The Information Value of Online Physician Ratings

Guodong (Gordon) Gao and Brad Greenwood
Robert H. Smith School of Business
University of Maryland

Jeffrey McCullough
School of Pblic Health
University of Minnesota

Ritu Agarwal
Robert H. Smith School of Business
University of Maryland

May 2011

Working paper: please do not quote or cite without permission

The Information Value of Online Physician Ratings

Abstract

Online ratings by consumers have spread widely on the Internet these days, however little is known about their value in informing consumers. This study examines the informativeness of online ratings in the case of consumer assessment of physicians. Specifically, we seek to examine the magnitude of two potential biases: (1) selection of physicians to rate; and (2) selection of opinions to express. Based on a novel data set, we find that patients are much less likely to talk about physicians with low perceived quality. Additionally, the opinions expressed online tend to be exaggerated compared to offline population opinions. These findings highlight the intrinsic biases associated with online ratings in general.

The Information Value of On-Line Physician Ratings

1. Introduction

In recent years, the Web 2.0 technologies have fostered the rapid growth of online ratings by consumers. This is especially true in the case of physician ratings. More than 40 companies have joined the bandwagon to allow patients to rate physicians online, including Angie's list, healthgrades.com, RateMds.com, Vitals.com, Yelp.com, and Zagat.

The growth of online physician ratings is welcomed by patients as a convenient channel to assess physician quality. Even though there is evidence that physician quality varies substantially (Gawande 2002), up to now there is limited information in the public domain about the quality of individual physicians. Advocates for transparency in healthcare quality have largely devoted their efforts to institutes such as hospitals or nursing homes (Jha et al. 2005, Harris and Buntin 2008) rather than individual doctors. Due to the lack of other channels to discern physician quality, online physician ratings are gaining popularity among patients. The public's demand for this information is striking: 76% of Angie's list users welcome physician rating information; and a recent survey found that 61% of US adults have looked online for health information, and among them 24% have consulted rankings or reviews online of physicians or other providers (PEW, 2009).

Despite the increasing popularity of online physician ratings, most physicians are understandably uneasy about being rated online by patients (Levine 2009). Professional societies such as the American Medical Association and some state governments have expressed concern that these ratings merely reflect unhappy patients' opinions and may ruin physicians' reputations. There are even lawsuits from physicians against these rating websites. Similarly, when the National Health Service (NHS) in the United Kingdom enabled physician rating function on an NHS-run website, it created significant controversy and heated debate (Bacon 2009, McCartney 2009).

The disagreement among policy makers, consumers, and physicians about online physician ratings centers on their value to inform patients about physician quality. Although it is a legitimate concern that these ratings are biased and therefore might do more harm than good, up to now there exists no study examining the information value of these ratings. This study seeks to fill this void, and provides one of the first analyses of the information value of the online physician ratings.

Based on an extensive literature survey, we first identify two biases that online ratings might suffer from: selection of physicians to rate, and selection of opinions to express. We then construct a novel dataset, which combines physician characteristics, their online ratings, and local demographic variables. Importantly, we match individual physician's online ratings to an extensive off-line patient survey. This dataset allows us to assess the above two selection biases while controlling for several confounding factors.

Our findings not only offer important implications for healthcare policy, they also contribute to the research discourse on online word-of-mouth (WOM). Despite the broad availability of online reviews on almost every online retailing website, we have surprisingly little knowledge about the nature of these reviews. This lack of understanding has prevented us from constructing more advanced metrics to associate reviews with sales, designing websites to better facilitate knowledge sharing among consumers, and helping firms manage online consumer reviews (Chen and Xie 2004). The insights generated from this study will shed light on these important issues as well.

2. Background

This study draws upon two streams of prior work: quality transparency in healthcare, and online word-of-mouth.

2.1 Quality transparency in healthcare

Transparent and accurate quality measures are crucial for both the selection and reimbursement of health care providers. Quality is difficult for consumers to measure as outcomes are uncertain, search is difficult, and consumers' typically lack specialized clinical knowledge (Arrow, 1963). Since physicians serve as patients' agents, providing knowledge and judgment, choosing a good physician may have wide-reaching clinical implications.

Despite physicians' central role in health care delivery, consumers have limited access to systematic physician quality information (Jha and Epstein, 2006; Harris and Buntin, 2008; Christianson *et al*., 2010). Traditionally, consumers relied on word-of-mouth references from friends and family when choosing physicians. These references undoubtedly convey measures of service quality and bedside manner but they may be poor predictors of clinical quality.

In the absence of formal quality ratings, the internet has provided a forum for consumers to fill this information vacuum. Nearly 60% of US adults have used online health information resources (Fox and Jones, 2009) – a three-fold increase from 2001 (California Healthcare Foundation, 2001). Furthermore, many consumers use online information to select physicians (Fox and Jones, 2009).

Online consumer ratings of physicians have rapidly grown. As of January, 2010 one in six US physicians were rated by RateMDs.com (Gao *et al*., 2010). While online consumer ratings are popular for many products and services, physician organizations such as the American Medical Association have raised concerns about online ratings of physicians (Dolan, 2008; Martin 2009). Consumers lack detailed clinical information and results may represent a disgruntled minority rather than reflecting average patient experiences. Consumer ratings could, in effect, provide negligible or even misleading information. While recent studies have found that online consumer ratings are predominantly positive (Lagu *et al*., 2010; Gao *et al*., 2010) we have limited understanding regarding the mechanisms through which online ratings occur or the representativeness of these ratings. Understanding the underlying rating process is crucial to determining the meaning and value of online rating systems.

We build on this literature by combining online physician rating data with patient surveys. We match ratings from multiple sources for individual physicians. These data are used to understand what physicians receive online ratings and whether these ratings are representative of the average patient population.

2.2 Online word-of-mouth

Word of Mouth (WOM) has long been recognized as an important determinant of consumer choice (Butler 1923; Katz and Lazarsfeld, 1955, Arndt 1967, Bass 1969). We always ask friends to recommend a restaurant when we move to a new place, and consult users of a product about their experience before buying a similar item. By its nature, word of mouth has been limited to people with close social connections (Brown and Reigen 1987). However, in recent years, the Internet has enabled online communities for consumers to exchange experience and opinions. Recent surveys by two leading online survey companies (Bizrate and DoubleClick) find that 44% of respondents consulted opinion websites before making a purchase, and 59% considered consumer generated reviews more valuable than expert reviews. For some products, product review websites outrank all other media in influencing customer purchase decisions. Consumer-generated ratings on a variety of products and services are now becoming commonplace.

As a result, the past decade has witnessed an increasing number of studies on online reviews by consumers (e.g., Godes and Mayzlin 2003; Dellarocas, Awad and Zhang 2004; Clemons, Gao and Hitt 2006; Duan, Gu and Whinston; Li and Hitt 2008; Dellarocas, Gao, and Narayan 2010, Zhu and Zhang 2010, Moe and Trusov 2010). All these studies have provided important insights on the relationship between online consumer reviews and sales. We note that a majority of these studies focus on establishing the value proposition of digital word-of-mouth by linking consumer reviews with financial performance. In this work, consumer reviews have been examined from various perspectives, such as volume and variance, and from multiple markets including books, CDs, movies, and TV shows. Important as they are, these studies largely focus on understanding how online ratings influence product sales. By contrast, there has been much less attention as to the information value of these ratings, which is the focus of this study.

3. Theory and Hypotheses

3.1 The nature of bias in online reviews

A key distinctive feature of online ratings is that they are voluntarily posted by users, i.e., they reflect intentional public disclosure of privately held information. The existence of a variety of biases in such public reporting of consumption and trading experiences has been acknowledged and documented in the literature. The biases are typically a result of the underlying social motivations and preferences of reviewers.

Reviewers who post their product judgments online do so in a social context where others observe them and possibly use the information to draw conclusions about the reviewer, thereby creating social pressure for the focal reviewer. Studies indicating that perceived social pressure influences ratings beyond what the consumer privately experiences include Moe and Trusov (2010) and Wang, Zhang, and Hann (2010). Moe and Trusov argue that, via mechanisms related to social influence, the valence of previously posted ratings affects future ratings behavior, and that these ratings dynamics, in turn, influence sales. Wang, Zhang, and Hann (2010) posit that social pressure experienced by raters from their online "friends" affects ratings: after controlling for homophily effects, they find that social influence is most potent when earlier ratings by friends are negative. Their study, conducted in the context of books, further reveals that the strength of social influence is conditioned by a variety of factors such as the popularity of the book, stage of the review cycle, experience of the user, and the size of the user's social network. Thus, to the degree that social influence modifies the truthful reporting of experienced product quality, it reduces social welfare by misleading users whose consumption choices are driven by the online reviews.

Researchers have also argued for and demonstrated the existence of temporal effects in ratings. Early experimental work by Schlosser (2005) revealed that the presence of negative reviews caused downward adjustments in product evaluations for future review contributors, causing them to not reveal their true assessments. She implicates such behavior to rater psychology whereby those holding a negative opinion are perceived to be more intelligent and discriminating than those with a positive evaluation. Temporal effects on ratings can be potentially problematic for consumers of the reviews, even when the veracity of the rating is not compromised and it reflects the reviewer's true experience. Arguing that the preferences of those who purchase early are systematically different from late buyers and hence, are a source of bias, Li and Hitt (2008) show that early consumer reviews of books on Amazon.com are overwhelmingly positive, and that subsequent purchases do not make adjustments for this positive bias when drawing inferences about product quality. The self-selection into early purchase caused by *ex ante* differences in preferences then has implications for future demand of the product.

Hu, Pavlou, and Zhang (2010) isolate two biases in online ratings that together result in a J-shaped distribution for product reviews, where an overwhelming majority of ratings are at the most positive end of the scale (5 stars), a small number at the low end (1 star) and very few ratings exist in between the two extremes. Noting that the shape of this distribution is inconsistent with the expectation of a normal curve with a large number of reviews, they characterize the self-selection biases that lead to the J-shaped distribution as a purchasing bias and an under-reporting bias. The former bias, echoing the notion of heterogeneity in *ex ante* consumer preferences for products identified by Li and Hitt (2008), arises because product purchases are made by those with higher product valuations, thereby causing the average rating for a product to be inflated. The under-reporting bias is a result of a higher propensity to rate a product amongst consumers that experience extreme satisfaction or dissatisfaction as compared with those who like or dislike the product only moderately. Hu et al. conclude that the average rating is not an accurate reflection of true product quality and recommend that consumers using reviews to inform their choices examine the entire distribution of the ratings.

To summarize, multiple underlying causes have been implicated in the bias observed in online reviews, including systematic product preference differences among those who choose to provide an online review and those who do not, and the social context within which reviews occur, such as the size of an individual's friendship network, and the valence and volume of prior reviews. To an extent, the presence of such bias is inevitable; however, it is important for researchers and policy makers to understand and circumvent the information loss caused by the bias. The predominant focus in prior work has been on self-selection into providing a review, rather than systematic differences related to the *object* of the review. We address this gap by investigating two biases in the context of physician ratings. The first, a selection bias, is related to consumers' choice about what physicians to rate online. Drawing on prior findings and related theory from marketing and psychology, we offer two alternative predictions with distinct underlying theoretical mechanisms for which physicians receive ratings online: a "sound of silence" versus "bad-mouthing" explanation. The second bias we explore is in the choice of what opinions to express online. We label this the "hyperbole effect." Specific hypotheses related to these biases are developed below.

3.2 "Bad mouthing" vs "Sound of silence" in choosing physicians to rate

In the absence of distortions in user choices about what physicians to rate online, online ratings should reflect broad population opinions about the quality of physicians.

In other words, the average rating that a doctor receives online from a subset of patients should not be significantly different from the rating provided by a broader, representative sample of patients. We note that both the online rating as well as the population rating is a subjective evaluation of quality from the patient's perspective; as discussed earlier, there is considerable controversy and ambiguity surrounding the measurement of physician quality and the appropriate objective metrics for quality assessment. Lacking appropriate clinical training and expertise required to evaluate the true quality of medical care, it may be the case that consumers' opinions simply reflect the quality of factors such as inter-personal interaction, "bedside" manner, communication skills, and timeliness of the healthcare provider. Nonetheless, these assessments constitute useful information for future consumers, and evidence suggests that they are used in consumers' choice processes (Fanjiang et al. 2007).

Prior research in marketing has documented the phenomenon of customer complaining behavior (CCB) in both online and offline settings extensively (e.g., Lee and Cude, 2011; Bearden *et al*., 1979; Cho *et al*., 2001, 2002b). Singh et al. (1988, p. 94) define CCB as "a set of multiple behavioral and non-behavioral responses) some or all of which are triggered by perceived dissatisfaction with a purchase episode." Prior research has also suggested that consumers' propensity to complain is positively associated with their level of dissatisfaction, the importance of the "purchase," the relative benefit-cost ratio of complaining, and their personal competence (Landon, 1977; Bearden et al. 1979, Cho *et al*., 2001, 2002b). Early work by Richins (1983) established that minor dissatisfaction evokes minimal response from consumers, while the severity of dissatisfaction was strongly correlated with the consumer response of "telling others."

To the extent that physicians who are lower in quality as judged by patients are more likely to evoke dissatisfaction, we can expect more CCB for this group of physicians. Indeed, critics worry that the internet rating sites for doctors will be a forum for disgruntled patients to vent their anger (McCartney 2009, Miller 2007). If this "bad-mouthing" assumption is correct, then we would expect to see a disproportionately higher number of low quality physicians being rated online. In this instance, the information value of the ratings for consumers is considerably degraded: even if a physician is rated high compared to others, consumers still need to exercise extra caution when making inferences about the quality of that physician because s/he is the "best of a bad lot." Those posting ratings for these physicians are likely motivated by the fact that the service consumed (a medical interaction) is highly consequential for the reviewer, there are low costs associated with the burden of posting online relative to the satisfaction of public venting, and the reviewer has a positive belief in her ability and competence to evaluate her physician. Thus, we predict:

H1a: Online ratings of physicians are subject to the phenomenon of "bad mouthing" in that physicians of lower quality are more likely to be rated online.

Drawing on findings in CCB, we have argued that dissatisfied patients are more likely to complain. On the other hand, there has been literature suggesting that people are much less likely to speak out about a negative experience (Dellarocas and Wood, 2008), which may yield a "sound of silence" for low quality physicians where individuals prefer to keep their damaging evaluations private. In their study of eBay's feedback mechanisms Dellarocas and Wood (2008) characterize a reporting bias as one where the propensity to reveal private information is a function of the outcome experienced by the reviewer. When the outcome is negative, traders choose not to provide low ratings for fear of retaliation from trading partners. Their 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 word-of-mouth is further corroborated by a number of other studies. Chevalier and Mayzlin (2006) find that consumer reviews of books are strikingly positive at Amazon.com and Barnedandnoble.com. East et al. (2007) review 15 studies and report that positive word of mouth is more pervasive than negative word of mouth in every case, with positive ratings/reviews exceeding negative ones by a factor of 3. Hu, Pavlou, and Zhang's (2009) identification of the J-shaped distribution of product ratings is, similarly, an outcome of a disproportionately large number of positive valuations.

There are a variety of reasons why a patient might be motivated to remain silent after a dissatisfying experience with a physician. It is generally accepted that the patient-physician relationship 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. In such situations, the patient experiences lower power and competence and may attribute the negative experience to her own shortcomings. When such an attribution of blame to self occurs, the patient is less likely to report the negative experience.

Additionally, unlike the consumption of goods such as movies or books where there is typically no relationship between sellers and buyers, the patient-physician relationship is generally deeply personal and enduring. Insurance requirements and other logistical barriers create significant switching costs to move from one physician to another. Public reporting of a negative experience may evoke the fear of physician retaliation in the patient's mind, much as in the case of trading partners in an online auction, albeit of a different form. Physicians' reprisals could be in the form of interminable delays in getting the next appointment at best or, at worst, litigation for public libel. Indeed, in the highly litigious profession of medicine, the latter form of reprisal is not implausible. Under the assumption of this form of behavior where patients experiencing negative physician encounters volitionally choose to remain silent, we would expect high quality physicians to be more likely to be rated online. Following this logic we predict:

H1b: Online ratings of physicians are subject to the phenomenon of "sound of silence" in that physicians of higher quality are more likely to be rated online.

We note that H1a and H1b are specifications of alternative biases that can influence consumers' choices about what physicians to rate online. Each bias derives from a plausible set of motivations driving raters, and prior literature or theory has not identified which bias is more dominant. We examine whether the "sound of silence" effect or "bad-mouthing" effect prevails in online physician ratings.

3.3 Hyperbole effects

The second bias arises from the valence of the rating itself. Substantial prior research has documented that individuals are more likely to talk about extreme experiences (Anderson, 1998; Hu, Pavlou Zhang, 2009, Dellarocas and Narayanm 2006). Anderson (1988) argued that the marginal utility a consumer derives from word-of-mouth activity increases as satisfaction or dissatisfaction with the product increases, resulting in a U-shaped curve for the valence of word of mouth activity. Empirical data from a cross-cultural study of customer satisfaction in the United States and Sweden supports the posited functional form.

Theoretically, the explanation for high levels of satisfaction evoking positive WOM has been ascribed to a variety of reasons, including altruistic motives (sharing useful information for the benefit of others), instrumental motives (the desire to appear well informed), and general cognitive biases that favor positive experiences (Dichter, 1966; Arndt, 1967; Holmes and Lett, 1977). Motivational mechanisms have also been implicated in high levels of dissatisfaction giving rise to negative word of mouth: Hu, Pavlou, and Zhang (2009) label this "moaning" about the experience. Consumers who are highly dissatisfied may seek to express their hostility (Jung, 1959), or seek vengeance and retribution (Richins, 1984). Cognitively, consumers who are highly satisfied or highly dissatisfied are argued to derive homeostase utility from sharing their experiences: "expressing positive emotions and venting negative feelings" (Henning et al. 2004, p. 44). Homestase utility is predicated on the fundamental tenets of balance theory or homeostatis, suggesting that individuals strive for balance and will seek to restore equilibrium after an "unbalancing experience." Highly positive and highly negative (i.e., extreme) consumption experiences threaten the individual's sense of balance and motivates them to externalize their feelings by expressing opinions about the consumption experience, thereby restores balance.

If only patients with extreme opinions are motivated to post ratings, we should expect large variation/dispersion in online opinions compared to the population evaluation. We label this the "hyperbole" effect to suggest that online physician ratings are more exaggerated (both positively and negatively) than what is observed by the population interacting with the physicians. Based on this we predict:

H2: Online physician ratings exhibit the "hyperbole effect" such that there is significantly greater dispersion in the online ratings of physician quality than in the population rating of quality.

4. Data

We draw on several different datasets to conduct our empirical investigation on the existence and magnitude of biases in physician ratings. The first dataset is an offline patient survey conducted by the consumer advocacy group Consumers' Checkbook. Due to data availability, we restrict our dataset to three major metropolitan areas within the United States (Denver, Memphis, and Kansas City). Our second dataset (the online ratings of physicians) is drawn from RateMDs.com. RateMDs is one of the nation's largest aggregators of consumers' online rating of physicians. The third dataset is the 2007 Economic Census (conducted by the US Census Bureau) from which we gather information regarding the population and median income of the area the physicians practice in. The final data source is the state medical board websites from which we draw physician characteristics such as board certification, physician experience, and specialty. To reduce the heterogeneity of the set, and control for the possibility of unobserved variable bias, we conduct our empirical analysis on physicians who are general practitioners (or family care physicians). The resulting dataset is composed of 1513 physicians with offline surveys completed (hereafter referred to as surveys), and 696 physicians who have been rated online (hereafter referred to as webratings). We next describe, and discuss the operationalization, of our specific variables.

A description of each variable can be found in table 1 (Variable Descriptions) with summary statistics in table 2 and correlations in table 3. *SurveyRating* (15) and *Webrating* (18) are the scores the physician receives from the survey administered by Consumers' Checkbook and the website RateMDs respectively. *isRated* (6) is Boolean indicator of if the physician has received a webrating or not. *Lower* (7) and *Upper* (8) are dummy indicators of the physician's presence in the lower or upper quartiles of surveys (respectively) received by Consumers' Checkbook. *Board* (1) is a dummy indicator if the physician is board certified to practice internal medicine by the state medical review board in which the physician practices. *Denver* (2) and *Memphis* (3) are dummy control variables which indicate which of the three metropolitan areas the physician practices in (Kansas City excluded). *Gender* (4) is a dummy control for the sex of the physician (0 being female; 1 being male). *SurveysFilled* (5) indicates the number of patients who completed the surveys administered by Consumers' Checkbook (note that 10 is the minimum number of surveys necessary for a physician to have their information reported by Checkbook). *Median* (9) is a control which indicates the median household income of the county the physician practices in (reported in 1000s). *ofRatings* (10) is the number of webratings a physician has received. *PeerRatings* (11) is an indicator of the number of times the physician has been endorsed by other physicians in the Checkbook survey. *Population* (12) is a control for the size of the county the physician practices in. *rawZero* (13) and *ratedZero* (14) are controls for the number of physicians who are currently practicing in the physician's zip code and the number of those physicians who have received webratings in that same zip code (respectively). *Urban* (16) and *largeUrban* (17) are dummy controls for the size of the town where the focal physician's practice is located. *Urban* areas are those cities with three or more zip codes and *largeUrban* are for those with ten or more zip codes. Finally, *Experience* (19) is a control for the number of years the physician has practiced medicine, operationalized as the difference between the physician's graduation year from medical school and 2010.

5. Analysis

We use a logistic estimation (logit) to determine if a physician has received a webrating and a Tobit model to determine the effect of survey scores on the value of webratings. The logit is selected as the dependent variable for our first empirical investigation is dichotomous and the Tobit model is selected as there is artificial truncation of the dependent variable (*webrating*) at the upper and lower bounds of 5 and 1. Robustness checks of the logit estimation were conducted using a probit regression and the results remain largely consistent (results available upon request). We test our first hypothesis using the following empirical model:

$$\text{Logit (isRated)} = \beta 0 + \beta 1a*\text{SurveyRating} + \beta 2*\text{Experience} + \beta 3*\text{Gender} + \beta 4*\text{SurveysFilled} + \beta 5*\text{Urban} + \beta 6*\text{largeUrban} + \beta 7*\text{rawZero} + \beta 8*\text{ratedZero} + \beta 9*\text{Population} + \beta 10*\text{Median} + \beta 11*\text{Board} + \varepsilon \quad (1)$$

There exists, however, a strong possibility that physicians of different *ex ante* quality have a different propensity to receive a webrating. As suggested by our hypothesis it is likely that physicians of higher or lower quality (compared to the mean) are more likely to have consumers self select into rating them. We therefore expand our model to capture the upper and lower quartiles of physicians whose patients have been surveyed regarding their quality. We test this expansion of the first hypothesis using the following empirical model:

$$\text{Logit (isRated)} = \beta 0 + \beta 1b{*}\text{Lower} + \beta 1c{*}\text{Upper} + \beta 2{*}\text{Experience} + \beta 3{*}\text{Gender} + \beta 4{*}\text{SurveysFilled} + \beta 5{*}\text{Urban} + \beta 6{*}\text{largeUrban} + \beta 7{*}\text{rawZero} + \beta 8{*}\text{ratedZero} + \beta 9{*}\text{Population} + \beta 10{*}\text{Median} + \beta 11{*}\text{Board} + \varepsilon \quad (2)$$

There exists, however, a possibility that those physicians who are in the extremes of the quality quartiles have a different propensity to be rated online than those physicians who are closer to the mean. To reflect this, we extend our model by interacting the survey score with the presence of the physician in the lower or upper quartile:

$$\text{Logit (isRated)} = \beta 0 + \beta 1a{*}\text{SurveyRating} + \beta 1b{*}\text{Lower} + \beta 1c{*}\text{Upper} + \beta 1d{*}\text{Lower}{*}\text{SurveyRating} + \beta 1e{*}\text{Upper}{*}\text{SurveyRating} + \beta 2{*}\text{Experience} + \beta 3{*}\text{Gender} + \beta 4{*}\text{SurveysFilled} + \beta 5{*}\text{Urban} + \beta 6{*}\text{largeUrban} + \beta 7{*}\text{rawZero} + \beta 8{*}\text{ratedZero} + \beta 9{*}\text{Population} + \beta 10{*}\text{Median} + \beta 11{*}\text{Board} + \varepsilon \quad (3)$$

We use the following Tobit model to test the second hypothesis on the relationship between population survey of opinions and webratings:

$$\text{Tobit (Webrating)} = \beta 0 + \beta 1a{*}\text{SurveyRating} + \beta 2{*}\text{PeerRating} + \beta 3{*}\text{ofRatings} + \beta 4{*}\text{Denver} + \beta 5{*}\text{Memphis} + \beta 6{*}\text{SurveysFilled} + \beta 7{*}\text{Urban} + \beta 8{*}\text{largeUrban} + \beta 9{*}\text{rawZero} + \beta 10{*}\text{ratedZero} + \beta 11{*}\text{Population} + \beta 12{*}\text{Median} + \beta 13{*}\text{Board} + \beta 14{*}\text{Experience} + \varepsilon \quad (4)$$

The results of this interaction model are shown in column 1 of Table 5.

As with the first hypothesis it is likely that the correlation between the survey rating and webrating is predicated on the location of the physician within the spectrum of survey ratings. We therefore, as in the first hypothesis, separate out physicians who in the lower and upper quartile:

$$\text{Tobit (Webrating)} = \beta 0 + \beta 1b{*}\text{Lower} + \beta 1c{*}\text{Upper} + \beta 2{*}\text{PeerRating} + \beta 3{*}\text{ofRatings} + \beta 4{*}\text{Denver} + \beta 5{*}\text{Memphis} + \beta 6{*}\text{SurveysFilled} + \beta 7{*}\text{Urban} + \beta 8{*}\text{largeUrban} + \beta 9{*}\text{rawZero} + \beta 10{*}\text{ratedZero} + \beta 11{*}\text{Population} + \beta 12{*}\text{Median} + \beta 13{*}\text{Board} + \beta 14{*}\text{Experience} + \varepsilon \quad (5)$$

5.1 Results

We first discuss the results with respect to the baseline probability of reception of a rating (column 1 of Table 4). As predicted we see a strong and significant correlation ($p<0.05$) between an increase in the survey rating and the probability of a physician receiving a webrating. Dissection of the model into quartiles supports this general assertion (column 2 of Table 4). In this model we see a significant correlation ($p<0.10$) between the lower quartile and a decrease in the probability of a rating being received suggesting general support for H1b and the presence of the sound of silence. The interaction of the quartiles with the survey rating (column 3 of table 4), however, provides more nuanced findings. Here we note that there is no statistically different likelihood of being rated between high or median quality physicians. However, being in the lower quartile has positive and significant effect ($p<0.10$) on the probability of being rated, suggesting general support for H1b. This effect is compounded by the interaction between physician existence in the lower quartile and the survey score ($p<0.10$). As the interaction suggests the increase in the probability of being rated in the lowest quartile is highest at the extreme end of the physician survey rating spectrum, and the increase in the probability of being rated decreases as the physician's survey rating rises. This result indicates support for H1a (bad mouthing), albeit with a limited scope. Differences in the *ex ante* probability of receiving funding are graphically represented in Figures 1 and 2.

We obtain interesting findings for hypothesis 2, the impact of population opinion on webrating. In the baseline model (column 1 of Table 5) we see a strong and significant correlation between the survey rating and the webrating ($p<0.001$), indicating strong support for H2. There also exists a difference in the impact of survey rating based on the location in which the physician resides in the spectrum of physicians, as evidenced by the quartile regressions (column 2 of Table 5). The strong and significant impact of being in the lower quartile ($p<0.001$) indicates an increased correlation between the informativeness of the reviews in the lower quartile and those in the median or upper quartiles. The "hyperbole" effect (H2) therefore appears to be stronger at the lower end.

6. Discussion and Conclusion

In this paper, we examined the information value of online physician ratings. Based on a survey of related literature, we identify two intrinsic sources of bias in online ratings: the selection of physicians to rate, and the selection of opinions to express. We then derived formal hypotheses related to these two biases, which were subsequently tested based on a dataset of 1513 primary care physicians in three metropolitan areas: Denver, Kansas city, and Memphis. This research represents one of the first efforts in quantifying the information value of online ratings.

We have two interesting findings. First, physicians with low survey ratings are less likely to be rated online, therefore supporting the "Sound of Silence" effect in selecting what physicians to rate. However, when the physician quality moves towards the extreme low end of the quality spectrum the probability of rating increases dramatically. Therefore, the "Sound of Silence" effect prevails across the quality spectrum of physicians, while the "Bad Mouthing" effect is evident among very low-end physicians.

Second, we find there is a strong correlation between the online ratings and offline population opinion. Additionally, the association is strongest in the lowest quartile of opinions. These results suggest that online ratings are more informative when identifying low-quality physicians, but not as effective in discerning high quality physicians from median ones. We also find that the "hyperbole effect" to be evident in low-end physicians.

We point out two major limitations of this study. The first is that our findings are based on primary care physicians, and therefore one should be cautious of generalizability towards other categories of products or services. Second, this study examines the relationship between online ratings and offline survey opinions of physicians. Although patient surveys provide important insights into the quality of physician, it will be ideal to extend our measures to the true clinical quality of physicians.

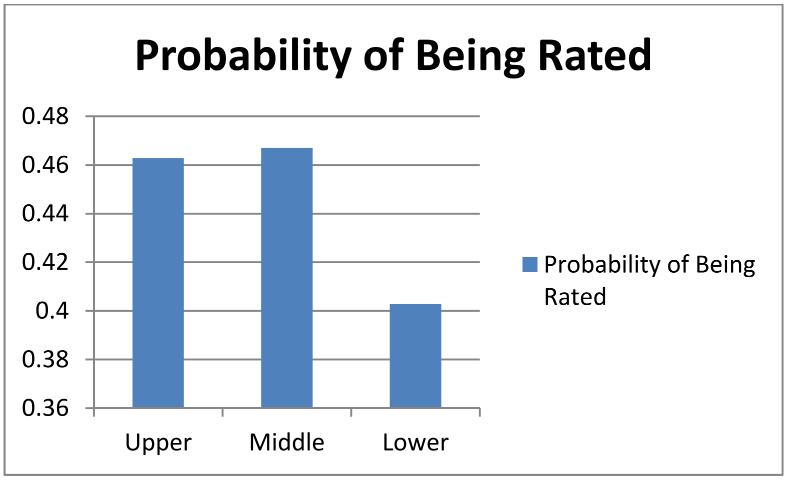

Figure 1: Probability of Webrating based on Quartile

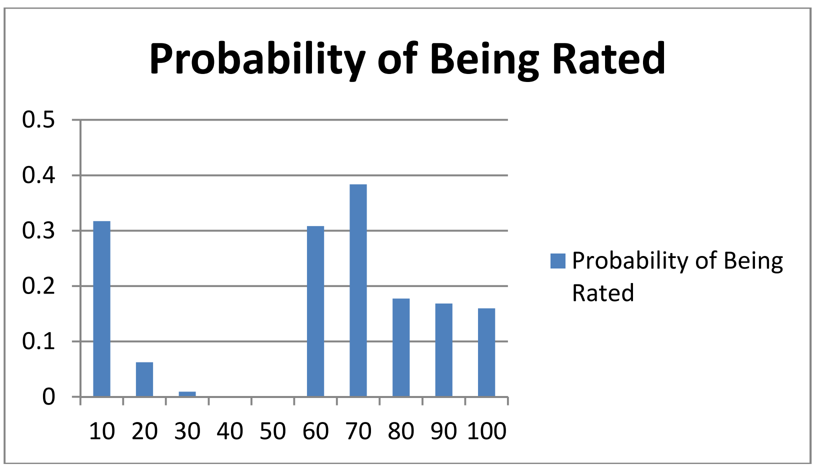

Figure 2: Probability of Webrating based on Quartile and Survey Interaction

Table 1: Variable Definitions

Number	Empirical Name	Specialization Description
1	Board	Board Certification of the Focal Physician
2	Denver	Denver Dummy
3	Memphis	Memphis Dummy
4	Gender	Physician Gender
5	SurveysFilled	Number of Patients to Complete Survey
6	isRated	Is Physician Rated Online (Dummy)
7	Lower	Physician in Lower Survey Quartile
8	Upper	Physician in Upper Survey Quartile
9	Median	Median Household Income for County
10	ofRatings	Number of Webratings Physicians Received
11	PeerRating	Peer Endorsements for Physician
12	Population	Physician County Population
13	rawZero	Number of Physicians in Physician Zipcode
14	ratedZero	Number of Webrated Physicians in Zipcode
15	SurveyRating	Physician Survey Rating
16	Urban	Urban Area Control
17	largeUrban	Large Urban Area Control
18	Webrating	Web Rating
19	Experience	Experience

Table 2: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Board	0.80	0.40	0	1
Denver	0.30	0.46	0	1
Memphis	0.27	0.44	0	1
Gender	0.68	0.47	0	1
SurveysFilled	53.95	16.95	10	105
isRated	0.45	0.50	0	1
Lower	0.22	0.41	0	1
Upper	0.28	0.45	0	1
Median	47.05	10.85	23.27	82.93
ofRatings	1.18	2.14	0	27
PeerRating	0.16	0.90	0	12
Population	520455.60	273311.70	8969	897472
rawZero	15.47	13.65	0	57
ratedZero	7.11	6.33	0	27
SurveyRating	79.67	8.26	34	98
Urban	0.55	0.50	0	1
largeUrban	0.42	0.49	0	1
Webrating	1.84	2.18	0	5
Experience	23.96	9.51	5	60

Table 3: Correlation Matrix

		1	2	3	4	5	6	7	8	9
1	Board									
2	Denver	0.02								
3	Memphis	-0.07	-0.40							
4	Gender	-0.08	-0.04	0.04						
5	SurveysFilled	0.08	-0.20	-0.07	0.10					
6	isRated	0.02	0.07	-0.02	-0.02	0.03				
7	Lower	-0.03	0.07	-0.07	-0.10	-0.32	-0.08			
8	Upper	0.07	-0.11	0.15	0.06	0.34	0.04	-0.31		
9	Median	0.05	0.41	-0.40	-0.07	-0.03	0.09	0.06	-0.07	
10	ofRatings	0.00	0.09	-0.05	-0.05	0.02	0.61	-0.03	0.02	0.13
11	PeerRating	0.06	0.01	-0.03	0.07	0.22	0.05	-0.08	0.13	0.01
12	Population	0.02	-0.23	0.55	-0.01	0.07	0.01	-0.11	0.14	-0.37
13	rawZero	0.12	-0.26	0.50	0.01	0.17	0.02	-0.15	0.19	-0.20
14	ratedZero	0.12	-0.19	0.43	0.00	0.14	0.08	-0.15	0.18	-0.08
15	SurveyRating	0.09	-0.11	0.12	0.10	0.46	0.07	-0.77	0.65	-0.09
16	Urban	0.05	0.07	0.05	0.01	-0.02	0.09	-0.05	0.02	-0.34
17	largeUrban	0.02	0.02	0.20	0.01	-0.03	0.06	-0.09	0.06	-0.38
18	Webrating	0.03	0.05	-0.02	0.01	0.08	0.93	-0.13	0.07	0.07
19	Experience	-0.27	-0.01	0.05	0.35	0.20	-0.04	-0.04	0.04	-0.1

		10	11	12	13	14	15	16	17	18
10	ofRatings									
11	PeerRating	0.03								
12	Population	0.01	0.05							
13	rawZero	0.00	0.07	0.51						
14	ratedZero	0.06	0.05	0.46	0.86					
15	SurveyRating	0.01	0.15	0.17	0.23	0.23				
16	Urban	0.06	0.03	0.39	0.12	0.24	0.07			
17	largeUrban	0.05	0.05	0.43	0.19	0.28	0.12	0.77		
18	Webrating	0.55	0.07	0.02	0.04	0.10	0.14	0.09	0.06	
19	Experience	-0.04	0.05	0.07	0.04	0.02	0.05	0.0751	0.10	-0.04

Table 4: Logit Estimation of a Physician Receiving a Webrating
Dependent Variable: isRated (1 – Reception / 0 – Non reception)

	1	2	3
	Baseline	Quartiles	Baseline and Quartiles Interacted
SurveyRating	0.015 *		0.033
	(0.008)		(0.024)
Lower		-0.261 +	3.956 +
		(0.139)	(2.282)
Upper		-0.013	1.614
		(0.124)	(4.132)
Lower			-0.057 +
* SurveyRating			(0.030)
Upper			-0.020
* SurveyRating			(0.048)
Experience	-0.012 +	-0.011 +	-0.012 +
	(0.006)	(0.006)	(0.006)
Gender	-0.018	0.009	-0.016
	(0.122)	(0.120)	(0.122)
SurveysFilled	0.002	0.006 +	0.002
	(0.004)	(0.003)	(0.004)
Urban	0.467 **	0.410 *	0.479 **
	(0.170)	(0.168)	(0.171)
largeUrban	0.000	0.070	-0.014
	(0.174)	(0.171)	(0.175)
rawZero	-0.017 +	-0.016 +	-0.017 +
	(0.009)	(0.009)	(0.009)
ratedZero	0.049 **	0.047 **	0.049 **
	(0.018)	(0.018)	(0.018)
Population	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Median	0.022 ***	0.024***	0.022***
	(0.006)	(0.006)	(0.006)
Board	-0.113	-0.116	-0.101
	(0.140)	(0.138)	(0.141)
Constant	-2.433 ***	-1.597 ***	-3.900 *
	(0.650)	(0.410)	(1.944)
N	1513	1513	1513
Chi^2	56.24	59.35	61.82

+ significant at 10%; * significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 5: Tobit Estimation of Survey Effects on Webrating
Dependent Variable: Webrating
Lower Bound – 1 Upper Bound – 5

	1	2
	Baseline	Quartiles
SurveyRating	0.063 ***	
	(0.012)	
Lower		-1.123 ***
		(0.219)
Upper		0.216
		(0.196)
Lower * SurveyRating		
Upper * SurveyRating		
PeerRating	0.120	0.127
	(0.092)	(0.090)
ofRatings	-0.116 ***	-0.121 ***
	(0.032)	(0.031)
Denver	-0.277	-0.186
	(0.218)	(0.215)
Memphis	-0.353	-0.271
	(0.293)	(0.289)
SurveysFilled	0.011	0.013 *
	(0.006)	(0.006)
Urban	-0.203	-0.246
	(0.268)	(0.266)
largeUrban	0.136	0.176
	(0.267)	(0.264)
rawZero	-0.007	-0.007
	(0.016)	(0.016)
ratedZero	0.017	0.020
	(0.031)	(0.031)
Population	0.000	0.000
	(0.000)	(0.000)
Median	-0.004	-0.004
	(0.010)	(0.010)
Board	0.090	0.138
	(0.219)	(0.216)
Experience	0.002	-0.001
	(0.010)	(0.009)
Constant	-0.406	4.664 ***
	(1.010)	(0.700)
Σ	1.965	1.961
	(0.084)	(0.083)
N	696	696
Chi^2	82.86	80.81
Left Censored	35	35
Right Censored	312	320

+ significant at 10%; * significant at 5%; ** significant at 1%; *** significant at 0.1%

References

Arrow K, “Uncertainty and the Welfare Economics of Medical Care.” *American Economic Review*. Vol 53, no 5 (1963): 941-973.

Bacon, N. (2009). Will doctor rating sites improve standards of care? Yes. *BMJ*, *338*(mar17 1), b1030-b1030. doi: 10.1136/bmj.b1030.

California Healthcare Foundation. Consumers and health care quality information: need, Availability, Utility: California Healthcare Foundation; 2001.

Christianson JB, Volmar KM, Alexander J, Scanlon DP. “A Report Card on Provider Report Cards: Current Status of the Health Care Transparency Movement.” *Journal of General Internal Medicine*. Vol 25, no 11 (2010): 1235-1241.

Dolan PL. Patients rarely use online ratings to pick physicians. *AMedNewscom* June 23, 2008.

Fox S, Jones S. The social life of health information: Americans' pursuit of health take place within a widening network of both online and offline sources: Pew Internat and American Life Project; 2009.

Gao G. *Working paper* 2010.

Harris KM, Butin MB. Choosing a health care provider: The role of quality information: Robert Wood Johnson Foundation; 2008.

Jha AK, Epstein AM. The predictive accuracy of the New York State coronary artery bypass surgery report-card system. *Health affairs.* Vol. 25, no 3 (2006): 844-55.

Levine, J. (2009). Welcoming transparency: some physicians find merit in online ratings. *The Journal of perinatal education : an ASPO/Lamaze publication*, *18*(3), 3-9. doi: 10.1624/105812409X461153.

Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' Evaluations of Health Care Providers in the Era of Social Networking: An Analysis of Physician-Rating Websites. *Journal of General Internal Medicine.* Vol. 25, no 9 (2010): 942-6.

Martin R. To doctors' dismay, web ratings grow. *St Petersburg Times* May 17, 2009.

McCartney, M. (2009). Will doctor rating sites improve the quality of care? No. *BMJ*, *338*(mar17 1), b1033-b1033. doi: 10.1136/bmj.b1033.