MIS Quarterly

RESEARCH ARTICLE

VOCAL MINORITY AND SILENT MAJORITY: HOW DO ONLINE RATINGS REFLECT POPULATION PERCEPTIONS OF QUALITY[1]

Guodong (Gordon) Gao
Robert H Smith School of Business, University of Maryland, 4325 Van Munching Hall,
College Park, MD 20742-1815 U.S.A. {ggao@rhsmith.umd.edu}

Brad N. Greenwood
Fox School of Business, Temple University, 1801 Liacouras Walk,
Philadelphia, PA 19122-6083 U.S.A. {greenwood@temple.edu}

Ritu Agarwal
Robert H Smith School of Business, University of Maryland, 4325 Van Munching Hall,
College Park, MD 20742-1815 U.S.A. {ragarwal@rhsmith.umd.edu}

Jeffrey S. McCullough
School of Public Health, University of Minnesota, 420 Delaware Street S.E.,
Minneapolis, MN 55455 U.S.A. {mccu0056@umn.edu}

Consumer-generated ratings typically share an objective of illuminating the quality of a product or service for other buyers. While ratings have become ubiquitous and influential on the Internet, surprisingly little empirical research has investigated how these online assessments reflect the opinion of the population at large, especially in the domain of professional services where quality is often opaque to consumers. Building on the word-of-mouth literature, we examine the relationship between online ratings and population perceptions of physician quality. We leverage a unique dataset which includes direct measures of both the offline population's perception of physician quality and consumer-generated online reviews. As a result, we are able to examine how online ratings reflect patients' opinions about physician quality. In sharp contrast to the widely voiced concerns by medical practitioners, we find that physicians who are rated lower in quality by the patient population are less likely to be rated online. Although ratings provided online are positively correlated with patient population opinions, the online ratings tend to be exaggerated at the upper end of the quality spectrum. This study is the first to provide empirical evidence of the relationship between online ratings and the underlying consumer-perceived quality, and extends prior research on online word-of-mouth to the domain of professional services.

Keywords: Online ratings, physician quality, online word-of-mouth, professional services, informativeness

[1]Kai Lim was the accepting senior editor for this paper. Bin Gu served as the associate editor.

Introduction

Recent years have witnessed a rapid growth of online ratings by consumers across many products and services. Extant literature has shown that these ratings can significantly affect sales (e.g., Chevalier and Mayzlin 2006; Clemons et al. 2006; Liu 2006), underscoring their importance in the online marketplace. However, when consumers rely on online ratings to make decisions, legitimate concerns emerge that the generation of online ratings may be subject to bias as a result of various environmental and behavioral factors, including the self-selection of early buyers, the influence of existing ratings, and different types of managerial intervention (e.g., Godes and Silva 2012; Gu and Ye 2014; Li and Hitt 2008; Moe and Trusov 2011; Muchnik et al. 2013; Wang et al. 2010). To the degree that online ratings are typically generated by a fraction of users, the ability of ratings to inform consumers about the quality of products or services is questionable. For example, not all products receive ratings, begging the question of how the quality of products with ratings compares with those that are not rated. If rated, how closely do online ratings reflect the perceived quality of the product? Further, does the discriminatory power of online ratings vary across the quality spectrum? Although answering these questions has important implications for the usefulness of online ratings for consumer choice, few studies have investigated these critical issues.

In this paper we provide insights into the relationship between online ratings and the underlying consumer population's perception of quality. While the majority of prior research attempts to infer product quality from online ratings (e.g., Li and Hitt 2008; Moe and Trusov 2011), our work is distinctive in that we examine how quality is reflected in online ratings by using direct measures of quality. We use detailed data on 1,425 primary care physicians in three metropolitan areas which contains, among other variables, population measures of physician quality as rated by patients, constructed using rigorous standards set by the Agency for Healthcare Research and Quality (AHRQ) of the U.S. Department of Health and Human Services. We match these data to online ratings from one of the largest physician rating websites as well as other demographic and economic variables. This unique dataset enables us to investigate the relationship between population quality measures and online ratings.

The research presented here makes several important contributions. First, to the best of our knowledge, this study is the first to provide empirical evidence of the relationship between online ratings and population perceptions of quality. Second, we extend research on online word-of-mouth into the domain of professional services. Although online reviews for consumer goods such as books and movies are well-known, online word-of-mouth (WOM) is increasingly expanding to include professional service providers such as auto mechanics, lawyers, and physicians. To date, however, studies on online WOM have focused limited attention on the expanding expert service market (a brief review of existing literature is provided in Table 1). Online ratings may play an even greater role in affecting consumer decisions for professional services as they often possess experience and credence qualities and have been characterized as "natural candidates" for WOM communication (Harrison-Walker 2001).

Background and Prior Literature

Word-of-Mouth and Social Communication

WOM has long been acknowledged as an important determinant of consumer choice (e.g., Arndt 1967; Bass 1969; Katz and Lazarsfeld 1955). Until approximately a decade ago, WOM was, by its very nature, limited to individuals with close social connections (Brown and Reingen 1987) and/or close physical proximity. In recent years, however, the notion of WOM has been extended to include a wide range of social communication enabled by technology, and growing empirical evidence suggests that these new forms of communication are highly influential in guiding consumer decisions. As a result, an increasing number of studies have examined the phenomenon of online consumer reviews (e.g., Clemons et al. 2006; Dellarocas et al. 2004; Dellarocas and Narayan 2006; Duan et al. 2008; Godes and Mayzlin 2004; Li and Hitt 2008; Moe and Trusov 2011; Muchnik et al. 2013; Sun 2012; Zhu and Zhang 2010).

Table 1 provides a brief summary of recent, representative empirical literature on WOM. Researchers have studied WOM across diverse product categories, ranging from video games to movies to beer, and the effect of these online reviews has been explored on a variety of outcomes, such as product sales, market penetration, and box-office performance. Extant literature on online reviews has also enriched theoretical models of the relationship between online WOM and outcomes with multiple moderating and mediating variables including product experience, the location of WOM, consumer characteristics, and trust.

Despite the breadth of literature in this area, three critical gaps exist that represent fruitful opportunities for research. First, there is limited work in the context of credence goods such as professional services. The dominant contexts studied in prior research include online retailing (Chen and Xie 2008; Chevalier and Mayzlin 2006; Dellarocas and Wood 2008; Li and Hitt 2008; Richins 1983), movies (Chintagunta et al. 2010;

Table 1. Overview of Major Empirical Studies in Online Word-of-Mouth			
Phenomenon	**Paper**	**Finding**	**Context**
Generation of WOM	Dellarocas and Narayan (2006)	Curvilinear Relationship between satisfaction and WOM	Film
	Berger and Iyengar (2012)	Synchronicity of conversation channel impacts which products receive WOM based on how interesting the object is	Experiment
	Dellarocas et al. (2010)	Niche products are more likely to receive online reviews	Film
Trust and value in WOM	Brown et al. (2007)	Social platforms are viewed as actors in social networks and have an effect on the perception and value of WOM	TV programs
	Mudambi and Schuff (2010)	Depth of review is correlated with helpfulness, however extremity is less helpful for experience goods	Books, CD, etc.
WOM and Ratings Dynamics	Godes and Mayzlin (2004)	Online conversation is an effective proxy for WOM and has explanatory power when measuring ratings	TV programs
	Dellarocas et al. (2004)	Online ratings are an effective proxy for WOM	Film
	Moe and Trusov (2011)	WOM affects sales and is subject to social dynamics in that ratings will affect future rating behavior	Bath, fragrance and beauty products
	Godes and Silva (2012)	As the number of ratings increase the average score decreases. This increases with positivity, which induces purchase errors and dissatisfaction	Books
	Muchnik et al (2013)	Social influence biases the rating dyanmics	News
Effect of WOM on Sales/ Performance	Chen et al. (2004)	Customer reviews are not correlated with sales, however recommendations are highly significant	Books
	Senecal and Nantel (2004)	Online recommender systems are more influential than expert reviews when determining customer product choice	Electronics and wine
	Liu (2006)	Posted reviews are highly correlated with box office sales	Film
	Chevalier and Mayzlin (2006)	Online Reviews positively relate to sales	Books
	Clemons et al. (2006)	Both the mean and dispersion of WOM are associated with sale growth	Craft Beers
	Li and Hitt (2008)	WOM is not an unbiased indication of quality and will affect sales	Books
	Chen and Xie (2008)	Consumer WOM can be used as a valuable tool for online marketing and can be strategically manipulated to increase sales	Cameras
	Duan et al. (2008)	Box office revenue is correlated with WOM, which in turn is correlated with performance	Film
	Forman et al. (2008)	Reviewer disclosure of identity is a strong indicator of both future sales and future geographic sales	Books
	Trusov et al. (2009)	WOM references increase social network tenure	Social Networks
	Chintagunta et al. (2010)	WOM is correlated with film box office performance	Film
	Zhu and Zhang (2010)	Consumer characteristics moderate the relationship between WOM and sales	Video Gaming
	Gu et al. (2011)	WOM on external review websites is a more effective indicator of sales for high-involvement products	Cameras
	Hennig-Thurau et al. (2004)	Increases in Micro WOM valence (as moderated by volume and customer characteristics) significantly impacts product adoption	Film
	Sun (2012)	When average rating is low, higher variance is associated with greater demand	Books

Dellarocas et al. 2010; Dellarocas and Narayan 2006; Duan et al. 2008), and financial markets (Hong et al. 2005). Professional services, however, are distinctive insofar as there is considerable information asymmetry between sellers and buyers, and consumers typically lack the specialized knowledge required to evaluate the quality of the service (Arrow 1963). In principle, online customer opinions in such contexts should create significant welfare because information acquisition is not only costly, but often infeasible. In healthcare settings in particular, the information deficit that consumers confront has been characterized as particularly acute, underscoring the need for rigorous research to promote additional information transparency (Christianson et al. 2010; Harris and Buntin 2008).

A second gap relates to the question of which products/services are more likely to receive online WOM: although researchers have studied *when* products are likely to receive reviews from consumers (e.g., Anderson 1998; Dellarocas et al. 2010; Dellarocas and Narayan 2006; Dellarocas and Wood 2008); the question of *which* products receive ratings has received limited attention. For example, research by Anderson (1998) and Dellarocas and Narayan (2006) examines how the marginal experience consumers have with a product is likely to affect their propensity to rate that product online. Dellarocas et al. (2010) address the question of whether niche or hit films are likely to receive more or fewer online reviews. Likewise, after observing that "less is known about its [WOM's] causes or what leads people to talk about certain products and brands rather than others," Berger and Iyengar (2012, p. 2) find that more interesting products and brands generate more online WOM but not face-to-face WOM. Although these studies shed light on the characteristics of products (e.g., popularity and interestingness, respectively) that lead to online WOM, the question of how product or service quality affects the availability of online ratings remains unanswered. To the degree that the main purpose of reading online ratings is to learn about quality and facilitate comparison, understanding how quality correlates with the availability of online ratings is vitally important. If, for example, most rated physicians came from the low-quality cohort, the utility of the ratings would be considerably degraded as the available information would be limited to one end of the quality spectrum.

The final gap we note relates to the implicit assumption in this body of literature that higher online ratings reflect higher quality, resulting in a higher likelihood of purchase (e.g., Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Zhu and Zhang 2010). Surprisingly, in spite of the pervasiveness of this assumption, little empirical evidence exists demonstrating the relationship between online ratings and product quality. Clearly, online ratings are contributed by self-motivated individuals. While it has been suggested that products yielding a customer experience that significantly deviates from *a priori* expectations are more likely to be rated (both in the case of positive or negative experiences) (Anderson 1998), this research does not provide an answer to the question of how representative these ratings are of the offline population's perception of the product quality. Given the documented selection biases in the decision to rate online, together with evidence of punitive or reinforcing effects of early reviews on long-term sales (Godes and Silva 2012; Li and Hitt 2008), there is a critical need to examine the relationship between online ratings and intrinsic quality.

In summary, our review of the literature on WOM reveals that a majority of studies focus on establishing the value proposition of WOM by relating reviews to financial measures. Consumer reviews have been examined from various perspectives, including the dispersion of WOM (Godes and Mayzlin 2004), the distribution of the ratings (Clemons et al. 2006, Sun 2012), product segment (Dellarocas et al. 2010), valence (Chintagunta et al. 2010), and consumer characteristics (Zhu and Zhang 2010). All of these studies have provided important insights into the relationship between online consumer reviews and sales; however, even though ratings are largely used by consumers to infer product quality and make consumption choices, there is surprisingly limited understanding of how product/service quality affects the likelihood of receiving ratings, as well as how ratings reflect quality. We address this gap and contribute to the existing knowledge about the value of online ratings to consumers (Mudambi and Schuff 2010).

Word-of-Mouth for Physicians

Our research setting, online ratings for physicians, is consequential for both policy and practice. Within the professional services spectrum, healthcare is the largest sector of the U.S. economy, accounting for nearly 18 precent of the U.S. GDP in 2010. More recently, healthcare has experienced a striking growth in online WOM (Gao et al. 2012; Lagu et al. 2010). These ratings may help resolve a crucial information asymmetry as, prior to their emergence, very little information about the quality of an individual physician made its way into the public domain.[2] From the consumer's perspective, the inability to learn about the characteristics and qualities of a physician prior to engaging in a "service" encounter is limiting, and can result in erroneous physician choice. Given

[2]To address the lack of transparency in physician quality, the federal government has initiated programs such as Physician Compare led by the Centers for Medicare & Medicaid Services. Despite this and other initiatives, consumers continue to have limited access to physician quality information.

the information deficits and the absence of other channels for distinguishing among physicians, it is not surprising that consumer response has been positive and the ratings are experiencing a striking rise in popularity. A survey by Deloitte (2011), for example, found that 47 percent of Internet users had searched information about their healthcare providers online, and 37 percent had visited physician rating websites.

Although consumers have welcomed the availability of online physician ratings, their emergence has generated considerable debate and controversy in the medical community. Most physicians are apprehensive about being rated online by patients (Levine 2009) for a variety of reasons. These include a widely shared concern that online information may be subject to bias, as well as general apprehension that not all rating websites require authentication of the reviewer, thereby raising the possibility of fraud in the reviews posted. In the United States, for example, professional societies such as the American Medical Association (AMA) and some state governments have expressed concern that online ratings merely reflect disgruntled patients' opinions and are not representative of the opinion of the broader population (Colwell 2010). In other words, they suggest that ratings are inherently biased and provide limited utility when informing patients. Fearing that the online ratings may ruin their reputations, some physicians have even resorted to legal action against the rating websites (Solomon 2008). Similar cases of marked disagreement exist internationally as well: in the United Kingdom, when the National Health Service (NHS) enabled a physician rating function on its official website, it created significant disagreement and heated argument (McCartney 2009) regarding the veracity, effectiveness, and content of these reviews. Our research aims to inform the often contentious discussion on the validity of online physician ratings by providing one of the first analyses of how online ratings reflect physician quality as perceived by a broader population of patients.

Hypotheses

Behavioral and Environmental Influences on Online Ratings

The multiple factors that may distort the relationship between online ratings and quality broadly fall into two categories: self-selection and social influence. First, because online ratings are voluntarily posted by users, they may reflect both the heterogeneity in *ex ante* consumer preferences for products, as well as motivations underlying the intentional public disclosure of privately held information. That is, ratings are vulnerable to self-selection that can lead to both over-reporting and under-reporting (Hu et al. 2009). The former arises because product purchases are made by those with higher product valuations, thereby inflating the average rating for a product (Li and Hitt 2008, Godes and Silva 2012), while the latter results from a higher propensity to rate a product among consumers that experience extreme satisfaction or dissatisfaction, as compared with those who like or dislike the product only moderately. The existence of self-selection concerns raises the possibility that online ratings may not accurately reflect true product quality (Hu et al. 2009).

Second, the social context of online reviews often influences individuals' decisions about what to post online. Reviewers who disclose their product judgments do so in a setting where others observe them and use the information to draw conclusions about not only the product, but also the reviewer. This might distort their evaluation and posting behavior, as has been demonstrated in various settings (e.g., Moe and Trusov 2011; Muchnik et al. 2013; Tucker and Zhang 2011; Wang et al. 2010). Thus, social influence could modify the truthful reporting of experienced product or service quality.

Because consumers mainly use the ratings to infer quality, how well online ratings reflect quality has important implications for consumer welfare. Motivated by this fact, our study addresses two research questions related to the relationship between quality and online ratings in the healthcare setting:

(1) How does population-perceived physician quality affect the likelihood of being rated online?

(2) Conditional on being rated, how is population-perceived physician quality reflected in online ratings?

Multiple factors can influence what types of patients are more likely to go online and post reviews of their physicians. Therefore, we derive two competing hypotheses, as detailed below.

First, there is widespread concern among healthcare providers that Internet rating sites for physicians will become forums for disgruntled patients to vent their anger (McCartney 2009; Miller 2007) and criticize their doctors. This phenomenon of "customer complaining behavior" (CCB) has been extensively documented by prior marketing research in both online and offline settings (Bearden and Mason 1979; Cho et al. 2001, 2002). Singh (1988, p. 94) defines CCB as responses "triggered by perceived dissatisfaction with a purchase episode." Prior research has also suggested that the propensity to complain is positively associated with level of dissatisfaction, the importance of the "purchase," the relative cost-benefit ratio of complaining, and consumers' personal competence (Bearden and Mason 1979; Cho et al. 2001, 2002;

Landon 1977). Minor dissatisfaction evokes minimal response, while more severe dissatisfaction is strongly correlated with the consumer response of "telling others" (Richins 1983).

Healthcare is a highly consequential, and deeply personal, service. Health-related interactions and decisions have also been identified as evoking intense emotion, as opposed to only prompting cognition (Anderson and Agarwal 2011). It is plausible that an unsatisfactory physician encounter will elicit a visceral reaction from patients who seek to externalize their frustration. To the extent that physicians who are lower in quality (as judged by patients) are more likely to evoke dissatisfaction, we expect more CCB for this group of physicians. Drawing on this logic we predict

H1a: Physicians of lower quality are more likely to be rated online than average and high-quality physicians.

Second, while the theoretical mechanism of CCB indicates that dissatisfied patients tend to complain more, other evidence suggests that consumers are less likely to speak out about a negative experience. In their study of eBay's feedback mechanisms, Dellarocas and Wood (2008) characterize a reporting bias in which the propensity to reveal private opinion is a function of the outcome experienced by the reviewer. When the outcome is negative, users choose not to provide low ratings for fear of retaliation from trading partners. Empirical results show that eBay traders are significantly more likely to report satisfactory outcomes than those that are moderately unsatisfactory. The limited presence of negative WOM is further corroborated by a number of other studies. Chevalier and Mayzlin (2006), for example, find that consumer reviews of books are strikingly positive at Amazon.com and BarnesandNoble.com. Further, East et al.'s (2007) review of 15 studies found that positive WOM is more pervasive than negative WOM in every case, with positive reviews exceeding negative by a factor of three. Finally, Hu et al. (2006) identify the J-shaped distribution of product ratings as being, similarly, an outcome of a disproportionately large number of positive valuations.

An alternative possibility is that a patient could choose to remain silent after a dissatisfying experience with a physician, yielding a "sound of silence" in online ratings. Theoretically, such an outcome could result from the quintessential nature of the provider-consumer relationship in the healthcare setting. It is well recognized that healthcare is characterized by substantial information asymmetry where the physician, by virtue of specialized knowledge and training, is the "expert" and the patient has limited understanding of medical procedures, diagnoses, and treatments. Moreover, society accords physicians a unique status because of the nature of their work: saving lives (Mishra et al. 2012). In such situations, the patient possesses less power and competence. Unsurprisingly, medical researchers have found evidence that patients often refrain from speaking up about negative experiences, as they believe it may have an adverse effect on their care (Alexander et al. 2004; Blenkinsopp et al. 2007) or even believe they may be perceived as a "bad patient" (Ward et al. 1993).

Unlike the consumption of goods such as movies or books, in which there is typically no personal relationship between sellers and buyers, the patient–physician relationship is often deeply personal and enduring. Insurance requirements and other logistical barriers create significant switching costs to move from one physician to another. In the patient's mind, public reporting of a negative experience may evoke the fear of physician retaliation, much as in the case of trading partners in an online auction, albeit in a different form. Physicians' reprisals could take the form of, at best, interminable delays in getting the next appointment, or, at worst, litigation for public libel. Indeed, in the highly litigious profession of medicine, the latter form of reprisal is not implausible.

While the anonymity of online reviews in this context potentially precludes direct retaliatory action by the physician, significant research has suggested that individuals still fear reprisal in anonymous contexts (Willard 2007). Recent lawsuits by physicians against patients for submission of reviews online (Lawler 2011; Raphael 2009) only underscore the fact that despite the anonymity granted by the Internet, the possibility of reprisals by the physician is present. Such fears may be amplified in contexts where the patient's choice set of physicians is significantly smaller, such as rural areas. Under the assumption that patients experiencing negative physician encounters voluntarily remain silent, we would expect high-quality physicians to be more likely to be rated online. Following this logic, we predict

H1b: Physicians of lower quality are less likely to be rated online than average and high-quality physicians.

Our theoretical arguments for which physicians are more likely to receive an online rating are predicated on prior literature and we acknowledge that other mechanisms may be responsible for the relationship between physician quality and the likelihood of being rated online. For example, low quality physicians might have a lower patient volume, which reduces the probability of being rated online, thereby supporting the relationship described in H1b. Similar to the documented self-selection of customers into purchase behavior (Li and Hitt 2008), patients choose physicians based on their own preferences. It is plausible that patients of low quality physicians are less active online, and therefore are less likely to rate their doctors (again supporting H1b). In addition, healthcare is highly regulated and physicians are required to go through extensive training and rigorous testing before they are eligible to practice. Therefore, it is possible that most physicians provide acceptable services which may not be sufficiently extreme, either positively or negatively, to motivate patients to post a review (Anderson 1998; Hu et al 2009). This mechanism supports H1b. Up to this point, neither literature nor theory has determined which of these hypotheses may be dominant. Our objective is not to isolate a specific mechanisms (as there are many), but rather to provide empirical insight into the overall relationship between physician quality and physician ratings.

Hyperbole Effects

The second effect we examine is related to the discriminatory power of the rating itself, that is, how closely does the valence of the online rating reflect population perceived quality? Substantial prior research has documented the fact that individuals are more likely to talk about extreme experiences (Anderson 1998; Dellarocas and Narayan 2006; Hu et al. 2009). Anderson (1998), for example, argued that the marginal utility a consumer derives from WOM activity increases as satisfaction or dissatisfaction with the product intensifies, resulting in a U-shaped curve for the valence of WOM activity.

What causes individuals to "speak out" after an extreme encounter, either positively or negatively? Theoretically, the sharing of positive experiences has been ascribed to a variety of motivations, including altruism (providing useful information for the benefit of others), instrumental motives (the desire to appear well informed), and general cognitive biases that favor positive incidents (Arndt 1967; Dichter 1966; Holmes and Lett 1977). Motivational mechanisms have also been implicated in high levels of dissatisfaction that give rise to negative WOM. Hu et al. (2006), for example, label this as "moaning" about the experience. Consumers who are highly dissatisfied may seek to express their hostility or even seek retribution (Richins 1983).

Consumers who are highly satisfied or highly dissatisfied are argued to derive homeostatic utility from communicating their experiences: "expressing positive emotions and venting negative feelings" (Hennig-Thurau et al. 2004, p. 44). Homeostatic utility is predicated on the tenets of balance theory, or homeostasis, suggesting that individuals strive for balance and will seek to restore equilibrium after an unbalancing experience. Highly positive and highly negative consumption experiences threaten the individual's sense of balance and motivate them to externalize their feelings by expressing extreme opinions about the consumption experience, which will then restore balance. We label this the *hyperbole effect* in online ratings. As physicians at the ends of the quality spectrum are more likely to evoke extreme opinions, or hyperbole, their ratings should be less informative.

As is the case with the likelihood of being rated online, alternative theoretical processes may contribute to the hyperbole effect, such as the rater's susceptibility to the influence of social context (Moe and Trusov 2011; Wang et al. 2010). Compared to a traditional WOM setting, one prominent feature of online WOM is that when a consumer contributes a review, the opinion is inevitably influenced by existing reviews displayed on the same page, which may lead to conformity (Tucker and Zhang 2011) through cognitive anchoring. If the existing ratings for a physician are predominantly negative or positive, which is likely the case for those of high or low quality, the new ratings are more likely to be consistent with the dominant opinion, in contrast to physicians with mixed or balanced reviews. This will strengthen the hyperbole effect. Therefore, we predict

H2: Online physician ratings exhibit the hyperbole effect, such that the online ratings for high-end and low-end physicians are less correlated with physician quality than they are for average physicians.

Methodology and Results

Data and Measures

We conduct empirical tests of our hypotheses using a novel dataset constructed from four sources of data. First, we use data from the consumer advocacy group Consumers' Checkbook which provides a representative sample of patients' perceptions of *physician quality* (based on patient experience). Consumers' Checkbook employs a rigorous survey methodology utilizing questionnaires and survey approaches developed and tested by the Agency for Healthcare Research & Quality (AHRQ). This measure of physician quality is widely utilized in health services research and is commonly accepted as a reliable indicator of patient perceived physician quality (Safran et al. 2000). Thus, these data provide a standard for understanding potential biases in consumer-generated online ratings (details about the patient survey and the rationale for the use of the physician quality measure are available in Appendix A). Second, our measure for online physician ratings comes from RateMDs.com, one of the nation's largest aggregators of consumers' online physician ratings. Established in 2004, RateMDs.com has accumulated over one million doctor ratings to date. Third, we use the 2007 Quinquennial Economic Census (conducted by the U.S. Census Bureau), which describes the population and economic conditions of individual physician practice areas. Finally, data from state medical boards provides information about physicians' accreditation, licensing, and disciplinary history.

Dependent Variables: Our empirical strategy uses two different dependent variables. For the first set of hypotheses, propensity to receive an online rating, we use a dichotomous measure equal to one if a physician has been *rated online* at least once and zero otherwise. For the second hypothesis, examining correlation between offline perception of physician quality and online ratings, we use the physician's average *online rating* from RateMDs*,* conditional on being rated.

Independent Variables: The primary independent variable of interest is the perceived physician quality measure (*physician quality*) based on patient surveys administered by Consumers' Checkbook. As discussed in Appendix A, this measure captures the mean, debiased, population perception of the physician's quality as rated by his/her patients.

To ensure empirical rigor, we include a robust set of controls to account for unobserved heterogeneity which may bias estimation. The first set of controls is physician-specific characteristics that may influence a physician's propensity to be rated online as well as the score received. These include dichotomous indicators for physician *gender* (equal to one if male and zero otherwise) and physician certification by the state medical board (*board*). Additionally, we control for length of time the physician has practiced medicine, *experience*, operationalized as the number of years elapsed from medical school graduation to the current year and the number of times the physician received endorsements (*peer rating*) from physicians in another Consumers' Checkbook survey.

As environmental factors may also affect online ratings, we include a second set of controls for physician practice locations. These environmental variables are measured at the county level. We include the *population* of the county in which the physician practices. Next, to account for both the geographic and digital competition the physician is facing, we control for the number of physicians who are surveyed in the focal physician's ZIP code, *physician count*, as well as the number of physicians in the focal physician's ZIP code who have received online ratings, *rated physician count*. As a physician's patients may be subject to economic constraints, we control for the *median income* of the county in which the physician practices. Finally, we control for the level of urbanization in the city in which the physician practices (*urban* and *large urban* indicating municipalities with more than 3 or 10 ZIP codes, respectively) as well as dummy variables indicating the metropolitan area in which the Checkbook survey was implemented (*Memphis* and *Denver*; Kansas City is the base category).

To further limit unobserved heterogeneity, we restrict the sample to general practitioners and family care physicians. To the extent that consumer choice related to specialist physicians for particular procedures (such as neurosurgery) occurs less frequently and may invoke a different emotional reaction, we eliminate this source of bias from the data. The resulting dataset comprises 1,425 physicians of which 794 of which have been rated online. Summary statistics are available in Table 2 and a correlation matrix is presented in Table 3.

Analysis

Likelihood of Being Rated Online

Despite the recent growth in online consumer ratings of physicians, a substantial portion of doctors (approximately 44 percent in our sample) have yet to receive an online rating. If the probability of being rated is uncorrelated with physician quality, the absence of a rating would be completely uninformative of physician quality. Alternatively, receiving an online rating could reflect patients' perceptions of physician quality. Hypotheses 1a and 1b offer alternative predictions for how a physician's quality affects the likelihood of being rated online.

We first explore our data visually. Figure 1 plots the *physician quality* distributions for physicians with and without online ratings. On average, *physician quality* for physicians with at least one online rating is higher than that of physicians who have not received an online rating. These descriptive patterns are inconsistent with Hypothesis 1a (that physicians of lower perceived quality are more likely to be rated), but provide initial support for Hypothesis 1b.

Based on the initial evidence from Figure 1, we formally test Hypotheses 1a and 1b by modeling the probability that a physician receives an online rating as

$$\log\left(\frac{\Pr(\textit{Rated Onlie})}{1-\Pr(\textit{Rated Online})}\right) = \alpha_1 + \beta_1 \textit{Physician Quality} + X'\delta_1 + M'\theta_1 + \nu \quad (1)$$

where X and M are vectors of physician and market characteristics, respectively. The terms $\{\alpha_1, \beta_1, \delta_1, \theta_1\}$ are param-

Table 2. Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Online Ratings				
Rated Online (yes = 1)	0.557	0.497	0	1
Online Rating value	4.021	1.215	1	5
Rating Count	1.772	2.749	0	24
Physician Characteristics				
Physician Quality	79.782	8.199	34	96
Board	0.799	0.401	0	1
Gender	0.681	0.466	0	1
Experience	23.968	9.292	5	60
Peer Rating	0.164	0.918	0	12
Market Characteristics				
Population (in thousands)	518.123	273.477	8.969	897.472
Physician Count	15.588	13.79	0	57
Rated Physician Count	7.161	6.321	0	27
Urban	0.535	0.499	0	1
Large Urban	0.403	0.491	0	1
Median Income	47.212	10.912	23.27	82.929
Denver	0.295	0.456	0	1
Memphis	0.269	0.443	0	1

Table 3. Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Online Rating value															
2	Rating Count	-0.093														
3	Physician Quality	0.329	-0.07													
4	Board	0.04	-0.027	0.078												
5	Gender	0.121	-0.069	0.119	-0.082											
6	Experience	0.063	-0.041	0.049	-0.269	0.347										
7	Peer Rating	0.083	0.005	0.144	0.057	0.093	0.059									
8	Population (in thousands)	0.051	-0.003	0.199	0.051	-0.027	0.083	0.052								
9	Physician Count	0.075	-0.012	0.227	0.091	0.011	0.074	0.08	0.532							
10	Rated Physician Count	0.074	0.044	0.236	0.093	0.014	0.044	0.066	0.462	0.853						
11	Urban	-0.018	-0.005	0.074	0.084	0.005	0.036	0.038	0.364	0.112	0.235					
12	Large Urban	-0.002	0.013	0.135	0.05	-0.016	0.062	0.049	0.419	0.193	0.276	0.745				
13	Median Income	0.006	0.164	-0.075	0.043	-0.054	-0.095	-0.011	-0.444	-0.248	-0.124	-0.357	-0.406			
14	Denver	-0.053	0.095	-0.112	0.04	-0.068	-0.019	0.016	-0.29	-0.308	-0.239	0.096	0.029	0.404		
15	Memphis	0.023	-0.043	0.139	-0.07	-0.013	0.091	-0.038	0.638	0.573	0.476	0.057	0.216	-0.421	-0.409	

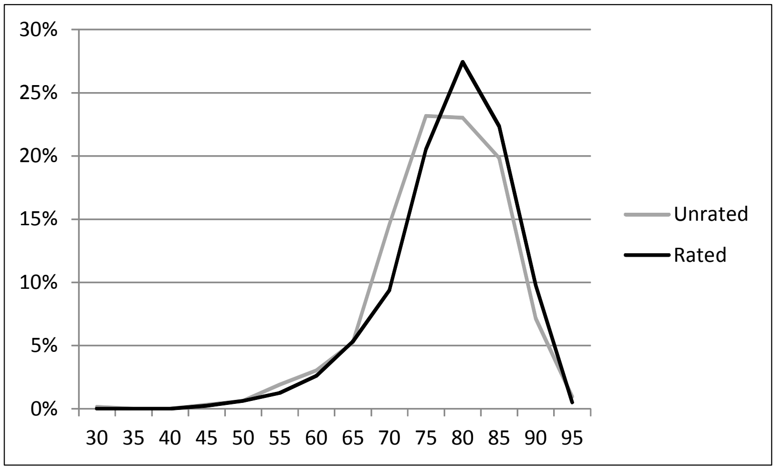


Figure 1. The Distributions of Patient-Perceived Physician Quality: Rated and Unrated Online

eters to be estimated and v represents the error term. The model is estimated via logistic regression with robust standard errors and results are presented in Table 4. As seen in Column 1, *physician quality* is positively and strongly correlated ($p < 0.01$) with the likelihood of rating. The result remains stable across a wide range of alternative specifications, as we show with robustness tests described below. Our finding suggests that as *physician quality* increases there is a corresponding increase in the probability of the physician receiving an online rating, thereby providing support for H1b.

The above estimation assumes that the relationship between *physician quality* and the likelihood of being rated online is constant over the distribution of *physician quality*. To examine the robustness of our results, we employ another specification allowing the probability of receiving an online rating to vary across the lower, middle two (the excluded category), and upper quartiles of *physician quality*. Results are reported in Column 2 of Table 4. Interestingly, we see a significant negative correlation between the lowest segment indicator and the probability of being rated online. The lower quartile coefficient is -0.380, which is statistically significant at $p < 0.001$ level. In contrast, the highest quartile has a coefficient of 0.120 and is not significantly different from the physician with average quality (the default group). Evaluated at the average values of other variables, a physician in the bottom quartile has a 9.43 percentage point reduction in the probability of being rated online compared to an average physician, thereby supporting H1b.[3]

Online Ratings and Patient Perceived Physician Quality

To examine how online ratings relate to population perceptions of physician quality for those physicians who are rated online, we first assess the correlation between online and offline quality ratings. We then formally test for hyperbole effects in the tails of the quality distribution. Figure 2 presents the distributions of *online rating* for physicians with low, medium, and high *physician quality*. On average, higher-quality physicians receive a higher *online rating*. For example, physicians in the highest tercile of *physician quality* overwhelmingly receive an *online rating* of 5. By contrast, high ratings are less prevalent for physicians in the middle and lowest terciles. This offers initial evidence that *online rating* is positively correlated with *physician quality* and that the former provides useful information for consumers.

We employ regression techniques to further study the relationship between online ratings and physician quality. We estimate a series of models based on the following equation:[4]

$$\begin{aligned} Online\ Rating = \alpha_2 + \beta_2 Physician\ Quality + X'\delta_2 + \\ M'\theta_2 + \mu \end{aligned} \quad (2)$$

We first estimate a set of simple models, imposing a linear relationship between *physician quality* and its parameter, β_2. Column 1 of Table 5 reports the OLS estimation of Equation (2). We detect a large, positive, and significant effect of *physician quality* ($p < 0.001$). A 10-point increase in *physician quality* (measured on a 100-point scale) leads to 0.466

[3]The results are consistent with an ordered logit where the dependent variable is the number of reviews received.

[4]As with the logistic regression, each estimation is done with robust Huber-White standard errors to limit the effect of heteroscedasticity

Table 4. Logistic Regression of Probability of a Physician Being Rated Online

	(1) Base Regression	(2) Quartile Regression
Physician Quality	0.0184***	
	(0.00697)	
Lower End of Physician Quality		-0.380***
		(0.132)
Upper End of Physician Quality		0.120
		(0.143)
Denver	0.236*	0.245*
	(0.143)	(0.143)
Memphis	-0.110	-0.114
	(0.169)	(0.169)
Urban	0.385**	0.387**
	(0.186)	(0.186)
Large Urban	-0.205	-0.220
	(0.183)	(0.184)
Population	0.000203	0.000212
	(0.000278)	(0.000278)
Median Income	0.0210***	0.0208***
	(0.00668)	(0.00668)
Physician Count	0.0187**	0.0187**
	(0.00893)	(0.00896)
Rated Physician Count	-0.0151	-0.0149
	(0.0188)	(0.0189)
Experience	-0.00829	-0.00851
	(0.00659)	(0.00659)
Physician Gender	-0.0140	-0.0173
	(0.126)	(0.126)
Board Certification	-0.206	-0.197
	(0.146)	(0.146)
Constant	-2.306***	-0.755*
	(0.662)	(0.411)
N	1,425	1,425

Robust standard errors in parentheses
***p < 0.01, **p < 0.05, *p < 0.1

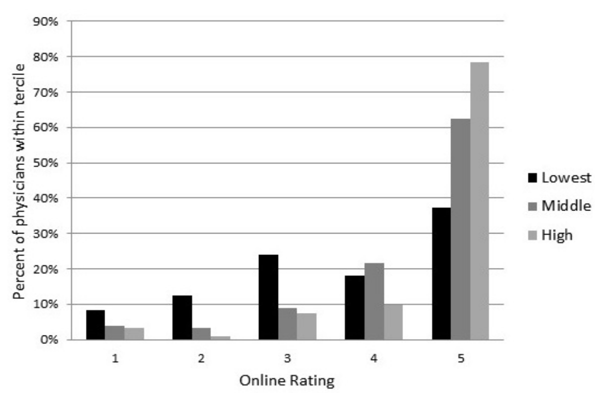


Figure 2. Distribution of Online Ratings by Tercile of Physician Quality

Table 5. Reflection of *Physician Quality* on *Online Rating* Values, Base Specifications

	(1) OLS	(2) Tobit	(3) Ordered Logit	(4) Quantile	(5) Normalized Tobit
Physician Quality	0.0466***	0.0736***	0.0735***	0.0513***	0.604***
	(0.00599)	(0.0101)	(0.00962)	(0.00614)	(0.0826)
Peer Rating	0.0342	0.129*	0.0784	-0.0172	0.129*
	(0.0222)	(0.0703)	(0.0637)	(0.0468)	(0.0703)
Rating Count	-0.0285***	-0.108***	-0.110***	-0.0831***	-0.108***
	(0.0106)	(0.0183)	(0.0176)	(0.0159)	(0.0183)
Denver	-0.0623	-0.170	-0.135	-0.0593	-0.170
	(0.104)	(0.189)	(0.164)	(0.121)	(0.189)
Memphis	-0.0973	-0.261	-0.124	0.0930	-0.261
	(0.160)	(0.290)	(0.249)	(0.163)	(0.290)
Urban	-0.0714	-0.0214	0.0572	-0.00267	-0.0214
	(0.136)	(0.253)	(0.215)	(0.154)	(0.253)
Large Urban	-0.0401	-0.105	-0.117	0.00475	-0.105
	(0.134)	(0.248)	(0.208)	(0.152)	(0.248)
Population	0.000148	0.000379	0.000185	-0.000111	0.000379
	(0.000262)	(0.000464)	(0.000409)	(0.000265)	(0.000464)
Median Income	0.00426	0.00945	0.00763	0.00136	0.00945
	(0.00491)	(0.00884)	(0.00749)	(0.00569)	(0.00884)
Physician Count	-0.000128	0.00347	0.00146	-0.00692	0.00347
	(0.00616)	(0.0124)	(0.0103)	(0.00736)	(0.0124)
Rated Physician Count	0.00214	-0.00399	-0.000325	0.0105	-0.00399
	(0.0133)	(0.0263)	(0.0218)	(0.0157)	(0.0263)
Experience	0.00417	0.00930	0.00863	0.00182	0.00930
	(0.00491)	(0.00925)	(0.00804)	(0.00577)	(0.00925)
Physician Gender	0.178*	0.299*	0.240*	0.174	0.299*
	(0.0970)	(0.169)	(0.144)	(0.108)	(0.169)
Board Certification	0.0717	0.0888	0.0768	0.0592	0.0888
	(0.111)	(0.196)	(0.170)	(0.124)	(0.196)
Constant	-0.107	-1.970**		0.357	3.904***
	(0.558)	(0.951)		(0.603)	(0.603)
N	794	794	794	794	794
R^2	0.125				0.125

Robust standard errors in parentheses

***p < 0.01, **p < 0.05, *p < 0.1

point increase in the *online rating* (measured on a five-point scale). Second, as the range of *online rating* is truncated between 1 and 5 and its distribution is skewed, we employ a Tobit model, reported in Column 2. As expected, the magnitude is even greater in the Tobit estimate (0.0736, $p < 0.01$). This indicates *online ratings* are informative and reflect population opinions. Third, as it is possible that consumers make use of the online ratings by *ranking* physicians rather than using absolute values, we estimate an ordered logit model reported in Column 3 of Table 5. This model yields almost the same finding as that of the Tobit model, both in magnitude and statistical significance. Fourth, to the degree that results may be driven by outliers in the data, we employ a quantile regression (Column 4 of Table 5). This estimation, once again, yields similar estimates in terms of both magnitude and significance. Finally, we normalize the physician quality measure for easier interpretation of the coefficients and reestimate the Tobit model (Column 5 of Table 5). The results show that if a physician is one standard deviation higher, the ratings will increase by 0.604 on a five-point scale. All estimated models provide strong and consistent evidence of a positive relationship between quality and online ratings.

Hypothesis 2 posits that the online ratings will have lower correlations with the underlying physician quality for low or high end physicians. To formally test this hypothesis we examine the relationship between *online rating* and *physician quality* across the *physician quality* distribution. We incorporate splines of *physician quality*, denoted by $f(s_1; \beta_3)$, in a series of regressions (Equation 3). Splines create a piecewise specification where the functions are spliced together at predefined intervals based on the distribution of *physician quality*. These models allow for a flexible relationship and are less sensitive to outliers in the tails of the distribution than polynomials. Knots in the linear spline models are set at the 25th and 75th percentiles of *physician quality*. These models allow for local flexibility without creating discontinuities (Kennedy 2003).

$$Online\ Rating = \alpha_3 + f\left(s_1;\beta_3\right) + X'\delta_3 + M'\theta_3 + \varepsilon \quad (3)$$

The results from models using linear splines are presented in Table 6. Specification 1 of Table 6 is estimated using OLS and generates separate parameter estimates for three segments of *physician quality*: the lower quartile, the middle half, and the top quartile. The *physician quality* coefficient is 0.0395 for the lower quartile, and is significant at $p < 0.01$, suggesting that a 10-point increase in *physician quality* leads to a 0.395-point increase in the *online rating*. Interestingly, the *physician quality* coefficient increases to 0.603 ($p < 0.01$) in the middle half segment. This is 53 percent higher than the coefficient for lower quartile, implying that *online rating* is most sensitive to *physician quality* among average quality physicians. We do not, however, find a significant correlation between *physician quality* and *online rating* in the highest quartile of quality where the coefficient is relatively small ($\beta = 0.0163$, $p = 0.52$). This indicates that online ratings are not sensitive to *physician quality* at the high end of the distribution.

One plausible explanation for the reduced sensitivity in the high end of physician quality is that the online ratings are capped at 5, thus violating the distributional assumptions of OLS. To examine this alternative explanation, we apply a Tobit model to reestimate Equation (3). The results, reported in Column 2 of Table 6, are even sharper. The coefficient of *physician quality* for average physicians is 0.0972 ($p < 0.01$), which is almost twice that of low-end physicians (0.0491). The coefficient of quality for high-end physicians remains statistically insignificant. Overall, the Tobit estimates provide strong support for Hypothesis 2 for doctors with high *physician quality* ratings. Next, we conduct this analysis using the ordered logit model (Column 3, Table 6). The results indicate that the finding is supported not only in the valence of ratings, but also the relative online rankings. Finally, we conduct our analysis using a quantile regression (Column 4) to reduce the effect of outliers. Once again, the results remain consistent.[5]

Identifying hyperbole effects depends not only on the correlation between *online rating* and *physician quality*, but also on the differences between these ratings across the *physician quality* distribution. This requires the joint examination of the intercept, slopes, and confidence intervals. This may be accomplished simply, by comparing the predicted values of *online ratings* across the distribution of *physician quality*, holding constant control variables and adjusting for differences in measurement scales. Such a comparison provides further evidence for Hypothesis 2. In an ideal scenario absent any bias, *online rating* would accurately reflect *physician quality*; thus, it is useful to compare the predicted values and a theoretical scenario with a linear mapping between the two ratings, which would have an intercept of 1 and a slope of 0.04 (as we are mapping physician quality with a 100-point scale, to online ratings with five-point scale). Predicted values and confidence intervals from both the OLS model and the Tobit model are plotted in Figure 3, along with the theoretical perfect information line.

The results in Figure 3 indicate that there is a positive correlation between *online rating* and *physician quality* across the quality spectrum. However, corroborating previous analysis, we observe a marked change in the correlation between *online rating* and *physician quality* (i.e., the slope), conditional on where in the quality spectrum the physician resides. When moving from low quality to average physicians, the slope increases significantly, indicating a stronger sensitivity to changes in *physician quality*. The slope then decreases in the high end segment, suggesting that the online ratings are less sensitive to *physician quality*. Moreover, as indicated by the size of confidence intervals, there is significantly greater noise in the ratings for both low and high quality physicians. Similar patterns are observed in the figure based on Tobit estimates. Note that since the observed online ratings reach a ceiling at the high end boundary of 5 out of 5, consumers will not be able to discern differences among physicians with predicted values exceeding 5. Overall, these results indicate that online ratings offer the greatest discriminatory power for physicians of average quality and are subject to strong hyperbole at either end of the quality distribution.

To further examine the robustness of the spline regression specification, we replicate the above analysis by setting knots

[5]Given the consistently negative and significant coefficient of number of ratings, we have further explored interactions between the number of ratings and *physician quality*. The results remain largely consistent with our reported results and are available from the authors upon request.

Table 6. Reflection of *Physician Quality* on *Online Rating* Values, Allowing Flexibility across *Quality* Splines Set by Physician Quartiles of Quality

	(1) OLS	(2) Tobit	(3) Ordered Logit	(4) Quantile
Spline-Lower End of Physician Quality	0.0395***	0.0491**	0.0497**	0.0671***
	(0.0148)	(0.0224)	(0.0209)	(0.0134)
Spline-Middle Half of Physician Quality	0.0603***	0.0972***	0.0982***	0.0774***
	(0.0130)	(0.0232)	(0.0210)	(0.0142)
Spline-Upper End of Physician Quality	0.0163	0.0666	0.0566	-0.00635
	(0.0233)	(0.0518)	(0.0402)	(0.0301)
Peer Rating	0.0411*	0.126*	0.0772	0.00936
	(0.0229)	(0.0703)	(0.0641)	(0.0463)
Rating Count	-0.0280***	-0.109***	-0.111***	-0.0674***
	(0.0106)	(0.0182)	(0.0173)	(0.0155)
Denver	-0.0579	-0.149	-0.113	-0.0321
	(0.105)	(0.190)	(0.166)	(0.119)
Memphis	-0.0911	-0.281	-0.141	0.0697
	(0.159)	(0.289)	(0.250)	(0.160)
Urban	-0.0634	-0.0196	0.0551	-0.000688
	(0.137)	(0.254)	(0.216)	(0.150)
Large Urban	-0.0474	-0.112	-0.118	0.0149
	(0.134)	(0.247)	(0.209)	(0.148)
Population	0.000142	0.000384	0.000193	-0.000109
	(0.000262)	(0.000464)	(0.000412)	(0.000259)
Median Income	0.00423	0.00900	0.00721	0.000746
	(0.00487)	(0.00882)	(0.00749)	(0.00555)
Physician Count	0.000267	0.00393	0.00216	-0.00155
	(0.00621)	(0.0125)	(0.0104)	(0.00719)
Rated Physician Count	0.00107	-0.00573	-0.00232	0.00176
	(0.0134)	(0.0264)	(0.0220)	(0.0154)
Experience	0.00421	0.00891	0.00839	0.00178
	(0.00494)	(0.00928)	(0.00810)	(0.00563)
Physician Gender	0.178*	0.310*	0.245*	0.0472
	(0.0967)	(0.168)	(0.144)	(0.106)
Board Certification	0.0778	0.0824	0.0731	0.0744
	(0.111)	(0.196)	(0.170)	(0.121)
Constant	0.347	-0.261		-0.904
	(1.101)	(1.686)		(1.016)
N	794	794	794	794
R^2	0.127			

Robust standard errors in parentheses
***p < 0.01, **p < 0.05, *p < 0.1

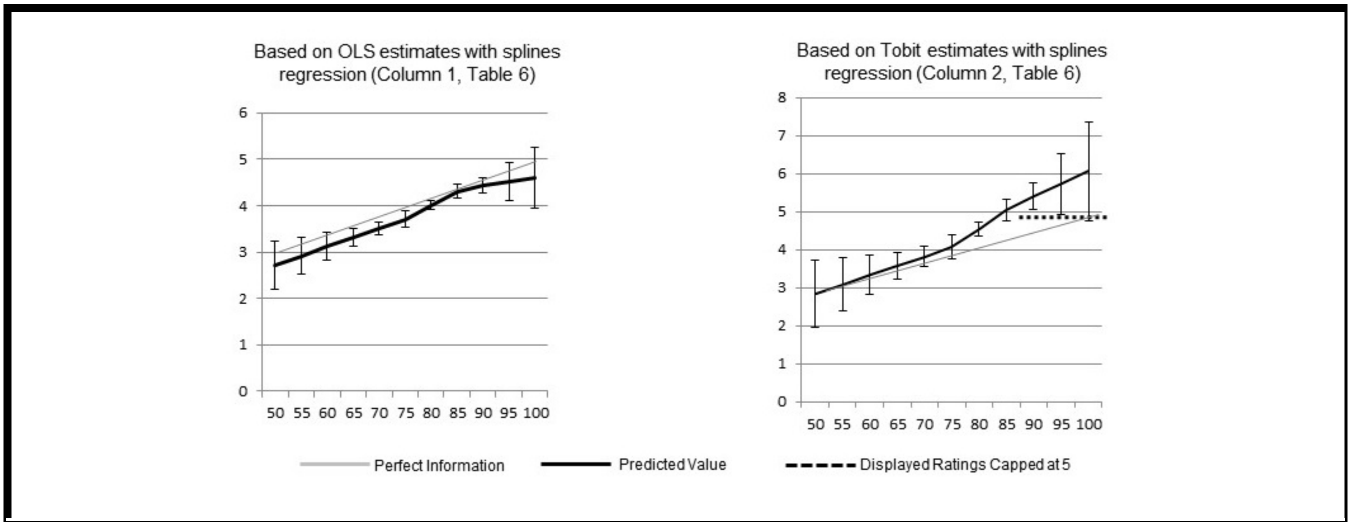


Figure 3. Predicted Online Ratings as a Function of Physician Quality with 95% Confidence Intervals

at the 33^{rd} and 66^{th} percentiles (Table 7), as opposed to quartiles. Once again, the results remain consistent. As expected, the marginal effect of an increase in offline physician ratings is more strongly correlated with an increase in the physician's online rating in the middle of the distribution (i.e., middle third) than either of the tails of the distribution. These results provide further support for H2 and indicate that changes in online ratings are much less sensitive to *physician quality* in the tails of the distribution compared to the middle of the distribution of physicians.

We also explore the insignificant correlation between *physician quality* and online ratings among high quality physicians. It could be argued that high quality physicians are likely to serve more customers and receive a higher volume of patient reviews, which in turn may lead to greater variance among the reviews resulting in the overall rating of the physician not being reflective of physician quality. In other words, it is plausible that larger volume and greater variance is driving the findings, not the hyperbole effect.[6] To eliminate this alternate explanation, we first compare the distribution of rating count across quality levels. The box plot is provided in Figure 4A. As can be seen, the distribution of ratings shows similar patterns across the quality spectrum, and especially so between the middle and upper end physicians. This is further confirmed with a formal t-test, where we find that the average rating count for middle-quality physicians is 2.56, and that for high-end physicians is 2.69. The difference is statistically insignificant ($p = 0.51$). We then examine whether there is significantly more dispersion in the value of online ratings between average and high-end physicians using a box plot (Figure 4B). It is clear that low end physicians receive lower ratings and have the greatest dispersion. The rating dispersion of high end physicians is the smallest among the three, and not statistically different from the middle group ($p = 0.71$). These results further corroborate our proposition that the hyperbole effect leads to the *saturation* of ratings among high-end physicians so that they are less sensitive to the underlying quality.

To summarize, we find that *online rating* is relatively effective at distinguishing physicians with average quality. However, *online rating* has almost no ability to distinguish quality variation within the top quartile of *physician quality* (where most online ratings are capped at 5 out of 5). We further find that *online ratings* exaggerate *physician quality* for physicians exceeding the 75^{th} (66^{th}) percentiles of *physician quality*. These results collectively provide support for Hypothesis 2 and lead us to conclude that online ratings of physicians are subject to the hyperbole effect, notably in the upper tail of the *physician quality* distribution.

Predictive Model

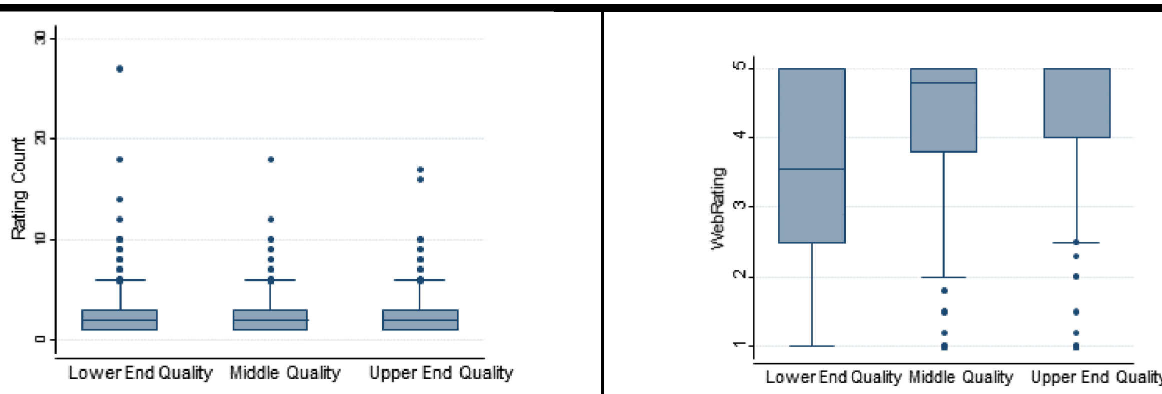
We next explore the relationship between online ratings and physician quality from a consumer's perspective. A consumer may, for example, wish to know how observed online rankings predict physician quality. We construct a predictive model to understand this relationship, measure the signal to noise ratio across the quality distribution, and reexamine hyperbole in this context.

[6]We thank the anonymous reviewer for this point.

Table 7. Reflection of *Physician Quality* on *Online Rating* Values, Allowing Flexibility across *Quality* Splines Set by Physician Terciles of Quality	**(1) OLS**	**(2) Tobit**	**(3) Ordered Logit**	**(4) Quantile**
Spline-Lower Third of Physician Quality	0.0386***	0.0486**	0.0495**	0.0600***
	(0.0140)	(0.0211)	(0.0197)	(0.0118)
Spline-Middle Third of Physician Quality	0.0788***	0.124***	0.124***	0.119***
	(0.0194)	(0.0345)	(0.0311)	(0.0198)
Spline-Upper Third of Physician Quality	0.0183	0.0542	0.0508*	-0
	(0.0155)	(0.0336)	(0.0266)	(0.0183)
Peer Rating	0.0406*	0.125*	0.0759	0.0111
	(0.0229)	(0.0709)	(0.0647)	(0.0420)
Rating Count	-0.0269**	-0.107***	-0.109***	-0.0667***
	(0.0107)	(0.0182)	(0.0173)	(0.0142)
Denver	-0.0591	-0.150	-0.114	-0
	(0.105)	(0.190)	(0.165)	(0.108)
Memphis	-0.0855	-0.271	-0.135	0
	(0.159)	(0.289)	(0.250)	(0.146)
Urban	-0.0555	-0.00932	0.0666	0
	(0.137)	(0.254)	(0.216)	(0.137)
Large Urban	-0.0526	-0.118	-0.123	0
	(0.134)	(0.247)	(0.209)	(0.135)
Population	0.000137	0.000373	0.000183	0
	(0.000261)	(0.000464)	(0.000411)	(0.000236)
Median Income	0.00428	0.00895	0.00707	0
	(0.00486)	(0.00879)	(0.00746)	(0.00507)
Physician Count	0.000647	0.00436	0.00277	0
	(0.00622)	(0.0125)	(0.0104)	(0.00657)
Rated Physician Count	0.000164	-0.00675	-0.00392	-0
	(0.0134)	(0.0265)	(0.0220)	(0.0141)
Experience	0.00403	0.00877	0.00834	-0
	(0.00493)	(0.00927)	(0.00812)	(0.00514)
Physician Gender	0.181*	0.312*	0.249*	0.0667
	(0.0966)	(0.168)	(0.144)	(0.0964)
Board Certification	0.0751	0.0826	0.0715	-0
	(0.111)	(0.196)	(0.170)	(0.111)
Constant	0.388	-0.246		-0.393
	(1.051)	(1.608)		(0.897)
N	794	794	794	794
R^2	0.129			

Robust standard errors in parentheses

***p < 0.01, **p < 0.05, *p < 0.1



4A. Rating Count

4B. Value of Ratings

Figure 4. A Comparison of the Distribution of Online Ratings Across the Quality Spectrum

To conduct this analysis we begin by regressing *physician quality* on a function of *online rating*. We focus on two specifications that regress *physician quality* on *online rating* and the second order polynomial of *online rating*.[7] From these regressions we calculate predicted *physician quality* conditional on *online quality* $(\hat{y})$ and the standard deviation of the prediction error . The regressions allow us to compute the signal-to-noise ratio $\left(\frac{\hat{y}}{\sigma}\right)$.

Table 8 presents results based on quartiles of *online ratings* for both specifications. As expected, *physician quality* increases in online rating: higher online ratings predict higher quality across the rating distribution. Moreover, the noise is higher in the tails; that is, the mean residual is larger in the lowest (highest) quartile than it is in the second lowest (highest) quartile. Furthermore, the signal-to-noise ratio (i.e., the mean amount of information which can be extracted from a change in the *online rating*) is lower in the fourth quartile than in the third quartile. This finding holds for both specifications.[8] This decrease in the signal-to-noise ratio in the upper tail is consistent with finding of hyperbole in the higher tail of the quality distribution.

[7]Note that we explored a variety of alternative specifications. We can reject the hypothesis that the data are better fit by higher-order polynomials and neither linear nor cubic splines could be estimated due to the heavily skewed nature of online ratings. Ultimately, the predictive models do not fit the data as well as our earlier models and the results are sensitive to specification.

[8]It is, however, important to note that the predictive model results are less robust than the results for Equations 1–3 and are sensitive to specification. The noise is a function of the prediction error across the distribution of *online rating*. While the parameter estimates and signals are both stable and consistent, the irregular and skewed distribution of the ratings (e.g., the large number of perfect 5's) yields wide fluctuations in σ.

Robustness

In executing the main analyses presented thus far we took multiple precautions to ensure robustness. First, acknowledging that the relationship between *online rating* and *physician quality* could plausibly be affected by several factors (e.g., the comparison of continuous and Likert scales), we executed a wide range of specifications to ensure that this is not biasing our estimates. Second, we employed several different estimation techniques, as demonstrated in Tables 5, 6, and 7, and estimated alternative spline functions to those reported in Table 6 and 7, choosing, for example, different numbers of knots and changing knot locations. Third, we used alternative specifications of f in Equation 3, including linear and categorical *physician quality* measures. Each specification supported the two crucial features of Figure 3: (1) the predicted line is monotonically increasing in quality and (2) *online rating* is less sensitive to *physician quality* at the high end of quality distribution. However, other potential sources of bias remain.

One source of bias relates to endogeneity, which may manifest in two ways. First, to the extent that physician ratings were available online preceding the Consumers' Checkbook survey, they could affect patients' survey responses. Therefore, rather than online ratings reflecting the physician quality, as we have argued, it is possible that the influence goes in the opposite direction. This is, in essence, the "reverse causality" argument.[9]

We exploit variation in the timing of online ratings to test and correct for this potential endogeneity issue. If indeed online

[9]We thank the anonymous reviewer for pointing this out.

Table 8. Predictive Model of Information Value in Online Ratings

Online Rating	Mean ($\hat{y}$)	Mean (Residual)	Mean (Signal to Noise Ratio)
Linear Model			
Lowest Quartile	76.29112	0.6119493	138.6782
2nd Lowest Quartile	80.0038	0.2896819	278.9981
2nd Highest Quartile	81.7238	0.2868486	285.6919
Highest Quartile	82.43422	0.3266393	252.3708
Quadratic Model			
Lowest Quartile	76.29376	0.7318256	116.9293
2nd Lowest Quartile	79.09261	0.4171059	193.3672
2nd Highest Quartile	81.66637	0.300411	272.1595
Highest Quartile	82.94365	0.3589014	231.1043

Table 9. Logistic Regression of Probability of a Physician Being Rated Online Robustness Checks Based on Survey Release Date

	(1) Removing Physicians Who Received Ratings Before the Survey	(2) Examining Rating Status by the Time of the Survey
Lower End of Physician Quality	-0.379**	-0.319**
	(0.156)	(0.153)
Upper End of Physician Quality	0.195	-0.0463
	(0.165)	(0.159)
Denver	0.130	0.210
	(0.166)	(0.155)
Memphis	-0.156	0.0338
	(0.199)	(0.197)
Urban	0.349	0.342*
	(0.219)	(0.207)
Large Urban	-0.320	-0.0239
	(0.217)	(0.196)
Population	0.000172	0.000202
	(0.000328)	(0.000332)
Median Income	0.0214***	0.0135*
	(0.00773)	(0.00737)
Physician Count	0.0213**	-0.00305
	(0.00990)	(0.00995)
Rated Physician Count	-0.00945	0.00247
	(0.0210)	(0.0208)
Experience	-0.00794	-0.00616
	(0.00781)	(0.00773)
Physician Gender	0.0164	-0.104
	(0.150)	(0.140)
Board Certification	-0.236	-0.0660
	(0.170)	(0.162)
Constant	-1.348***	-1.674***
	(0.478)	(0.470)
N	1,058	1,425

Robust standard errors in parentheses

***p < 0.01, **p < 0.05, *p < 0.1

ratings are affecting the offline survey, then those physicians with zero online ratings before the results of the Consumers' Checkbook survey are released (July 22, 2009) should be less affected. Among the 1,425 physicians, 1,058 have no ratings prior to the survey. We therefore reestimated Equation 1 using the subsample of physicians for whom the Consumers' Checkbook survey preceded their first online rating. The results are presented in Table 9. Parameter estimates for physician quality are consistent with those presented in Table 4. Lower quality physicians remain significantly less likely to receive online ratings compared with average- and high-quality physicians, thereby providing strong evidence for H1b.

The second source of endogeneity relates to receiving the survey from Consumers' Checkbook. Offline perceptions of quality were gathered in the middle of our sample and, therefore, it is possible that the simple act of receiving a survey might affect how the patient rates the physician online. We are able to eliminate this potential confounding factor by setting our sample period to end before the survey was conducted.[10] In other words, we examine whether a physician is rated online before the survey is conducted. In this subsample, all ratings are posted before the survey, so they should not be affected by the survey. Regression results are reported in Column 2 of Table 9 and show a similar effect of quality on the likelihood of being rated online. Overall, we find consistent evidence that low-quality physicians are less likely to be rated online. Moreover, we do not detect a significant difference in the probability of being rated online in other ranges of *physician quality*.

As an additional check, we conduct a two-sample t-test, which compares the online ratings which physicians receive before and after the survey, to examine whether there is any systematic difference in the online ratings. The results are available in Table 10. In the first t-test we compare the mean score of all ratings, across physician, which were submitted before and after the survey was released. Results suggest that there is a modest decrease in the ratings submitted after the survey was conducted, which is marginally significant at the $p < 0.1$) level (Table 10, Panel A). We note that this result might simply reflect the fact that earlier ratings are more likely about physicians with higher quality, as H1b suggested. If we perform a within-subjects comparison of ratings (i.e., constraining our analysis to only physicians who were rated in both time periods), we see no significant effect of survey release ($p = 0.551$), as reported in Table 10, Panel B. Besides aggregated physician ratings, we also perform one further robustness check at the individual rating level. We compare the individual ratings posted 6 months before and 6 months after the survey. As evident in Table 10, Panel C, we see no significant difference ($p = 0.254$).[11] We therefore conclude that the Consumers' Checkbook survey is not significantly influencing online rating behavior.

Discussion and Conclusion

Key Results and Contributions

The importance of online ratings of products and services provided by consumers has been widely acknowledged by both practitioners and researchers; however, significant gaps remain in our understanding of the nature of these ratings. We empirically test the relationship between online ratings and underlying offline perceptions of quality. Based on a review of related literature, we examine how quality is associated with the likelihood of a physician being rated online and, conditional on being rated, how well online ratings reflect physician quality. Our analysis was conducted using a unique dataset of 1,425 primary care physicians in three metropolitan areas: Denver, Kansas City, and Memphis.

This study represents one of the first efforts at quantifying the relationship between online ratings and the underlying quality. We have three notable findings. First, physicians with low patient-perceived quality are less likely to be rated online. Second, we find a positive correlation between online ratings and physician quality. Third, while the association is strongest in the middle of physician quality spectrum, we find an exaggerated relationship between online and offline quality (hyperbole) for physicians with relatively high patient-perceived quality. These results suggest that online ratings are more informative for physicians in average quality segments, but are less effective in differentiating quality among high-end physicians.

We contribute novel findings to the domain of WOM in professional services; an area that has received limited attention from researchers. Given the difficulty of measuring the quality of professional services, online WOM plays a critical role in consumer choices. Further, while most WOM studies focus on how online ratings are associated with sales, we seek to understand the factors that determine the availability of online ratings in the first place. Finally, although prior work has alluded to the presence of a variety of selection biases in online ratings and online WOM (Hu et al. 2006; Li and Hitt 2008), none of these studies have empirically tested

[10]The first wave of surveys was mailed on November 10, 2008, in Denver and Kansas City, and on January 7, 2009, in Memphis.

[11]Robustness checks were conducted using 1-month and 3-month intervals. Results remain consistent.

Table 10. Robustness Check on the Potential Influence of Survey on Online Ratings		
Panel A: Aggregated Physician Ratings Before and After Survey is Mailed (across all physicians who are ever rated)		
	All Ratings	
	Before Survey	**After Survey**
M	3.99	3.82
Σ	1.44	1.48
N	367	451
t-value	1.66	
P-value	0.096	
Panel B: Aggregated Physician Ratings Before and After Survey is Mailed (across physicians who are rated both before and after the survey)		
	Before Survey	**After Survey**
M	3.83	3.73
Σ	1.44	1.44
N	182	182
t-value	0.59	
P-value	0.551	
Panel C: Individual Ratings Posted Before and After the Survey		
	6 months Before Survey	**6 Months After Survey**
μ	3.96	3.86
σ	1.59	1.56
N	186	223
t-value	-0.66	
P-value	0.254	

the correlation between the opinions of online raters and underlying quality. An assumption implicit in the prior literature is that online ratings are reasonable indicators of the quality of products and services. Our results provide empirical support for this assumption. Importantly, we unveil a more complex picture of how online ratings are associated with quality. We are able to identify the areas in the quality distribution where consumers may glean the most, and the least, useful information and insight from differences in ratings. This study is the first to empirically demonstrate the effects of hyperbole on the ability of consumers to ascertain quality.

In addition to expanding our understanding of WOM, this study contributes to the emerging health IT literature. With the infusion of health IT into the consumer domain and the focus on patient-centered healthcare, there is a compelling need to expand the realm of health IT research beyond the hospital setting (Agarwal et al. 2010). One of the most dynamic and rapidly evolving areas in need of exploration is how the Internet, especially Web 2.0 tools, affect health. Studies have examined how patients seek information on social networks (Goh et al. 2012; Yan and Tan 2011). We investigate the other side of the question: how patients decide to share their assessments of physicians online. We believe that the importance of online health information will continue to grow in the coming years, and there is an urgent need for research in this area, especially on issues related to information quality (Agarwal et al. 2010).

Implications for Policy and Practice

Three major implications for policy and practice follow from our findings. First, as noted, the AMA and other physician advocacy groups have been the staunchest opponents of the online aggregation of physician ratings. While their concerns that patients lack the training to evaluate physician quality are yet to be verified and remain a matter of speculation, the concern that ratings aggregation sites will become digital soapboxes for disgruntled patients appears to be unfounded. Not only are the user ratings overwhelmingly positive, but the fact that online reviews reflect offline population opinion suggests that these reviews are informative.

The second implication for policy is the pressing need for a national repository for online ratings. Much of the recent discourse in healthcare transformation is focused around the notion of a patient-centric healthcare system (Krist and Woolf 2011), in which the patient is viewed as an informed and engaged actor in managing his/her healthcare and well-being. To the degree that such a vision demands transparency and easily accessible information to drive patient decision making in consequential areas such as physician choice, and because studies suggest that patients value the opinions of others more than clinical quality measures (Fanjiang et al. 2007), repositories can facilitate important comparisons and trade-offs. These data may complement clinical quality measures the AMA has proposed for the future physician quality reporting system.

Finally, our empirical findings have implications on how to use online ratings in choosing physicians. First, a patient can infer that the expected quality of a physician with no ratings is likely to be low. Second, our analysis confirms that there is a positive correlation between online ratings and quality; however, the correlation gets weaker in the high end. Therefore, the online ratings are more valuable in separating low-end physicians from the average and high-end practitioners. Our findings further caution patients to not ascribe too much significance to minor differences in online ratings among high-end physicians. For example, a rating of 4.9 versus 5.0 might not reflect a meaningful quality difference.

Limitations and Future Work

We point out three limitations of this study that represent fruitful opportunities for future work. First, our findings are based on primary care physicians on one major rating website, and therefore caution must be exercised in generalizing results to apply to other product or service categories, including other forms of professional services. Extending this study to different specialties in healthcare services, such as cardiologists and obstetricians, would provide initial support for the robustness of the findings, and extending the study to additional professional services would help further illuminate the nature of biases in online ratings in this important category of services. While RateMDs.com remains a major source of online doctor ratings (Lagu et al. 2010) , more websites have begun providing this service in recent years. Therefore, it will be interesting for future studies to aggregate doctor ratings from various data sources. In addition, given the relatively small number of ratings for each physician, our online rating data represents the opinions of early contributors (i.e., the vocal minority). The findings might change when a larger number of reviews have been accumulated.

Second, future studies could extend the measure of physician quality. The offline survey was conducted using a scientific approach with sufficient sample size and robust response rates, thereby constituting a reasonable basis for measuring physician quality from a patient perspective. A natural extension of this work would be to examine the relationship between online ratings and physician clinical quality.

Third, the focus of this study is to provide empirical insights into the relationship between online ratings and quality. While we discuss multiple underlying mechanisms, it is beyond the scope of this study to isolate the effect of each mechanism, which we leave to future work. In addition, there are other factors that might complicate the relationship between online ratings and quality. One example is the possibility of fraudulent reviews, which may emerge in our dataset and contribute to the high-end ratings (Newman 2011; Streitfeld 2011). Our study cannot distinguish between fraudulent entries and other drivers of hyperbole. While we are unable to determine the extent to which fraudulent reviews may be driving our findings, their existence underscores the importance of studies such as this one that investigate how meaningful or useful online reviews are for the general population.

Acknowledgments

The authors thank the review team for their useful and constructive comments. They also thank Checkbook.org and RateMDs.com for providing industry background and the essential data. Gordon Gao is partially supported by NSF Career Award #1254021. All errors remain the authors' responsibility.

References

Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. "Research Commentary—The Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.

Alexander, G. C., Casalino, L. P., Tseng, C. W., McFadden, D., and Meltzer, D. O. 2004. "Barriers to Patient-Physician Communication About Out-of-Pocket Costs," *Journal of General Internal Medicine* (19:8), pp. 856-860.

Anderson, C. L., and Agarwal, R. 2011. "The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information," *Information Systems Research* (22:3), pp. 469-490.

Anderson, E. W. 1998. "Customer Satisfaction and Word of Mouth," *Journal of Service Research* (1:1), pp. 5-17.

Arndt, J. 1967. "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research* (4:3), pp. 291-295.

Arrow, K. J. 1963. "Uncertainty and the Welfare Economics of Medical Care," *The American Economic Review* (53:5), pp. 941-973.

Bass, F. 1969. "A New Product Growth Model for Consumer Durables," *Management Science* (15:5), pp. 215-227.

Bearden, W. O., and Mason, J. 1979. "Elderly Use of In-Store Information Sources and Dimensions of Product Satisfaction/ Dissatisfaction," *Journal of Retailing* (55:1), pp. 79-91.

Berger, J., and Iyengar, R. 2012. "How Interest Shapes Word-of-Mouth Over Different Channels," Marketing Department, University of Pennsylvania (available at SSRN: http://ssrn.com/ abstract=2013141 or http://dx.doi.org/10.2139/ssrn.2013141).

Blenkinsopp, A., Wilkie, P., Wang, M., and Routledge, P.A. 2007. "Patient Reporting of Suspected Adverse Drug Reactions: A Review of Published Literature and International Experience," *British Journal of Clinical Pharmacology* (63:2), pp. 148-156.

Brown, J., Broderick, A. J., and Lee, N. 2007. "Word of Mouth Communication Within Online Communities: Conceptualizing the Online Social Network," *Journal of Interactive Marketing* (21:3), pp. 2-20.

Brown, J. J., and Reingen, P. H. 1987. "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research* (14:3), pp. 350-362.

Chen, P.-y., Wu, S.-y., and Yoon, J. 2004. "The Impact of Online Recommendations and Customer Feedback on Sales," in *Proceedings of the 25th International Conference on Information Systems*, R. Agarwal, L. Kirsch, and J. I. DeGross (eds.), Washington, DC, December 12-15, pp. 711-723.

Chen, Y., and Xie, J. 2008. "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science* (54:3), pp. 477-491.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. 2010. "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science* (29:5), pp. 944-957.

Cho, Y., Im, I., Hitlz, S., and Fjermestad, J. 2001. "Causes and Outcomes of Online Customer Complaining Behavior: Implications for Customer Relationship Management," in *Proceedings of the Seventh Americas Conference on Information Systems*, pp. 900-907.

Cho, Y., Im, I., Hiltz, R., and Fjermestad, J. 2002. "An Analysis of Online Customer Complaints: Implications for Web Complaint Management," in *Proceedings of the 35th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press.

Christianson, J., Volmar, K., Alexander, J., and Scanlon, D. 2010. "A Report Card on Provider Report Cards: Current Status of the Health Care Transparency Movement," *Journal of General Internal Medicine* (25:11), pp. 1235-1241.

Clemons, E. K., Gao, G. G., and Hitt, L. M. 2006. "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems* (23:2), pp. 149-171.

Colwell, J. 2010. "Should Doctors Worry About Online Ratings?," *ACP Internist* (30:9), pp. 1, 16. (http://www.acpinternist.org/ archives/2010/09/online.htm).

Dellarocas, C., Awad, N., and Zhang, X. 2004. "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," in *Proceedings of the 25th International Conference on Information Systems*, R. Agarwal, L. Kirsch, and J. I. DeGross (eds.), Washington, DC, December 12-15, pp. 379-386.

Dellarocas, C., Gao, G., and Narayan, R. 2010. "Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?," *Journal of Management Information Systems* (27:2), pp. 127-157.

Dellarocas, C., and Narayan, R. 2006. "A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth," *Statistical Science* (21:2), pp. 277-285.

Dellarocas, C., and Wood, C. A. 2008. "The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias," *Management Science* (54:3), pp. 460-476.

Deloitte Development LLC. 2011. "2011 Survey of Health Care Consumers in the United States: Key Findings, Strategic Implications," Deloitte Center for Health Solutions (available at http://unitedstates/local20assets/documents/us_chs_2011consumersurveyglobal_062111.pdf; accessed January 16, 2015).

Dichter, E. 1966. "How Word-of-Mouth Advertising Works," *Harvard Business Review* (44:6), pp. 147-161.

Duan, W., Gu, B., and Whinston, A.B. 2008. "The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry," *Journal of Retailing* (84:2), pp. 233-242.

East, R., Hammond, K., and Wright, M. 2007. "The Relative Incidence of Positive and Negative Word of Mouth: A Multi-Category Study," *International Journal of Research in Marketing* (24:2), pp. 175-184.

Fanjiang, G., von Glahn, T., Chang, H., Rogers, W., and Safran, D. 2007. "Providing Patients Web-Based Data to Inform Physician Choice: If You Build It, Will They Come?," *Journal of General Internal Medicine* (22:10), pp. 1463-1466.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.

Gallagher, P., Ding, L., Ham, H. P., Schor, E. L., Hays, R. D., and Cleary, P. D. 2009. "Development of a New Patient-Based Measure of Pediatric Ambulatory Care," *Pediatrics* (124:5), pp. 1348-1354.

Gao, G. G., McCullough, J. S., Agarwal, R., and Jha, A. K. 2012. "A Changing Landscape of Physician Quality Reporting: Analysis of Patients' Online Ratings of Their Physicians Over a 5-Year Period," *Journal of Medical Internet Research* (14:1), p. e38.

Godes, D., and Mayzlin, D. 2004. "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science* (23:4), pp. 545-560.

Godes, D., and Silva, J. C. 2012. "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science* (31:3), pp. 448-473.

Goldstein, E., Farquhar, M., Crofton, C., Darby, C., and Garfinkel, S. 2005. "Measuring Hospital Care from the Patients' Perspective: An Overview of the CAHPS® Hospital Survey Development Process," *Health Services Research* (40:6, Pt. 2), pp. 1977-1995.

Goh, J. M., Gao, G., and Agarwal, R. 2012. "The Creation of Social Value: Can Online Communities Reduce Rural Urban Health Disparities?," Working Paper, University of Maryland.

Gu, B., Park, J., and Konana, P. 2011. "Research Note—The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products," *Information Systems Research* (23:1), pp. 182-196.

Gu, B., and Ye, Q. 2014. "First Step in Social Media-Measuring the Influence of Online Management Responses on Customer Satisfaction," *Productions and Operations Management* (23:4), pp. 570-582.

Harris, K., and Buntin, M. 2008. "Choosing a Health Care Provider: The Role of Quality Information," *Policy* (1:14).

Harrison-Walker, L. J. 2001. "The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment as Potential Antecedents," *Journal of Service Research* (4:1), pp. 60-75.

Hennig-Thurau, T., Wiertz, C., and Feldhaus, F. 2004. "Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?," *Journal of Interactive Marketing* (18) pp. 38-52.

Holmes, J., and Lett, J. 1977. "Product Sampling and Word of Mouth," *Journal of Advertising Research* (17:5), pp. 35-40.

Hong, H., Kubik, J. D., and Stein, J. C. 2005. "Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers," *The Journal of Finance* (60:6), pp. 2801-2824.

Hu, N., Pavlou, P. A., and Zhang, J. 2006. "Can Online Reviews Reveal a Product's True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication," in *Proceedings of the 7th ACM Conference on Electronic Commerce*, New York: ACM, pp. 324-330.

Hu, N., Zhang, J., and Pavlou, P. A. 2009. "Overcoming the J-Shaped Distribution of Product Reviews," *Communications of the ACM* (52:10), pp. 144-147.

Jha, A. K., Orav, E. J., Zheng, J., and Epstein, A. M. 2008. "Patients' Perception of Hospital Care in the United States," *New England Journal of Medicine* (359:18), pp. 1921-1931.

Kane, R. L., Maciejewski, M., and Finch, M. 1997. "The Relationship of Patient Satisfaction with Care and Clinical Outcomes," *Medical Care* (35:7), pp. 714-730.

Katz, E., and Lazarsfeld, P. 1955. *Personal Influence*, Glencoe, IL: The Free Press of Glencoe.

Kennedy, P. 2003. *A Guide to Econometrics*, Cambridge, MA: The MIT Press.

Krist, A., and Woolf, S. 2011. "A Vision for Patient-Centered Health Information Systems," *Journal of the American Medical Association* (305:3), pp. 300-301.

Lagu, T., Hannon, N., Rothberg, M., and Lindenauer, P. 2010. "Patients' Evaluations of Health Care Providers in the Era of Social Networking: An Analysis of Physician-Rating Websites," *Journal of General Internal Medicine* (25:9), pp. 942-946.

Landon, L. 1977. "A Model of Consumer Complaint Behavior," in *Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, R. L. Day (ed.), Division of Research, Indiana University, pp. 31-35.

Lawler, C. 2011. "Duluth Doctor Appealing Judge's Decision to Toss Out Defamation Suit," *Duluth News Tribune*, Duluth, MN, June 25 (http://www.duluthnewstribune.com/content/duluth-doctor-appealing-judges-decision-toss-out-defamation-suit-0; accessed January 16, 2015).

Levine, J. 2009. "Welcoming Transparency: Some Physicians Find Merit in Online Ratings," *The Journal of Perinatal Education* (18:3), p. 7.

Li, X., and Hitt, L. M. 2008. "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research* (19:4), pp. 456-474.

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact of Box Office Revenue," *Journal of Marketing* (70:3), pp. 74-89.

McCartney, M. 2009. "Will Doctor Rating Sites Improve the Quality of Care? No," *BMJ* (338:b1033).

Miller, C. 2007. "What Would Hippocrates Do?," in *Forbes Magazine*, August 13, p. 48.

Mishra, A. N., Anderson, C., Angst, C. M., and Agarwal, R. 2012. "Electronic Health Records Assimilation and Physician Identity Evolution: An Identity Theory Perspective," *Information Systems Research* (23:3), pp. 738-760.

Moe, W. W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research* (48:3), pp. 444-456.

Muchnik, L., Aral, S., and Taylor, S. J. 2013. "Social Influence Bias: A Randomized Experiment," *Science* (341:6146), pp. 647-651.

Mudambi, S. M., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly* (34:1), pp. 185-200.

Newman, J. 2011. "Cornell Software Learns How to Spot Fake Online Reviews," *PCWorld*, July 27, p. 57.

O'Malley, A. J., Zaslavsky, A. M., Elliott, M. N., Zaborski, L., and Cleary, P. D. 2005. "Case-Mix Adjustment of the CAHPS® Hospital Survey," *Health Services Research* (40:6), pp. 2162-2181.

Raphael, J. 2009. "Posting Negative Reviews Could Get You Sued," *PC World*, January 7 (http://www.pcworld.com/article/156605/Yelp.html).

Richins, M. L. 1983. "Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study," *The Journal of Marketing* (47:1), pp. 68-78.

Safran, D. G., Rogers, W. H., Tarlov, A. R., Inui, T., Taira, D. A., Montgomery, J. E., Ware, J. E., and Slavin, C. P. 2000. "Organizational and Financial Characteristics of Health Plans: Are They Related to Primary Care Performance?," *Archives of Internal Medicine* (160:1), pp. 69-76.

Senecal, S., and Nantel, J. 2004. "The Influence of Online Product Recommendations on Customers' Online Choices," *Journal of Retailing* (80:2), pp. 159-169.

Singh, J. 1988. "Consumer Complaint Intentions and Behavior: Definitional and Taxonomical Issues," *The Journal of Marketing* (52:1), pp. 93-107.

Solomon, S. 2008. "Doc Seeks $12 Million for RateMDs Comments," *National Review of Medicine* (5:5), Policy & Politics (http://www.nationalreviewofmedicine.com/issue/2008/05/5_policy_ politics03_5.html).

Sun, M. 2012. "How Does the Variance of Product Ratings Matter?," *Management Science* (58:4), pp. 696-707.

Streitfeld, D. 2011. " In a Race to Out-Rave, 5-Star Web Reviews Go for $5," *The New York Times*, Technology, August 19 (http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?_r=0).

Trusov, M., Bucklin, R. E., and Pauwels, K. 2009. "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing* (73:5), pp. 90-102.

Tucker, C., and Zhang, J. 2011. "How Does Popularity Information Affect Choices? A Field Experiment," *Management Science* (57:5), pp. 828-842.

Wang, C., Zhang, X., and Hann, I.-H. 2010. "The Value of Social Dynamics in Online Product Ratings Forums," paper presented at the Workshop on Information Systems Economics, St. Louis, MO.

Ward, S. E., Goldberg, N., Miller-McCauley, V., Mueller, C., Nolan, A., Pawlik-Plank, D., Robbins, A., Stormoen, D., and Weissman, D. E. 1993. "Patient-Related Barriers to Management of Cancer Pain," *Pain* (52:3), pp. 319-324.

Werner, R. M., Greenfield, S., Fung, C., and Turner, B. J. 2007. "Measuring Quality of Care in Patients with Multiple Clinical Conditions: Summary of a Conference Conducted by the Society of General Internal Medicine," *Journal of General Internal Medicine* (22:8), pp. 1206-1211.

Willard, N. E. 2007. *Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress*, Champaign, IL: Research Press.

Yan, L., and Tan, Y. 2011. "Online Conversations and Patients' Perceived Treatment Beliefs Updating," paper presented at the Workshop on Health IT and Economics, Washington DC.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), pp. 133-148.

About the Authors

Guodong (Gordon) Gao is an associate professor in the Decision, Operations, and Information Technology Department at the R. H. Smith School, University of Maryland, College Park. Gordon received bachelor degrees from Tsinghua University, his MBA from the Tsinghua–MIT Sloan Joint Program, and his Ph.D. from the Wharton School of the University of Pennsylvania. His research focuses on the role of IT in transforming healthcare and improving quality transparency. His research has been published in leading journals including *Management Science*, *Information Systems Research*, *Manufacturing and Service Operations Management*, *Journal of Management Information Systems*, *Journal of Medical Internet Research*, and presented at conferences such as the International Health Economics Association Conference, as well as the American Society of Health Economists Conference. He has won several teaching and research awards, including the CAREER award from the National Science Foundation.

Brad N. Greenwood is an assistant professor in the Management Information Systems Department at Temple University's Fox School of Business. He received his doctorate from the Robert H. Smith School of Business at University of Maryland, College Park. His research investigates the role of information availability on decision making in the digital age (with a specific focus on healthcare, public health, and entrepreneurship). His research has been published in leading outlets including *MIS Quarterly* and *Productions and Operations Management* and he has presented his work at the International Conference on Information Systems, INFORMS, the Academy of Management, the Workshop on Information Systems Economics, and the Workshop on Health IT Economics. He also holds a B.S. in Information Technology from Rensselaer Polytechnic Institute, a Master's in Information Technology from Virginia Polytechnic Institute and State University, and an MBA from the University of Notre Dame.

Ritu Agarwal is Professor and the Robert H. Smith Dean's Chair of Information Systems at the Robert H. Smith School of Business, University of Maryland, College Park. She is also the founder and co-director of the Center for Health Information and Decision Systems (CHIDS) at the Smith School. She has published over 90 papers in journals such as *Information Systems Research*, *MIS Quarterly*, *Management Science*, *Communications of the ACM*, *Journal of Management Information Systems*, *Journal of the American Medical Informatics Association, Health Affairs, IEEE Transactions,* and *Decision Support Systems*. Her current research is focused at the intersection of information technology and healthcare, with a particular emphasis on mobile technologies, social media, digital information privacy, and evidence-based decision making. She is a Fellow of Association for Informatoin Systems and the INFORMS Information Systems Society.

Jeffrey S. McCullough is an associate professor in the Division of Health Policy and Management, School of Public Health, at the University of Minnesota. He has a Ph.D. in health economics from the Wharton School and his work focuses on the economics of technology and innovation in health care. In particular, Jeffrey's work has addressed health information technology. His work studies the adoption of health IT and its effects on health care quality and productivity. This work further explores how health IT, health system organization, and clinician effort interact to create value. Other studies explore competition on the federal health insurance exchange, modeling the dynamic consequences of treatment decisions using electronic health record data, and the application of big data techniques to improve risk adjustment and risk classification models.

Appendix A

Construction of Physician Perceived Quality Measure

The *physician quality* measure used in this study is constructed based on patient surveys conducted by an independent, not-for-profit organization, Consumers' Checkbook, in collaboration with five major health plans , viz.Aetna, United Healthcare, CIGNA Health Care, Blue Cross and Blue Shield of Kansas City, and Blue Cross and Blue Shield of Tennessee,[12] in three major metropolitan areas within the United States: Denver, Memphis, and Kansas City. Consumers' Checkbook employs a rigorous survey methodology to ensure the representativeness of the sample and further subjects the data to extensive debiasing procedures, as explained below.

The survey procedure and the survey instrument were designed by the U.S. Agency for Healthcare Research and Quality (AHRQ) to ensure the representativeness of the patient sample for each physician. This is accomplished in the following manner. First, the surveys were sent, on average, to 145 patients of each physician.[13] Physicians who did not possess a minimum of 113 active patients were dropped from the sample. Patients then rated their physician on a variety of different metrics (ranging from attentiveness and responsiveness to thoroughness and knowledgeability). The output of the individual questionnaires is converted to 100-point scale and, following the survey methodology of Safran et al. (2000), patients are pooled by health plan claims data for each of the physicians across major health plans participating in the survey. On average, roughly 50 surveys per physician were returned, a response rate ranging from 32 percent in Denver to 42 percent in Kansas City, yielding coverage of 99.14 percent of the physicians investigated.

While the survey methodology itself ensures a significant patient pool for each physician, one further concern is the self-selection of patients into returning the survey. To the extent that patient-level characteristics can influence both the perception of quality and the likelihood of participating in the survey, these data are subject to an extensive debiasing procedure to remove intrinsic biases that have historically emerged in healthcare survey data. The debiasing procedure employed is the methodology developed by the Consumer Assessments of Healthcare Providers and Systems (CAHPS) (O'Malley et al. 2005) which has been applied extensively in prior work (Gallagher et al. 2009; Goldstein et al. 2005; Jha et al. 2008). At its core, this methodology aims to ensure that patient characteristics do not affect the evaluations of physicians and hospitals. After the debiasing procedure was applied, none of the patient characteristics (age, self-reported health status, race, education, wave of response by the patients, or methodology of response) were statistically significant when the dependent variable, perceived physician quality, was regressed upon them. In summary, the measure used here reflects the current standard for assessing physician quality from the subjective patient experience perspective.

We note that physician quality in this study is a subjective measure based on the experiences of a population of patients. Such a measure is appropriate for this study for a variety of reasons. First, because the measure is based on patient experience, it offers the advantage of being directly comparable to online ratings, allowing us to address our research questions. Second, as discussed, there is considerable controversy and ambiguity surrounding the measurement of physician clinical quality and the appropriate objective metrics for quality assessment (Werner et al. 2007). Lacking suitable clinical training and expertise required to evaluate the true quality of medical care (Arrow 1963), it may be the case that consumer opinions simply reflect the quality of factors such as interpersonal interaction (e.g., bedside manner, communication skills, and punctuality). However, these assessments still constitute useful information for future consumers, and evidence suggests that they guide consumers' choice processes (Fanjiang et al. 2007). Moreover, patient satisfaction with the relationship between them and their healthcare provider has been argued to have positive effects on clinical outcomes (Kane et al. 1997) via multiple mechanisms. These include the ability of the patient to confide in the physician (thereby assisting with proper diagnosis and information transfer), ensuring compliance with prescribed treatment, and the eventual long-term clinical outcome of the patient experience.

[12]Consistent with the approach followed in other studies (Safran et al. 2000), we do not include Medicare patients.

[13]Due to confidentiality agreements, Consumers' Checkbook was unable to disclose to the research team the number of active patients the physician is treating.