Medical specialty FE	Yes		Yes		Yes	
Constant	0.173	0.012	−6.603***	0.135	−6.432***	0.140
N	242,319		242,319		242,319	

Note. This table presents the second-stage instrumental variable regression results where waiting time variables are instrumented by local road density, local physician density, and average waiting time of the same specialty within the same zipcode. Ease of Appointment is instrumented by the average Ease of appointment rating of the same specialty within the same zipcode. SEs clustered at zip code level are reported. ****$p < .1$; ***$p < .001$; **$p < .01$; *$p < .05$.

they are not correlated with the error terms in the model. The canonical test for instrument exogeneity in over-identified models is the Sargan test. We conduct a Sargan test for our model; the results indicate that we cannot reject the null hypothesis that our instrumental variables are exogenous ($\chi^2 = 1.108$, p-value = .253).

Our second approach to deal with the reverse causality issue is to restrict our sample to the physicians who have only one rating and re-run our analyses on the restricted sample. Because there is only one rating for these physicians, the patients who gave these ratings would not be able to see any prior ratings about their physicians from Vitals.com, and thus our estimations would suffer the least degree from the reverse causality problem. All of our qualitative results remain the same.

4.5.2 | Self-selection bias

A patient's choice of physicians may be related to her unobserved characteristics. For example, some patients may highly value physicians' quality, which could make these patients more tolerant of longer waiting time. Some other patients may highly value timeliness, and thus tend to be intolerant of longer waiting time and are more likely to give lower ratings. If this is true, these patients self-select into seeing their physicians and would give lower ratings for these physicians even if they did not experience a long waiting time, and thus makes it hard to identify the causal relationship between waiting time and physicians' overall ratings.

To mitigate the self-selection problem, we have adopted the following econometric adjustments and tests. First, we have controlled for a substantive set of factors that may be related to a patient's choice of physicians, including physician-related variables (e.g., medical specialty, number of reviews, years of experience, technical quality, interpersonal quality, and administrative quality) and geographic-based variables (e.g., zip code level income and population). Second, to the degree that the unobserved factors that determine a patient's choice of physicians are correlated with waiting time or ease of appointment, our 2SLS approach has shown that our main results are all consistent.

Third, we use four topic proportions extracted from the textual reviews (i.e., technical quality, interpersonal quality, timeliness, and cost and billing), and add them as patient-related control variables in our regressions. This allows us to control, for example, how much patients emphasize or value the technical quality of the health-care provider—which is directly related to the self-selection rule. We find that the results are highly consistent.

Fourth, we conduct a propensity score matching (PSM) analysis to check our results against possible bias caused by sample selection.[8] The goal of PSM is to construct a control group that has similar characteristics to the treatment group. We employ a single nearest neighbor matching using the *psmatch2* module in Stata. Specifically, we estimate a logit model where the dependent variable is binary, indicating whether a physician receives the treatment of high waiting time (greater than 17 min). The independent variables are the observable covariates of the physicians (aspect ratings and control variables). We then match each physician in the high waiting time group with a physician in the low waiting time group who has the smallest absolute difference in the propensity scores, which is the predicted probabilities from the logit model. We confirm that the procedure resulted in comparable treatment and control groups. For example, the differences in the aspect ratings between the two groups are significantly reduced, from 0.512 before matching (average of six dimensions) to 0.008 after matching. In other words, the physicians with long and short waiting times in the PSM matched sample have nearly identical characteristics. Table 9 reports results from the regression analysis using the matched sample. The coefficients are consistent with our main results in terms of both the magnitude and statistical significance, indicating that selection bias is unlikely a concern.

TABLE 9 Results from propensity score matched sample

	Matched tobit Log (overall rating)		Matched ordered logit Overall rating	
Variables	**Coefficient**	**SE**	**Coefficient**	**SE**
Operational efficiency (treatment)				
Waiting time > 17 min	−0.014***	0.001	−0.159***	0.012
Technical quality				
Accurate diagnosis	0.102***	0.003	1.042***	0.029
Interpersonal quality				
Courteous staff	0.019***	0.002	0.248***	0.021
Bedside manner	0.121***	0.003	1.420***	0.037
Spends time with me	0.019***	0.002	0.727***	0.032
Administrative quality				
Ease of appointment	0.013***	0.002	0.094***	0.016
Postvisit follow-up	0.035***	0.002	0.458***	0.021
Controls				
Number of reviews	0.000	0.000	−0.005***	0.001
Years of experience	−0.000***	0.000	−0.003***	0.001
Log (income)	−0.001	0.001	−0.023****	0.012
Log (population)	−0.001	0.001	−0.025*	0.011
Medical specialty FE	Yes		Yes	
Constant	0.082***	0.018	-	-
N	235,626		235,626	
(pseudo) R^2	0.391		0.356	

Note. Robust SEs adjusting for heteroskedasticity and zipcode-level clustering are used. ****$p < .1$; ***$p < .001$; **$p < .01$; *$p < .05$.

4.5.3 | Other robustness checks

We conduct several additional robustness checks. First, to account for the potential structural difference in service quality between physicians with more reviews and those with a significantly less number of reviews, we restrict our sample to those with at least 10 reviews and find the results to be highly consistent. Second, recognizing that family practitioners and specialists are typically considered as two groups of health-care providers that provide different types of services (i.e., general vs. specialized), the perceived quality of care by patients may systematically differ between the two groups. We address this issue by splitting our sample into primary care doctors and specialists and re-running our analyses. We find that results from both sub-samples are qualitatively the same as our main results. Third, patient satisfaction may be consistently higher in some medical specialties. The fact that physician's ratings may be correlated within each medical specialty could distort the variance of our estimators. To make sure that our results are robust to specialty-level clustering, we re-ran our regression models adjusting standard errors for all medical specialties. Again, our results remain qualitatively unchanged.

5 | DISCUSSION AND CONCLUSIONS

Online physician ratings and textual reviews are influencing patient decisions, elevating the importance of patient-centered health care, and changing the culture of health care. For physicians, these reviews offer valuable performance feedback for learning and improving; they can also improve health-care delivery and strengthen patient-physician relationships (Lee, 2017). However, little light has been shed on how health-care researchers, practitioners, and administrators can utilize this valuable data source. In this study, we take the first step by quantifying the effects of operational efficiency on patient satisfaction using a comprehensive online physician review data set spanning a diverse geography and multiple physician systems. We find that operational efficiency is a key determinant of how patients rate physicians, and that a significant proportion of textual reviews revolves around operations-related issues. Given the consistent results from our endogeneity and robustness analyses, the relationship between waiting time and overall rating is likely to be causal, indicating that improvement on waiting time will lead to tangible improvements in patient satisfaction.

While prior literature has established the importance of operational efficiency for clinical quality (Anderson, Gao, & Golden, 2014; Dobson, Hasija, & Pinker, 2011; Price et al., 2011), we provide empirical evidence that operational efficiency is critical to experiential quality as well: a waiting time longer than 17 min will, on average, reduce the odds of high patient satisfaction by more than 14%. Such finding is especially relevant at the intersection of two transformative trends—the priority of patient-centered health care and the ubiquity of online reviews. Waiting time has already been recognized to be tied with a physician's earnings (Gupta & Denton, 2008); waiting time will assume greater importance as patients increasingly turn to online ratings and textual reviews for choosing physicians, similar to how customers rely on online ratings for choosing products and services (Hanauer et al., 2014). The medical literature has also laid out compelling reasons for physicians to establish and manage their online presence in the era of social media (Gilbert et al., 2015). Hence, physicians and managers of health-care systems need to consider this added challenge when managing inefficient medical practices.

Given the importance of operational efficiency, what can health-care providers do to improve? Queuing theory suggests that physicians can reduce waiting time by taking three measures: reducing physician's utilization, smoothing out the variability of service time and arrival time, or reducing the average service time.[9] In this capacity, health care OM literature has provided many models such as optimal appointment scheduling and patient flow planning (Drupsteen, van der Vaart, & Pieter van Donk, 2013; LaGanga & Lawrence, 2012; Zacharias & Pinedo, 2014). Our results complement the existing literature by offering several novel insights. First, as Robinson and Chen (2011) point out, many models hinge on the fundamental trade-offs between patient waiting time and physician utilization. Our empirical results show that, although the negative impact of long waiting time is clear, the magnitude of the effect varies across specialties. Our analysis also shows the relative cost of the waiting time for different specialties, which can be critical for finding an optimal operational policy. For example, for specialties with long waiting time cost, physicians could ease utilization by using less aggressive overbooking or by seeing fewer patients. Hospitals can divert more operations-related resources to these specialties. To improve patient satisfaction, policymakers and the system as a whole can calibrate the payment scheme to balance the trade-offs between cost and patient volume for different specialties.

Second, our findings underscore the importance of two administrative activities that physicians can leverage to mitigate the negative effects brought on by waiting time. We show that the effect of waiting time does not depend on how well the physicians treat the patients *during* their visit, both technically and interpersonally. In other words, patients appear to compartmentalize long waiting time during a visit, even if other numeric ratings are satisfactory. Therefore, it is insufficient to focus on in-clinic activities alone. Our regression model suggests that a *previsit* activity (*Ease of Appointment*) and a *postvisit* activity (*Postvisit Follow-up*) are effective strategies that not only lead to higher patient satisfaction, but also relieve frustrations associated with longer waiting as it positively moderates the relationship between waiting time and patient satisfaction. This is an important insight as physicians do not necessarily have to balance a challenging task having to trade-off between physician utilization and patient volume; physicians can mitigate negative waiting time effects experienced by patients by focusing their attention to activities outside patients' appointment visit. For example, physicians can engage in previsit and postvisit activities such as completion of a Previsit Planning form in preparation for a patient's next visit or setting up regular reminders regarding adherence to prescribed home health care. The former sets expectations and streamlines service processes for the next encounter, while the latter provides a custom-tailored approach to health care. Overall, physicians could “worry less” about the negative effects brought on by waiting time by developing previsit and postvisit strategies that do not compete for physician utilization and service timeliness. We believe this novel and unexpected finding offers a path for new research stream previously unexplored, while proposing alternatives for physicians to improve their health-care service to patients.

Third, to develop a holistic view of the patient experience, physicians should heed both quantitative measures such as aspect ratings and qualitative information such as textual reviews from the patients. Our results suggest that textual reviews are complementary to the quantitative measures because they reveal what patients believe to be important. According to our guided LDA model, 18.2% of the total textual reviews are related to timeliness of the physician. Patients also devote 15.9% of the reviews on Cost and Billing, a dimension that is often overlooked in prior studies and standardized surveys. Moreover, we find that what is important to patients (e.g., technical quality vs. interpersonal quality) moderates the effect of waiting time. These novel discoveries suggest that physicians need to go beyond simply catering to addressing the objective measures of the encounter; they need to become more service-centric and adopt a customer-oriented "care" method by approaching a patient as an individual. This echoes the "marketing perspective" of patient satisfaction proposed by Cleary and McNeil (1988), which calls for a better understanding of patients' values, attitudes, and beliefs. Physician reviews provide a natural data source for such efforts. Another initiative may be for physicians to ask their patients to fill out a survey about their individual disposition and how much they value different health-care qualities. This may assist physicians optimize how best to deliver service and offer a customized solution that addresses both technical and interpersonal quality needs demanded by their patients.

We suggest several approaches to leverage our findings. Strategically, physicians should proactively monitor reviews from patients. Because online physician review sites are usually not affiliated with a physician and the reviews are completely voluntary, the narratives often offer clues that physicians can interpret much more constructively than just a standardized survey score or a star rating. Physicians can then consider communicating with dissatisfied patients directly on the online platform—an effective service recovery strategy according to Gu and Ye (2014). Moreover, practitioners should employ new machine learning tools such as LDA. The novel tools coupled with large-scale unstructured data can effectively extract performance measures from textual reviews. They can yield new insights for a better understanding of the health-care business, much like our study, and allow physicians and policymakers to detect shortcomings in health system performance.

Topic modeling can also be an important addition to the toolbox of OM researchers. There is a call for a new data-driven research stream that "let the data identify the specific issues, opportunities, and models that the organization should focus on" (Simchi-Levi, 2013). A major challenge in data-driven research is extracting insights from a large volume of unstructured, textual documents. LDA can be considered as an emergent clustering method to obtain a soft (multi-membership) clustering of documents (Brusco, Steinley, Cradit, & Singh, 2012). It is suitable for both exploratory content analysis (as in Lee, Qiu, & Whinston, 2018) and theory guided content analysis by incorporating prior knowledge in the model fitting process, as we demonstrated in our research.

This study is not without its limitations. First, the online physician reviews may not completely reflect the opinion of the patient population at large. For instance, López et al. (2012) find that patients who complete online reviews are younger and more affluent, who are less likely to respond to traditional patient satisfaction assessments. Future studies can complement our findings using provider-initiated patient survey data. Second, our data set is a cross-sectional data set constructed at the physician level, and thus does not capture the dynamic relationship between operational efficiency and patient satisfaction. One interesting direction for future research is to collect longitudinal data and explore how this relationship may evolve over time. Finally, the mechanism of how previsit and postvisit activities can best be operationalized is unclear. An in-depth examination could be made to explore the details of the appointment or follow-up systems such as timing, channel (telephone or email), protocols (best practices), effectiveness (verifying telephone numbers, obtaining best contact times, and informing patients that they will be contacted), and prediction (estimating successful postdischarge follow-up across demographic categories). Studying these questions will help us devise effective strategies to improve operational performance and thus deliver better patient-centered health care.

ACKNOWLEDGEMENT

We thank the participants of the Workshop on Health IT and Economics (2015), the 25th Workshop on Information Technologies and Systems (2015), University of Cincinnati and Cincinnati Children's Hospital Medical Center Biomedical Informatics Research Seminar (2016), and China Summer Workshop on Information Management (2017) who provided feedback on earlier drafts of this manuscript.

ENDNOTES

1 The current Vitals system uses a 5-point scale for ratings. At the time of data extraction, Vitals used a 4-point scale for ratings.

2 Only about 2.41% of the physicians in our sample have reported average waiting time of 60 min.

3 Available from https://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

4 Anecdotal evidence suggests that such decrease is substantial in terms of online ratings. Uber, for example, typically deactivates drivers if their customer ratings fall below 4.6/5.0 (Zahn, 2018).

5 The coefficients for waiting time in our logit regressions is −0.153 (Model 2). The corresponding odds ratio associated with high waiting time is $e^{-0.153} = 0.858$, that is, a 14.2% decrease in odds ratio.

6 Dagger (2007)'s model includes another dimension environment quality that describes the physical elements such as the design and layout of the clinic. However, our model was not able to find topics related to the definition of this dimension, with or without guidance.

7 UMass coherence score for a fitted topic t is defined as $C(t) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} \log \frac{D\left(w_m^t, w_l^t\right)+1}{D\left(w_l^t\right)}$, where $\left(w_1^t, \ldots w_M^t\right)$ are the top probable words for topic t in descending order, $D(w_1)$ is the number of reviews containing w_1, $D(w_1, w_2)$ is the number of reviews containing both w_1 and w_2. The coherence score for the entire model is the average of the individual topic's coherence scores. The UMass coherence score has another advantage over another commonly-used

measure, perplexity, because the coherence measure is an empirical measure based on the results of topic model. It does not depend on the likelihood of a held-out set, which is intractable when we guide the LDA using seed words.

[8] We thank an anonymous reviewer for this suggestion.

[9] For example, in a G/G/1 queue (a single-server queue with general arrival and service distribution), the following relationship, known as the Kingman's formula, holds: $W = V \times U \times T$. In the equation, W is the waiting time, V is the variability in service time and arrival time, U is the server utilization, and T is the average service time.

ORCID

Dawei Zhang https://orcid.org/0000-0002-6688-2961

REFERENCES

Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, *24*(6), 975–990.

Agency for Healthcare Research and Quality. (2011). Public reporting as a quality improvement strategy: A systematic review of the multiple pathways public reporting May influence quality of health care. Retrieved from http://www.effectivehealthcare.ahrq.gov/ehc/products/343/763/CQG-Public-Reporting_Protocol_20110817.pdf.

Ancarani, A., Di Mauro, C., & Giammanco, M. D. (2011). Patient satisfaction, managers' climate orientation and organizational climate. *International Journal of Operations & Production Management*, *31*(3), 224–250.

Anderson, D., Gao, G. G., & Golden, B. (2014). Life is all about timing: An examination of differences in treatment quality for trauma patients based on hospital arrival time. *Production and Operations Management*, *23*(12), 2178–2190.

Attree, M. (2001). Patients' and relatives' experiences and perspectives of 'Good' and 'not so Good' quality care. *Journal of Advanced Nursing*, *33*(4), 456–466.

Beadles, C. A., Ellis, A. R., Lichstein, J. C., Farley, J. F., Jackson, C. T., Morrissey, J. P., & Domino, M. E. (2015). First outpatient follow-up after psychiatric hospitalization: Does one size fit all? *Psychiatric Services*, *66*(4), 364–372.

Bensing, J. (1991). Doctor-patient communication and the quality of care. *Social Science & Medicine*, *32*(11), 1301–1310.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boyer, K. K., Gardner, J. W., & Schweikhart, S. (2012). Process quality improvement: An examination of general vs. outcome-specific climate and practices in hospitals. *Journal of Operations Management*, *30*(4), 325–339.

Brodie, R. J., Hollebeek, L. D., Jurić, B., & Ilić, A. (2011). Customer engagement. *Journal of Service Research*, *14*(3), 252–271.

Brusco, M. J., Steinley, D., Cradit, J. D., & Singh, R. (2012). Emergent clustering methods for empirical OM research. *Journal of Operations Management*, *30*(6), 454–466.

Camacho, F., Anderson, R., Safrit, A., Jones, A. S., & Hoffmann, P. (2006). The relationship between Patient's perceived waiting time and office-based practice satisfaction. *North Carolina Medical Journal*, *67*(6), 409–413.

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, *12*(4), 519–549.

Centers for Medicare and Medicaid Services. (2011). Physician compare web site overview–information for professionals. Retrieved from https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/physician-compare-initiative/.

Chan, H. K., Wang, X., Lacka, E., & Zhang, M. (2016). A mixed-method approach to extracting the value of social media data. *Production and Operations Management*, *25*(3), 568–583.

Chandrasekaran, A., Senot, C., & Boyer, K. K. (2012). Process management impact on clinical and experiential quality: Managing tensions between safe and patient-centered healthcare. *Manufacturing & Service Operations Management*, *14*(4), 548–566.

Chen, H., Zheng, Z. E., & Ceran, Y. (2016). De-biasing the reporting bias in social media analytics. *Production and Operations Management*, *25*(5), 849–865.

Cleary, P. D., & McNeil, B. J. (1988). Patient satisfaction as an indicator of quality care. *Inquiry*, *25*(1), 25–36.

Correia, A. W., Peters, J. L., Levy, J. I., Melly, S., & Dominici, F. (2013). Residential exposure to aircraft noise and hospital admissions for cardiovascular diseases: Multi-airport retrospective study. *British Medical Journal*, *347*(f5561), 1–11.

Dagger, T. S., & Sweeney, J. C. (2007). Service quality attribute weights. *Journal of Service Research*, *10*(1), 22–42.

Dagger, T. S., Sweeney, J. C., & Johnson, L. W. (2007). A hierarchical model of health service quality scale development and investigation of an integrated model. *Journal of Service Research*, *10*(2), 123–142.

Dixon, M., Freeman, K., & Toman, N. (2010). Stop trying to delight your customers. *Harvard Business Review*, *88*(7/8), 116–122.

Dobrzykowski, D. D., Callaway, S. K., & Vonderembse, M. A. (2015). Examining pathways from innovation orientation to patient satisfaction: A relational view of healthcare delivery. *Decision Sciences*, *46*(5), 863–899.

Dobson, G., Hasija, S., & Pinker, E. J. (2011). Reserving capacity for urgent patients in primary care. *Production and Operations Management*, *20*(3), 456–473.

Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA: The Journal of the American Medical Association*, *260*(12), 1743–1748.

Douglas, T. J., & Fredendall, L. D. (2004). Evaluating the deming management model of Total quality in services. *Decision Sciences*, *35*(3), 393–422.

Drupsteen, J., van der Vaart, T., & Pieter van Donk, D. (2013). Integrative practices in hospitals and their impact on patient flow. *International Journal of Operations & Production Management*, *33*(7), 912–933.

Fisher, M., Newman, G. E., & Dhar, R. (2018). Seeing stars: How the binary bias distorts the interpretation of customer ratings. *Journal of Consumer Research*, *45*(3), 471–489.

Fox, S. & Jones, S. (2012). The social life of health information. Pew internet & American life project 2009. Retrieved from http://www.pewinternet.org/2009/06/11/the-social-life-of-health-information/.

Froehle, C. M., & Magazine, M. J. (2013). Improving scheduling and flow in complex outpatient clinics. In *Handbook of healthcare operations management: Methods and applications*, (pp. 229–250). New York: Springer.

Fung, C. H., Elliott, M. N., Hays, R. D., Kahn, K. L., Kanouse, D. E., McGlynn, E. A., … Shekelle, P. G. (2005). Patients' preferences for technical versus interpersonal quality when selecting a primary care physician. *Health Services Research*, *40*(4), 957–977.

Gao, G., Greenwood, B., McCullough, J., & Agarwal, R. (2015). Vocal minority and silent majority: How do online ratings reflect population perceptions of quality? *MIS Quarterly*, *39*(3), 565–589.

Gilbert, K., Hawkins, C. M., Hughes, D. R., Patel, K., Gogia, N., Sekhar, A., & Duszak, R. (2015). Physician rating websites: Do radiologists have an online presence? *Journal of the American College of Radiology*, *12*(8), 867–871.

Gill, L., & White, L. (2009). A critical review of patient satisfaction. *Leadership in Health Services*, *22*(1), 8–19.

Goldstein, S. M. (2003). Employee development: An examination of service strategy in a high-contact service environment. *Production and Operations Management*, *12*(2), 186–203.

Greaves, F., Pape, U. J., Lee, H., Smith, D. M., Darzi, A., Majeed, A., & Millett, C. (2012). Patients' ratings of family physician practices on the internet: Usage and associations with conventional measures of quality in the English National Health Service. *Journal of Medical Internet Research*, *14*(5), e146.

Green, L. V. (2012). OM forum—The vital role of operations analysis in improving healthcare delivery. *Manufacturing & Service Operations Management*, *14*(4), 488–494.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Grondahl, V. A., Wilde-Larsson, B., Karlsson, I., & Hall-Lord, M. L. (2013). Patients' experiences of care quality and satisfaction during hospital stay: A qualitative study. *European Journal for Person Centered Healthcare*, *1*(1), 185–192.

Gu, B., & Ye, Q. (2014). First step in social media: Measuring the influence of online management responses on customer satisfaction. *Production and Operations Management*, *23*(4), 570–582.

Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, *40*(9), 800–819.

Gutman, J. (1982). A means-end chain model based on consumer categorization processes. *Journal of Marketing*, *46*(2), 60.

Haas-Wilson, D. (1994). The relationships between the dimensions of health care quality and price: The case of eye care. *Medical Care*, *32*(2), 175–182.

Hanauer, D. A., Zheng, K., Singer, D. C., Gebremariam, A., & Davis, M. M. (2014). Public awareness, perception, and use of online physician rating sites. *JAMA: The Journal of the American Medical Association*, *311*(7), 734–735.

Haraden, C., & Resar, R. (2004). Patient flow in hospitals: Understanding and controlling it better. *Frontiers of Health Services Management*, *20*(4), 3–15.

Harris, K., & Buntin, M. (2008). Choosing a health care provider: The role of quality information. *Policy*, *1*, 14.

Harvey, J. (1998). Service quality: A tutorial. *Journal of Operations Management*, *16*(5), 583–597.

Hay, M. C., Strathmann, C., Lieber, E., Wick, K., & Giesser, B. (2008). Why patients go online: Multiple sclerosis, the internet, and physician-patient communication. *The Neurologist*, *14*(6), 374–381.

Heath, S. (2018). Long appointment wait time a detriment to high patient satisfaction. Patient engagement HIT. Retrieved from https://patientengagementhit.com/news/long-appHeat.

Huang, X.-M. (1994). Patient attitude towards waiting in an outpatient clinic and its applications. *Health Services Management Research*, *7*(1), 2–8.

Hui, M. K., Thakor, M. V., & Gill, R. (1998). The effect of delay type and service stage on consumers' reactions to waiting. *Journal of Consumer Research*, *24*(4), 469–480.

Jagarlamudi, J., Daumé III, H., & Udupa, R. (2012). *Incorporating lexical priors into topic models*. Paper presented at the Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL'12). Avignon, France.

Jha, N., Premarajan, K. C., Nagesh, S., Khanal, S., & Thapa, L. B. (2005). Five-star doctors for the 21st century: A BPKIHS endeavour for Nepal. *Journal of Health Management*, *7*(2), 237–247.

Johnson, P. F., Klassen, R. D., Leenders, M. R., & Awaysheh, A. (2007). Utilizing E-business technologies in supply chains: The impact of firm characteristics and teams. *Journal of Operations Management*, *25*(6), 1255–1274.

Kadry, B., Chu, L. F., Kadry, B., Gammas, D., & Macario, A. (2011). Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of Medical Internet Research*, *13*(4), e95.

Kulkarni, S. S., Apte, U. M., & Evangelopoulos, N. E. (2014). The use of latent semantic analysis in operations management research. *Decision Sciences*, *45*(5), 971–994.

LaGanga, L. R., & Lawrence, S. R. (2012). Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management*, *21*(5), 874–888.

Lam, H. K. S., Yeung, A. C. L., & Cheng, T. C. E. (2016). The impact of firms' social media initiatives on operational efficiency and innovativeness. *Journal of Operations Management*, *47–48*, 28–43.

Larrabee, J. H., & Bolden, L. V. (2001). Defining patient-perceived quality of nursing care. *Journal of Nursing Care Quality*, *16*(1), 34–60.

Larson, R. C., Larson, B. M., & Katz, K. L. (1991). Prescription for waiting–in line blues: Entertain, enlighten and engage. *Sloan Management Review*, *32*(2), 44–55.

Lee, S. Y., Qiu, L., & Whinston, A. (2018). Sentiment manipulation in online platforms: An analysis of movie tweets. *Production and Operations Management*, *27*(3), 393–416.

Lee, V. (2017). Transparency and trust—Online patient reviews of physicians. *New England Journal of Medicine*, *376*(3), 197–199.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96.

Long, C., Tsay, E. L., Jacobo, S. A., Popat, R., Singh, K., & Chang, R. T. (2016). Factors associated with patient press Ganey satisfaction scores for ophthalmology patients. *Ophthalmology*, *123*(2), 242–247.

López, A., Detz, A., Ratanawongsa, N., & Sarkar, U. (2012). What patients say about their doctors online: A qualitative content analysis. *Journal of General Internal Medicine*, *27*(6), 685–692.

Lu, Y., Musalem, A., Olivares, M., & Schilkrut, A. (2013). Measuring the effect of queues on customer purchases. *Management Science*, *59*(8), 1743–1763.

Luxford, K. (2012). What does the patient know about quality? *International Journal for Quality in Health Care*, *24*(5), 439–440.

Maister, D. H. (1985). The psychology of waiting lines. In J. A. Czepiel, M. R. Solomon, & C. F. Surprenant (Eds.), *The service encounter* (pp. 113–123). Lexington, MA: Lexington Books.

Marley, K. A., Collier, D. A., & Meyer Goldstein, S. (2004). The role of clinical and process quality in achieving patient satisfaction in hospitals. *Decision Sciences*, *35*(3), 349–369.

Maute, M. F., & Forrester, W. R., Jr. (1991). The effect of attribute qualities on consumer decision making: A causal model of external information search. *Journal of Economic Psychology*, *12*(4), 643–666.

McCarthy, K., McGee, H. M., & O'Boyle, C. A. (2000). Outpatient clinic waiting times and non-attendance as indicators of quality. *Psychology, Health & Medicine*, *5*(3), 287–293.

Merlino, J. I. (2016). How to improve patient satisfaction scores by using data. Retrieved from https://www.healthcatalyst.com/insights/patient-satisfaction-outcomes-data.

Merlino, J. I., & Raman, A. (2013). Health care's service fanatics. *Harvard Business Review*, *91*(5), 108–116.

Meyer, A. N. D., Payne, V. L., Meeks, D. W., Rao, R., & Singh, H. (2013). Physicians' diagnostic accuracy, confidence, and resource requests. *JAMA Internal Medicine*, *173*(21), 1952.

Meyer, T. (1994). Subjective importance of goal and reactions to waiting in line. *The Journal of Social Psychology*, *134*(6), 819–827.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *2*, 1–9.

Milburn, M., Baker, M.-J., Gardner, P., Hornsby, R., & Rogers, L. (1994). Nursing care that patients value. *British Journal of Nursing*, *4*(18), 1094–1098.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). *Optimizing semantic coherence in topic models*. Paper presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland.

Moore, A. D., Hamilton, J. B., Krusel, J. L., Moore, L. G., & Pierre-Louis, B. J. (2016). Patients provide recommendations for improving patient satisfaction. *Military Medicine*, *181*(4), 356–363.

Mühlbacher, A., Johnson, F. R., Yang, J.-C., Happich, M., & Belger, M. (2016). Do you want to hear the bad news? The value of diagnostic tests for Alzheimer's disease. *Value in Health*, *19*(1), 66–74.

Murphy, J., Chang, H., Montgomery, J. E., Rogers, W. H., & Gelb Safran, D. (2001). The quality of physician–patient relationships. *The Journal of Family Practice*, *50*(2), 123–129.

Nair, A., Nicolae, M., & Narasimhan, R. (2013). Examining the impact of clinical quality and clinical flexibility on cardiology unit performance—Does experiential quality act as a specialized complementary asset? *Journal of Operations Management*, *31*(7), 505–522.

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, Price premiums, and seller differentiation. *Information Systems Research*, *17*(4), 392–414.

Pedraza-Martinez, A. J., & Van Wassenhove, L. N. (2016). Empirically grounded research in humanitarian operations management: The way forward. *Journal of Operations Management*, *45*, 1–10.

Price, C., Golden, B., Harrington, M., Konewko, R., Wasil, E., & Herring, W. (2011). Reducing boarding in a post-anesthesia care unit. *Production and Operations Management*, *20*(3), 431–441.

Queenan, C. C., Angst, C. M., & Devaraj, S. (2011). Doctors' orders—If They're electronic, do they improve patient satisfaction? A complements/substitutes perspective. *Journal of Operations Management*, *29*(7), 639–649.

Rao, M., Clarke, A., Sanderson, C., & Hammersley, R. (2006). Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: Cross sectional study. *British Medical Journal*, *333*(7557), 19.

Robinson, L. W., & Chen, R. R. (2011). Estimating the implied value of the customer's waiting time. *Manufacturing & Service Operations Management*, *13*(1), 53–57.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320.

Saad, L. (2014). The '40-Hour' workweek is actually longer—by seven hours. Gallup. Retrieved from http://www.gallup.com/poll/175286/hour-workweek-actually-longer-seven-hours.aspx.

Salzarulo, P. A., Bretthauer, K. M., Côté, M. J., & Schultz, K. L. (2011). The impact of variability and patient information on health care system performance. *Production and Operations Management*, *20*(6), 848–859.

Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, *190*(2), 212–221.

Schlesinger, M., Grob, R., & Shaller, D. (2015). Using patient-reported information to improve clinical practice. *Health Services Research*, *50*(S2), 2116–2154.

Schlesinger, M., Grob, R., Shaller, D., Martino, S. C., Parker, A. M., Finucane, M. L., … Rybowski, L. (2015). Taking patients' narratives about clinicians from anecdote to science. *The New England Journal of Medicine*, *373*(7), 675–679.

Schmitt, B. H. (2011). Experience Marketing: Concepts, Frameworks and Consumer Insights. *Foundations and Trends in Marketing*, *5*(2), 55–112.

Segal, J., Sacopulos, M., Sheets, V., Thurston, I., Brooks, K., & Puccia, R. (2012). Online doctor reviews: Do they track surgeon volume, a proxy for quality of care? *Journal of Medical Internet Research*, *14*(2), e50.

Senot, C., Chandrasekaran, A., & Ward, P. T. (2016). Role of bottom-up decision processes in improving the quality of health care delivery: A contingency perspective. *Production and Operations Management*, *25*(3), 458–476.

Senot, C., Chandrasekaran, A., Ward, P. T., Tucker, A. L., & Moffatt-Bruce, S. D. (2015). The impact of combining conformance and experiential quality on hospitals' readmissions and cost performance. *Management Science*, *62*(3), 829–848.

Sharma, L., Chandrasekaran, A., Boyer, K. K., & McDermott, C. M. (2016). The impact of health information technology bundles on hospital performance: An econometric study. *Journal of Operations Management*, *41*, 25–41.

Sharma, N., & Patterson, P. G. (1999). The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services. *Journal of Services Marketing*, *13*(2), 151–170.

Simchi-Levi, D. (2013). OM forum—OM research: From problem-driven to data-driven research. *Manufacturing & Service Operations Management*, *16*(1), 2–10.

Sousa, R. (2003). Linking quality management to manufacturing strategy: An empirical investigation of customer focus practices. *Journal of Operations Management*, *21*(1), 1–18.

Spake, D. F., & Megehee, C. M. (2010). Consumer sociability and service provider expertise influence on service relationship success. *Journal of Services Marketing*, *24*(4), 314–324.

Stewart, A. L., Nápoles-Springer, A. M., Gregorich, S. E., & Santoyo-Olsson, J. (2007). Interpersonal processes of care survey: Patient-reported measures for diverse groups. *Health Services Research*, *42* (3p1), 1235–1256.

Tang, C. S. (2015). The past, present, and future of manufacturing & service operations management. *Manufacturing & Service Operations Management*, *17*(1), 1–3.

The Economist. (2014). DocAdvisor: Patients around the World are starting to give doctors a piece of their mind. The result should be better care. Retrieved from http://www.economist.com/news/international/21608767-patients-around-world-are-starting-give-doctors-piece-their-mind-result.

Thirumalai, S., & Sinha, K. K. (2011). Customization of the online purchase process in electronic retailing and customer satisfaction: An online field study. *Journal of Operations Management*, *29*(5), 477–487.

Thomas, S., Glynne-Jones, R., & Chait, I. (1997). Is it worth the wait? A survey of patients' satisfaction with an oncology outpatient clinic. *European Journal of Cancer Care*, *6*(1), 50–58.

Tirunillai, S., & Tellis, G. (2014). Mining marketing meaning from online chatter. *Journal of Marketing Research*, *51*(4), 463–479.

Veeraraghavan, S. K., & Debo, L. G. (2011). Herding in queues with waiting costs: Rationality and regret. *Manufacturing & Service Operations Management*, *13*(3), 329–346.

Verhoef, L. M., Van de Belt, T. H., Engelen, L. J., Schoonhoven, L., & Kool, R. B. (2014). Social media and rating sites as tools to understanding quality of care: A scoping review. *Journal of Medical Internet Research*, *16*(2), e56.

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer experience creation: Determinants, dynamics and management strategies. *Journal of Retailing*, *85*(1), 31–41.

Xu, Y., Armony, M., & Ghose, A. (2016). The effect of online reviews on physician demand: A structural model of patient choice. Working Paper. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2778664.

Zacharias, C., & Pinedo, M. (2014). Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, *23*(5), 788–801.

Zahn, M. (2018). Uber university: Where drivers learn to achieve a 5.0. *Money*, *47*(8), 20–21.

Zandbelt, L. C., Smets, E. M. A., Oort, F. J., Godfried, M. H., & De Haes, H. C. J. M. (2004). Satisfaction with the outpatient encounter: A comparison of Patients' and Physicians' views. *Journal of General Internal Medicine*, *19*(11), 1088–1095.

Zohar, E., Mandelbaum, A., & Shimkin, N. (2002). Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science*, *48*(4), 566–583.

How to cite this article: Ko D-G, Mai F, Shan Z, Zhang D. Operational efficiency and patient-centered health care: A view from online physician reviews. *J Oper Manag*. 2019;1–27. https://doi.org/10.1002/joom.1028

APPENDIX A: TOPIC MODELING

We describe the latent Dirichlet allocation (LDA) model (Blei et al. 2003) and its extension, the guided LDA model (Jagarlamudi et al., 2012) in detail. The LDA model assumes the reviews are generated from a probabilistic process characterized by latent random variables. To make the inference process trackable, LDA makes the following assumptions about the generation process: (a) words contained in each review are generated from a mixture of T topics that the patients deemed important; (b) each topic has a probability distribution over a fixed word vocabulary V (i.e., some words are more likely to be used to describe a specific topic); and (c) the ordering of the words does not matter. This is commonly known as the "bag-of-words" assumption in text analysis—an unrealistic assumption on paper but one that works well for many natural language processing tasks. The generation process proceeds as follows:

1. For each topic $k \in 1 \ldots T$, draw a vocabulary mixture $\boldsymbol{\phi}_k$ for a topic from Dirichlet ($\boldsymbol{\beta}$).
2. A patient would write each review $d \in 1 \ldots D$ following the process:
3. Sample topic proportions $\boldsymbol{\theta}_d$ from a Dirichlet distribution with parameters $\boldsymbol{\alpha}$.
4. For each of the $n \in 1 \ldots N_d$ words in d:

a. Sample a topic assignment $z_{d,\,n}$ from Multinomial ($\boldsymbol{\theta}_d$), where $z_{d,\,n}$ is a topic index between $1 \ldots T$.
b. Choose the word $w_{d,\,n}$ from Multinomial $\left(\boldsymbol{\phi}_{z_{d,n}}\right)$.

The Bayesian model has two hyperparameters $\boldsymbol{\alpha},\boldsymbol{\beta}$ that encode our prior knowledge about the topic proportions and how the words are related to each topic. We observe the actual outputs of the generative process: each review and the words contained therein. Given the output, the Bayesian estimation infers (a) the parameters for the topic proportions $\boldsymbol{\theta_d}$ to determine which dimensions are important and (b) the parameters for the topic-word distribution $\boldsymbol{\phi}_k$ to determine the keywords associated with each topic. In other words, LDA allows us to attribute important words to topics and decides whether and how much to allocate the content of review to a topic.

Guided LDA allows the LDA model to learn topics of specific interest. In our application, the topics of interests are initialized using a theoretical model of health-care services (Dagger et al., 2007). For each topic of interest, we provide the algorithm with a list of seed words (see Section 4.2). To guide the LDA model to the topics of interests that the seed words have higher weights, we need to adjust the two hyperparameters $\boldsymbol{\alpha},\boldsymbol{\beta}$ of the prior probability distributions. In a regular LDA, both hyperparameters are symmetric, that is, the numbers in each element in a vector are the same. This indicates that we do not have any prior knowledge about which topics have higher proportions, and we also do not have any prior knowledge about which words are associated with any topic. Guided LDA essentially tunes up the prior probability of seed words in $\boldsymbol{\beta}_k$ so that they are more likely to be associated with a specific topic k by changing the generation process in the following way:

1. For each topic $k \in 1 \ldots T$, draw
   - **a.** a regular vocabulary mixture $\boldsymbol{\phi}_k^r$ for a topic from Dirichlet ($\boldsymbol{\beta}_r$).
   - **b.** a seed vocabulary mixture $\boldsymbol{\phi}_k^s$ for a topic from Dirichlet ($\boldsymbol{\beta}_s$).
   - **c.** a number π_k between (0, 1) from Uniform (0, 1).
2. A patient would write each review $d \in 1 \ldots D$ following the process:
3. Sample topic proportions $\boldsymbol{\theta}_d$ from a Dirichlet distribution with parameters $\boldsymbol{\alpha}$.
4. For each of the $n \in 1 \ldots N_d$ words in d:
   - **a.** Sample a topic assignment $z_{d,\,n}$ from Multinomial ($\boldsymbol{\theta}_d$), where $z_{d,\,n}$ is a topic index between $1 \ldots T$.
   - **b.** Select an indicator x_i from Bernoulli ($\pi_{z_{d,n}}$).
   - **c.** If $x_i = 0$, choose the word $w_{d,\,n}$ from Multinomial $\left(\boldsymbol{\phi}^r{}_{z_{d,n}}\right)$.
   - **d.** If $x_i = 1$, choose the word $w_{d,\,n}$ from Multinomial $\left(\boldsymbol{\phi}^s{}_{z_{d,n}}\right)$.

The key difference between the guided LDA and the regular LDA is that each topic k is associated with two word-distributions: $\boldsymbol{\phi}_k^s$ only contains the seed words, and $\boldsymbol{\phi}_k^r$ contains all the words in the vocabulary. When deciding how a word is generated from a review (Step 4), the generation process first chooses a topic k just as a regular LDA would. Then a biased coin is flipped. If the coin lands on the head, the word can only be chosen from seed words of the topic; if the coin lands on the tail, all words in the vocabulary can be chosen. This small change enables the model to gather the synonyms of the seed words to the same topic. The model can also generate seed topics from theory (technical quality, interpersonal quality, timeliness in our paper) and discover "free" topics (cost and billing, family members' experience in our article) at the same time by letting the seed words of the "free" topics be the entire vocabulary.

Before fitting the topic models, we preprocess the corpus using the following pipeline. We first prepare the text in the reviews using several preprocessing steps that transform the unstructured text into components that the LDA algorithm can accept as inputs. We then conduct lemmatization to transform words into their root forms and remove variance; for example, *nurses* becomes *nurse*, and *gone* becomes *go.* This step would help the LDA inference by reducing the dimensionality of vocabulary. We then filter out stop words, most common terms (50 words in each specialty's reviews), and rare terms (occurrence <10) to reduce the noise in the model outputs. These words carry little information and filtering them out is a common procedure in computational linguistics. As stated in the article, we also filter out the sentiment-related words in the MPQA lexicon to avoid any tautological issue in our regression analysis. Finally, we find common phrases such as answer question and bedside manner in the text and treat them as single words using the heuristic described in Mikolov et al. (2013).