When Similarity Beats Expertise - Differential Effects of Patient and Expert Ratings on Physician Choice
Field and Experimental Study

Kranzbühler, Anne; Kleijnen, Mirella; Verlegh, P.W.J.; Teerling, M.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Original Paper

When Similarity Beats Expertise—Differential Effects of Patient and Expert Ratings on Physician Choice: Field and Experimental Study

Anne-Madeleine Kranzbühler[1], PhD; Mirella H P Kleijnen[2], PhD; Peeter W J Verlegh[2], PhD; Marije Teerling[3], PhD

[1]Department of Product Innovation Management, Delft University of Technology, Delft, Netherlands

[2]Department of Marketing, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[3]Eneco, Rotterdam, Netherlands

Corresponding Author:
Anne-Madeleine Kranzbühler, PhD
Department of Product Innovation Management
Delft University of Technology
Landbergstraat 15
Delft, 2628 CE
Netherlands
Phone: 31 152783451
Email: a.kranzbuhler@tudelft.nl

Abstract

Background: Increasing numbers of patients consult Web-based rating platforms before making health care decisions. These platforms often provide ratings from other patients, reflecting their subjective experience. However, patients often lack the knowledge to be able to judge the objective quality of health services. To account for this potential bias, many rating platforms complement patient ratings with more objective expert ratings, which can lead to conflicting signals as these different types of evaluations are not always aligned.

Objective: This study aimed to fill the gap on how consumers combine information from 2 different sources—patients or experts—to form opinions and make purchase decisions in a health care context. More specifically, we assessed prospective patients' decision making when considering both types of ratings *simultaneously* on a Web-based rating platform. In addition, we examined how the influence of patient and expert ratings is conditional upon rating volume (ie, the number of patient opinions).

Methods: In a field study, we analyzed a dataset from a Web-based physician rating platform containing clickstream data for more than 5000 US doctors. We complemented this with an experimental lab study consisting of a sample of 112 students from a Dutch university. The average age was 23.1 years, and 60.7% (68/112) of the respondents were female.

Results: The field data illustrated the moderating effect of rating volume. If the patient advice was based on small numbers, prospective patients tended to base their selection of a physician on expert rather than patient advice (profile clicks beta=.14, $P<.001$; call clicks beta=.28, $P=.03$). However, when the group of patients substantially grew in size, prospective patients started to rely on patients rather than the expert (profile clicks beta=.23, SE=0.07, $P=.004$; call clicks beta=.43, SE=0.32, $P=.10$). The experimental study replicated and validated these findings for conflicting patient versus expert advice in a controlled setting. When patient ratings were aggregated from a high number of opinions, prospective patients' evaluations were affected more strongly by patient than expert advice ($mean_{patient\ positive/expert\ negative}=3.06$, SD=0.94; $mean_{expert\ positive/patient\ negative}=2.55$, SD=0.89; $F_{1,108}=4.93$, $P=.03$). Conversely, when patient ratings were aggregated from a low volume, participants were affected more strongly by expert compared with patient advice ($mean_{patient\ positive/expert\ negative}=2.36$, SD=0.76; $mean_{expert\ positive/patient\ negative}=3.01$, SD=0.81; $F_{1,108}=8.42$, $P=.004$). This effect occurred despite the fact that they considered the patients to be less knowledgeable than experts.

Conclusions: When confronted with information from both sources simultaneously, prospective patients are influenced more strongly by other patients. This effect reverses when the patient rating has been aggregated from a (very) small number of individual opinions. This has important implications for how to present health care provider ratings to prospective patients to aid their decision-making process.

Introduction

Background

The rise of the internet, with its abundance of information, has changed human decision making [1,2]. To reduce their uncertainty when facing a choice, people have always sought advice from other people [3-5]. Web-based information sources that provide electronic word of mouth have become ubiquitous, which has also influenced people's health care decisions [6-8]. When choosing a physician, for example, many patients consult health care rating platforms to obtain information about the quality of different providers [9,10]. Such platforms may be provided by governments (eg, United States: medicare.gov, United Kingdom: nhs.uk) or companies (eg, healthgrades.com, betterdoctor.com) and allow people to consult other patients' ratings of institutions, physicians, or specific procedures [11]. Parallel to this increased use of ratings by consumers, these ratings have also gained importance in public health systems; for example, in the United States [12,13]. Medicare social insurance program financially penalizes hospitals that perform poorly in their patient ratings [14,15].

Such uses of patient ratings might be questionable though, because health care, similar to other professional services, is complex and marked by credence qualities [16,17]. Even after it has been performed, it is difficult for patients to assess the actual quality of a health service, especially because laypeople often lack the skills needed to evaluate the complex offerings [18]. Accordingly, when prior patients evaluate a physician on a Web-based rating platform, they tend to highlight easily observable features related to the care, such as a physician's convenient location or friendly office staff [7,19], rather than cure-related outcomes [20]. A positive service experience is integral to health care, in that it contributes to psychological and physiological health [21], but such assessments ignore another essential element: the technical quality of health care procedures [13]. Moreover, though consumers often use their service experience as a proxy for technical quality, these factors are at best related, not substitutive, components of the same assessment [22,23]. To address this problem, many rating platforms include outcome-focused expert judgments based on statistics, such as readmission rates, as complementary sources of information. However, little is known about how consumers combine information from the 2 different sources—namely, similar patients or experts (with detailed knowledge)—to form opinions and make purchase decisions.

This study aims to fill this gap. On the basis of source effects literature, we investigated prospective patients' decision making when simultaneously exposed to other patients' and expert ratings on Web-based platforms. Whereas previous research primarily focused on the effects of *either* user or expert ratings on decision making in various contexts (eg, as seen in some studies [24-27]), we focused on situations where both types of ratings are provided *simultaneously*. This design feature of

Web-based rating platforms is not only of increasing practical interest, but it also allows for disagreement between the patient and expert advice, which is of theoretical interest because it forces prospective patients to make a decision about which source they want to follow. The literature on source effects suggests that patients may overly rely on information from similar sources, so that patients may base their decisions more strongly on advice from other patients (who are more similar to them) than from experts, although the latter may be seen as more competent to give advice in the respective situation [27].

We further investigated how the influence of patient and expert ratings is dependent on the number of patient opinions that contributed to the aggregated patient rating (hereafter referred to as "rating volume"). Rating volume has been widely studied in the literature on Web-based reviews and ratings (eg, [11,28-30]), but little is known about how website users make use of this additional cue when simultaneously confronted with different sources of advice. A common design practice of Web-based rating platforms is to present only 1 expert rating, whereas the patient rating typically is based on the aggregated opinion of multiple patients. On the basis of social influence theory, we argue that the influence of patient over expert ratings is dependent on the number of underlying patient opinions that contributed to the overall rating. This is not only an important insight for practitioners who have to decide about the design of such a platform, but it also helps to understand the psychological processes underlying patients' preferences for advice from other patients or experts.

Source Effects

To determine the usefulness of a recommendation, a receiver must evaluate its source [31]. A source's perceived expertise and perceived similarity to the receiver offer the best predictors of its impact on decision making [32,33]. Although other patients can be expected to be perceived as more similar than experts, they most likely lack the experts' detailed knowledge.

Similarity

Source similarity describes the extent to which people resemble one another on certain attributes. Demographic similarity refers to attributes such as gender, age, or socioeconomic status; attitudinal similarity instead pertains to values, attitudes, or lifestyles [32]. Similar sources tend to be particularly influential because they share similar needs and preferences with the receiver and thus deliver relevant information [34,35]. Similarity also leads to greater attractiveness, trust, and understanding [36], which facilitates communication [37]. Therefore, information from similar others is more persuasive and has a stronger influence on decision making than information from dissimilar others (eg, [35,38]). However, socially meaningful similarities can create a perceived bond between people, in addition to the fact that they share a similar experience or situation [39]. In the present health care context, the user might realize that other patients have been in the same situation of

needing a certain treatment. Experts instead might be perceived as being less similar as they do not share the same situation and probably approach the assessment of a service in a more abstract and technical manner.

Expertise

Expertise is the source's "ability to perform product-related tasks successfully" [40]. Perceptions of a source's expertise stem from evaluations of its knowledge, experience, or occupation [41]. In a Web-based environment, this evaluation relies on the limited available cues [42], although most consumers accept an expert source presented on a website and do not investigate the basis of its expertise further [43,44]. The influence of expert sources stems from their expansive knowledge base, compared with nonexpert sources [32,45]. They have greater knowledge about various, alternative offerings because of their enduring involvement in a certain product or service category [46]. In turn, experts' opinions appear to be of high quality, and their advice has a strong, positive impact on receivers (eg, [32,47,48]).

Comparing the Impact of Similarity Versus Expertise

Both perceived expertise and perceived similarity enhance the persuasiveness of a source (eg, [49]), but we do not know which source has the stronger influence: the expert or the similar peer. Research so far compares their impact only indirectly, by employing survey methods or confronting participants with either type. An early study found that consumers are more inclined to follow the advice of a similar but inexperienced salesperson compared with a dissimilar but experienced one [34]. In line with this, expert reviews have been found to reduce consumers' willingness to visit a restaurant's website, whereas consumer reviews increase it [50]. Similarly, when looking for a new doctor, most participants in a US study sought advice from similar sources rather than those with medical expertise [51]. Yet, other studies indicate instead that source expertise is a better predictor of influence [32,52]. A recent meta-analysis affirms a greater impact on sales elasticities from a review provided by an expert rather than a consumer source [53]. Thus, there is conflicting evidence from various contexts, which leads to our main research question:

> *What is the impact of expert and patient ratings when simultaneously provided on a Web-based rating platform?*

The Moderating Role of Rating Volume Information

In the present context of a *simultaneous* presentation of patient and expert ratings, website users might look for additional cues, such as volume information, to help them make inferences about the sources [54,55]. Rating platforms very often not only simultaneously provide star ratings from experts and patients but also inform website users about how many individual patient opinions have been aggregated to compute a rating. However, this ancillary information is usually only given for patient ratings, whereas expert ratings are not further qualified by specific volume information (eg, betterdoctor.com, nhs.uk, and independer.nl). Thus, this additional cue might impact the way website users construe the patient rating in relation to the simultaneously provided expert rating. This design feature leads

to the theoretically profound question—whether information about rating volume (ie, the number of opinions of other patients that underlie a physician rating) may constitute an important moderator for the influence of such ratings on other patients' service evaluations.

Social influence literature states that the likelihood that an individual imitates a certain behavior increases with the number of other individuals who display this behavior [56,57]. The high number of such individuals increases the perceived benefits of adopting a certain behavior compared with its perceived costs [58,59]. Numerical information has been found to be automatically encoded by the human brain and thus easy to process relatively independent of individual differences [60]. The numerical dominance of a majority opinion signals its correctness [61,62]. Humans have been found to be especially prone to the influence from a large number of others when it comes to making purchase decisions (eg, [63-65]). Humans engage in such imitative behavior because a high volume increases the diagnosticity of a message's valence and thus its persuasiveness [29]. This might be due to the fact that high consensus signals that those others might possess a piece of information that is not yet available to the decision maker, (ie, "they must all know something I don't") [63,66]. In line with that, the credibility as well as the impact of a group's judgment on an individual have been found to be positively correlated with the group's size—both Web-based and offline (eg, [28,67]). Hence, rating volume is likely to amplify the effect of advice from similar patients (compared with expert advice) on physician evaluation and decision making. In other words: the larger the number of patients who rate a physician, the more influential the (averaged) patient rating. We thus hypothesize the following:

> *The influence of patient (versus expert) ratings on(1) evaluations and (2) usage intentions of a health service is moderated by rating volume, such that with increasing numbers of underlying patient opinions, the influence of the patient over the expert rating increases.*

Methods

Study 1: Field Study

Study 1 relies on clickstream data from an actual US health care rating platform to assess the impact of simultaneously provided patient and expert ratings as well as the moderating role of explicit rating volume information. The US-based rating platform provides expert and patient ratings on more than 1 million doctors, dentists, and eye doctors throughout the United States to help patients find a suitable service provider. For each doctor, an algorithm calculates an expert rating on the basis of the physician's education, experience, training, and referrals by other doctors. It also displays the average Yelp (patient) rating for each doctor. Both rating types are presented next to each other on overview and search results pages from which patients can access the individual doctor profiles. We assessed the impact of both rating types on 2 behavioral actions: the number of times website users viewed a certain doctor's profile and the number

of times they clicked a button to obtain the doctor's contact information.

Data Collection

We obtained clickstream data for 5299 doctors during May 1 to July 31, 2015 from the rating platform's Web analytics tool and aggregated the click-level data (21,897 profile clicks and 1842 call clicks, see below) to the doctor-level (n=5299). We only included doctors whose profiles featured both expert and patient ratings and whose profiles had been viewed at least once by a prospective patient during the data collection period. We extracted how many times each doctor's profile had been viewed and how many times users clicked to obtain the doctor's contact information. Furthermore, for each doctor, we extracted the expert star rating, the Yelp (patient) star rating, and the rating volume; the overall Yelp rating was based on at the moment of the click. In addition, we collected several control variables (see below).

Measures

We have 2 dependent variables: number of profile clicks and number of call clicks, that is, we counted the number of times each doctor profile was viewed by prospective patients in the data collection period. We also determined how many times visitors to a specific doctor's profile clicked on a button to obtain this doctor's phone number. This behavioral measure provides a proxy for doctor choice.

The independent variables reflected expert and patient ratings. Expert ratings were the professional star ratings provided for each doctor, which ranged from 1 to 5 in 0.5 steps (half stars). The average expert rating was very positive (mean 4.27, SD 0.94). Patient ratings were provided in the form of Yelp star ratings on each doctor's profile. The Yelp ratings could change during the data collection period; we therefore used the respective Yelp rating at the moment of each individual click to calculate an average patient rating for each doctor over the data collection period. Patient ratings also varied from 1 to 5, in 0.5 steps, and the average patient rating in our sample was also positive (mean 3.89, SD 1.21). To be able to test the effect of rating volume, we included the number of individual patient opinions the Yelp star rating was based on in our model. This number could also change over the course of the data collection period; therefore, we calculated an average rating volume for each doctor again (the correlations can be found in Multimedia Appendix 1).

Finally, to isolate the effects of expert and patient ratings on the click-based dependent variables, we included doctor-specific control variables. We controlled for the number of doctor referrals. On the basis of 160 million Medicare referrals from 2009 to 2012, this number indicated how many patients were referred by other doctors to a specific doctor. Doctor referrals only appear on a profile when they are greater than 0, so we included a dummy variable for whether doctors had any referrals or not (0=no, 1=yes). Furthermore, we controlled for each doctor's specialty; the number of practices in which she or he is employed; and whether the doctor has a profile image (0=no, 1=yes), photo gallery (0=no, 1=yes), or Web-based booking system incorporated in the profile (0=no, 1=yes), as well as

whether a doctor has a premium profile (0=no, 1=yes) on the platform.

Model Specification

Most doctors in our sample received few profile or call clicks during the data collection period, and a few doctors had many visitors. Thus, both dependent count variables are overdispersed (mean$_{profile clicks}$ 4.12, variance 85.89; mean$_{call clicks}$ 0.34, variance 3.03), and we therefore modeled both dependent variables with a negative binomial regression [68]. To account for systematic differences between different kinds of doctors, we included a random intercept for each of the 57 doctor specialties:

$$\log(\text{profile clicks}_i) = \alpha_0 + \alpha_j + \beta_1(\text{expert rating}_i) + \beta_2(\text{consumer rating}_i) + \beta_3(\text{consumer rating}_i \times \text{rating volume}_i) + \Omega X_i + \varepsilon_i,$$

where i indexes the doctor and j the doctor specialty, X_i is the vector of doctor-specific control variables, and ε_i is the error. The model specification for our second dependent measure, call clicks, is analogous. For our call clicks model, we controlled for all variables; for the profile clicks model, we only controlled for doctor specialty, practice count, profile image, and premium profile, as the other control variables only become visible to consumers once they have clicked on a doctor's profile (and not on the search results page).

Although the analysis of real clickstream data is useful for getting first insights into the effect of simultaneously provided expert and patient ratings, these are only descriptive in nature. As we are interested in the causal process, we complement our first study with an experimental study in a controlled setting. In study 2, we systematically manipulated expert and patient ratings as well as rating volume. As such, study 2 aimed at finding further support for the causal influence of the number of underlying patient opinions (H1) by replicating the findings from study 1 in an experimental setting.

Study 2: Experimental Study

Procedure and Sample

We employed a 2×2 between-subjects design, in which we manipulated the valences of the expert and patient rating (positive vs negative) and the number of individual opinions the patient rating is based on (high volume vs low volume). We included only manipulations of conflicting ratings (ie, expert positive and patient negative and vice versa-we refer to this variable as "positive source type," as it can be "patient positive" or "expert positive"). Our objective was to compare a very high and a very low rating volume to be able to effectively contrast the respective effects. Drawing from prior research [64], observations of a Dutch health care rating platform (independer.nl), and especially from our study 1 dataset, we set the high rating volume to 142 (among top 1% of number of underlying opinions in our dataset). We set the low rating volume to 3 so that the aggregated patient rating resembled more than just a single but still only very few opinions. We recruited 125 undergraduate students from a Dutch university in exchange for course credit and randomly assigned them to 1 of the 4 experimental conditions in a computer lab. We excluded 13 participants who did not recall that there was a patient and

expert star rating on the site. In our sample (n=112), participants had a mean age of 23.1 years, and 60.7% (68/112) were female.

Participants imagined that they were looking for a hospital to perform a surgery, so they consulted a rating platform to help make their decision (the manipulation can be found in Multimedia Appendix 2). On the following page, the evaluation of a fictional hospital, on a fictional health care rating platform, featured both an expert and a patient star rating. Those ratings were conflicting and either negative (1.5 out of 5 stars) or positive (4.5 out of 5 stars). The patient rating was either based on 3 or 142 individual opinions. Next, participants indicated their attitudes toward the hospital [69,70] and their usage intentions [71]. These variables correlated (r=.83, P<.01), so we combined them into a single evaluative index ("hospital evaluation"; alpha=.96). We also measured participants' evaluations of experts and patients with regard to their expertise (experts: alpha=.94; consumers: alpha=.83 [72]) and trustworthiness (experts: alpha=.95; consumers: alpha=.88 [73]). All these measures featured 5-point scales (the correlations can be found in Multimedia Appendix 1; the reliability of scales are presented in Multimedia Appendix 3).

Results

Study 1: Field Study

Table 1 shows the results of the regression models. The analyses for the profile clicks and call clicks models exclude 5 doctors (profile clicks model) and 1213 doctors (call clicks model) because of missing values for the predictors, resulting in a sample size of 5294 for the profile clicks model and 4086 for the call clicks model. In our 2 initial models (models 1 and 3), we found a significant positive effect of the expert rating on both profile clicks (beta=.13, P<.001) and call clicks (beta=.32, P<.001) and of the patient rating on profile clicks (beta=.03, P=.02) but not on call clicks (beta=.04, P=.76). The interactions of patient rating and rating volume (mean centered [74]) on profile clicks (beta=.01, P<.001) and call clicks (beta=.01, P<.001) were, however, significant. To investigate these interaction effects further and assess the effect of the patient

rating on the dependent variables with different numbers of underlying individual ratings, we repeated the analyses with a mean-centered value for the rating volume plus 3 SDs (models 2 and 4). As expected, the effect of the patient rating increased with the number of individual reviews aggregated to create the rating. As we noted previously, the effects of patient ratings were only significant for profile but not for call clicks for an average rating volume (Yelp rating based on 9 opinions), but their effects were significant and comparable with those of expert ratings when the underlying number of patient opinions was 1 SD higher (Yelp rating based on 27 opinions; profile clicks beta=.14, P<.001; call clicks beta=.28, P=.03). The effects of expert and patient ratings were not significantly different from each other: profile clicks beta=.01, SE 0.04, P=.37; call clicks beta=−.05, SE 0.15, P=.38. When the patient rating was based on the average rating volume plus 3 SDs (Yelp rating based on 63 opinions), the effects on both dependent variables (profile clicks beta=.35, P<.001; call clicks beta=.75, P=.02) were even stronger. With the number of underlying patient opinions increasing to 63, the effect of the patient rating becomes (marginally) significantly stronger than the effect of the expert rating (difference profile clicks beta=.23, SE 0.07, P=.004; call clicks β=.43, SE 0.32, P=.10).

The results of the analysis of field data thus confirm rating volume (ie, the number of underlying patient reviews) to be an important moderator of the impact of patient versus expert advice. Specifically, we find that when confronted with both sources simultaneously, prospective patients tend to base their evaluations of a physician on expert rather than patient advice in case the patient advice is based on small numbers. However, when the group of patients substantially grows in size, prospective patients start to rely more on patients rather than the expert in making a decision, which supports H1. Thus, patients are indeed influenced more strongly by advice from other patients than by advice from experts, but this occurs only when the patient advice is based on a large number of individual opinions. This finding is in line with prior research stating that volume increases the perceived diagnosticity and the effect of Web-based reviews (eg, [28,29]).

Table 1. Regression results of study 1.

Independent variables	Dependent variable: log profile clicks				Dependent variable: log call clicks			
	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Expert rating	0.13[a]	0.02	0.13[a]	0.02	0.32[a]	0.06	0.32[a]	0.06
Patient rating × rating volume (centered at mean)	0.01[a]	0.00	__[b]	—	0.01[a]	0.01	—	—
Patient rating (at mean level of rating volume)	0.03[c]	0.01	—	—	0.04[a]	0.05	—	—
Rating volume	0.00[a]	0.00	—	—	–0.00[a]	0.00	—	—
Patient rating × rating volume (centered at mean + 3 SD)	—	—	0.01[a]	0.00	—	—	0.01[a]	0.01
Patient rating (at mean level of rating volume + 3 SD)	—	—	0.35[a]	0.07	—	—	0.75[a]	0.31
Rating volume	—	—	0.00[a]	0.00	—	—	–0.01[a]	0.00
Controls								
Doctor referrals	—	—	N/A	—	–0.00[a]	0.11	–0.00[a]	0.11
Doctor referrals count	—	—	N/A	—	0.00[b]	0.00	0.00[a]	0.00
Practice count	0.00[a]	0.01	0.00[b]	0.01	0.07[a]	0.02	0.07[a]	0.02
Profile image	0.03[a]	0.03	0.03[b]	0.03	–0.07[a]	0.10	–0.07[a]	0.10
Photo gallery	—	—	—	—	0.64[a]	0.15	0.64[a]	0.15
Web-based booking	0.19[a]	0.03	0.19[a]	0.03	0.05[a]	0.11	0.05[a]	0.11
Premium profile	2.03[a]	0.12	2.03[a]	0.12	0.98[a]	0.44	0.98[a]	0.44
Intercept	0.72[a]	0.07	84[a]	0.09	–3.34[a]	0.29	–3.45[a]	0.30
Log-likelihood	–12910.8	—	—	—	–2257.7	—	—	—
Akaike information criterion	25843.5	—	—	—	4543.4	—	—	—

[a]Significant at the $P<.01$ level.

[b]Not applicable.

[c]Significant at the $P<.05$ level.

Study 2: Experimental Study

To check whether participants indeed perceived the experts and patients in our stimuli as such, we investigated whether expert and patient ratings were perceived differently with regard to expertise and trustworthiness, and whether this depended on the additional rating volume information. Experts were generally perceived as higher in expertise (mean 3.95, SD 0.73) than patients (mean 2.57, SD 0.69; $F_{1,110}=190.33$, $P<.001$), which was independent of the number of underlying patient opinions in the experimental condition ($P>.85$). Trustworthiness did not differ between experts and patients, again independent of the number of underlying patient opinions (all P values>.20).

The results of an analysis of variance indicated no significant main effects of positive source type (patient positive vs expert positive) and rating volume (142 vs 3) on hospital evaluation ($F_{\text{positive source type}1,108}=0.17$, $P=.68$; $F_{\text{volume}1,108}=0.52$, $P=.47$). However, as expected, we found a significant interaction effect of these 2 variables on hospital evaluation ($F_{1,108}=13.06$, $P<.001$). Further analyses demonstrated that when patient ratings are aggregated from a high number of opinions, prospective patients' evaluations are affected more strongly by patient than expert advice (mean$_{\text{patient positive/expert negative}}$ 3.06, SD 0.94; mean$_{\text{expert positive/patient negative}}$ 2.55, SD 0.89; $F_{1,108}=4.93$, $P=.03$). Conversely, when patient ratings are aggregated from a low volume, participants are affected more strongly by expert compared with patient advice (mean$_{\text{patient positive/expert negative}}$ 2.36, SD 0.76; mean$_{\text{expert positive/patient negative}}$ 3.01, SD 0.81; $F_{1,108}=8.42$, $P=.004$; Figure 1). Thus, study 2 finds further support for H1.

Figure 1. Study 2: Interaction effect of positive source type and rating volume on hospital evaluation.
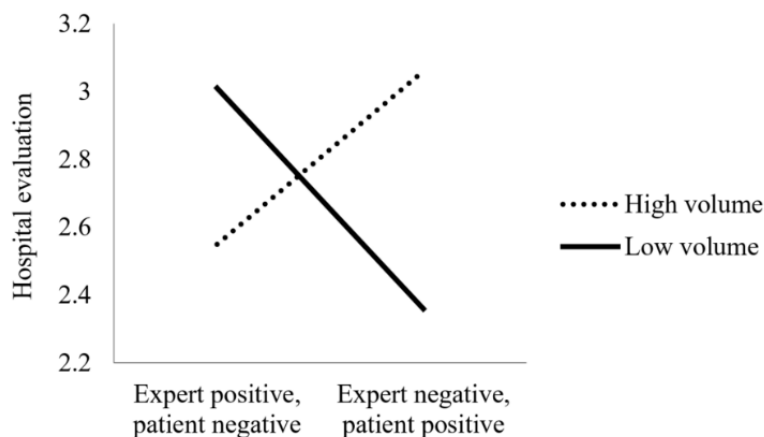
The results of study 2 thus further support the findings from study 1: prospective patients are influenced more strongly by other patients when the patient evaluation is based on a larger number of individual opinions, but not when it is based on only a few observations. When the patient rating has been aggregated from only a small number of individual opinions, website users are instead more inclined to follow the expert advice. However, we find that participants still perceive experts to be higher in expertise, even than a large number of other patients. Moreover, and in contrast to prior findings [75], website users do not seem to trust experts less than other patients or assume any biases or ulterior motives. Consequently, prospective patients seem to be aware of the fact that a large number of other patients still lack the expertise to judge the objective outcome quality of a health service and also do not trust them more than an expert giving a rating. Yet, they are more inclined to follow their peers' compared with an expert's advice when choosing a health service.

Discussion

Principal Findings

This study examined how the evaluations and usage intentions of a health service is influenced by the simultaneous exposure to conflicting patient versus expert ratings and how this relationship is moderated by rating volume. The findings support the hypothesis that with increasing numbers of underlying patient opinions, the influence of the patient over the expert rating increases. In case the patient advice is based on small numbers, prospective patients tend to base their evaluations on an expert rather than patient advice. However, when the group of patients substantially grows in size, prospective patients start to rely more on patients rather than experts in making a decision. These patients are considered to be more similar yet less knowledgeable than experts. Thus, in line with social influence and source effects literature, patients are influenced more strongly by advice from other patients than by advice from experts but only when the patient advice is based on a large number of individual opinions. This has important implications for practitioners who design health care rating platforms. These implications hold special importance for how and which ratings to present to prospective patients to aid their decision-making process. It further helps us to understand the psychological

processes underlying patients' preferences for advice from other patients or experts.

To the best of our knowledge, this study is the first to analyze the impact of the *simultaneous* presence of conflicting ratings from patient and expert sources on patient decision making. Research that provides participants with 1 source offers conflicting evidence for the preference for peer over expert advice (eg, [28,50]), but such studies cannot explicate the disaggregated effects of advice from different sources on a receiver or how consumers decide "which electronic word-of-mouth messages to adopt and which ones to reject" [76]. In 1 field study and 1 experimental study, we shed light on the differential effects of patient and expert advice in a health care context.

Limitations

In providing initial insights into the use of conflicting advice from different sources in patient decision-making, this study features several limitations that provide directions for further research. First, in our experimental study, we did not specify who the "experts" were. Prior research indicates that users mainly look for authority cues when assessing a website's credibility, but they rarely investigate who the expert sources actually are [44]. The participants in our studies perceived the experts as such. However, it would be interesting to explore whether the findings change when the advice comes from various experts, such as family doctors or specialist physicians.

Second, our findings only lay the foundation for understanding how users make a decision when confronted with different sources of information simultaneously. Although our analysis of clickstream data is high in external validity, it only produces correlational results. The sample was further not taken at random and should thus be interpreted with caution. Study 2 can likewise not quantify the exact effects of expert and patient ratings. Future research should investigate this further by making users choose between different offerings. Doing so will enable us to quantify the relative importance of different pieces of information.

Third, we did not analyze the potential moderating roles of other rating and platform characteristics. Platforms tend to feature both numerical ratings and textual reviews [77]. Other important characteristics, including language use, salience of the valence,

or further information about the sources and their reputation [76] thus deserve further inquiry.

Fourth, this study investigated Web-based ratings, one of many potential information sources a consumer employs to make health service decisions. Offline sources such as friends and family also strongly influence health service choice (eg, [78]). Further research might examine the differential effects of offline versus Web-based word of mouth and their interaction across the different stages of the decision-making process.

Conclusions

First, this study enhances the understanding of patients' use of Web-based decision support tools. More specifically, we shed light on the integration of simultaneously provided information from other patients and experts. We found that the opinion of a moderate number of other patients strongly influenced users' evaluations and choices of physicians, even overruling the conflicting opinions of experts. In this sense, choosing a health care provider does not seem to differ much from purchase choices for (for example) movies or restaurants. Even for credence services such as health care, others' subjective ratings of service experiences have a greater impact on decision making than more objective ratings of the service. Although the subjective patient experience is also important, it often fails to

acknowledge the actual outcome quality of the service which patients have difficulties evaluating even after it has been performed [79]. These 2 aspects should be considered complementary for well-informed decisions, but our study suggests that this combination is not the route most consumers take when considering different sources of quality information on a Web-based rating platform. Even in health care contexts, users turn to consumers over expert advice, and this negligence of objective outcome quality information might lead them to make suboptimal choices, with considerable consequences for their health and society at large.

Second, for rating platform providers, our results support the trend of providing advice from both expert and patient sources. Experts and patients differ in their expertise and access to information, so they often emphasize different features in their judgments. These 2 sources of advice therefore should be regarded as complementary instead of substitutive input. Furthermore, the platform providers need to include the number of individual opinions underlying a patient rating because users might assume a rating is based on many opinions if the information is not evident. When this information is present, it acts as an important moderator, preventing an overemphasis on a collection of just a few patient ratings.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The correlations of study 1 and 2.

[PDF File (Adobe PDF File), 152KB - jmir_v21i6e12454_app1.pdf]

Multimedia Appendix 2

Screenshots of a fictional Web-based health care rating platform used as the manipulations in study 2.

[PDF File (Adobe PDF File), 182KB - jmir_v21i6e12454_app2.pdf]

Multimedia Appendix 3

The reliability of scales.

[PDF File (Adobe PDF File), 132KB - jmir_v21i6e12454_app3.pdf]

References

1. Pan Y, Zhang JQ. Born unequal: a study of the helpfulness of user-generated product reviews. J. Retail 2011 Dec;87(4):598-612 [FREE Full text] [doi: 10.1016/j.jretai.2011.05.002]
2. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862 [FREE Full text] [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
3. DeAndrea DC, van der Heide B, Vendemia MA, Vang MH. How people evaluate online reviews. J Commun Res 2015;45(5):719-736 [FREE Full text] [doi: 10.1177/0093650215573862]
4. Grasso KL, Bell RA. Understanding health information seeking: a test of the risk perception attitude framework. J Health Commun 2015;20(12):1406-1414 [FREE Full text] [doi: 10.1080/10810730.2015.1018634] [Medline: 26161622]

5.   Verlegh PW, Ryu G, Tuk MA, Feick L. Receiver responses to rewarded referrals: the motive inferences framework. J Acad Mark Sci 2013 Jan 29;41(6):669-682 [FREE Full text] [doi: 10.1007/s11747-013-0327-8]

6.   Walther JB, Jang JW, Hanna Edwards AA. Evaluating health advice in a web 2.0 environment: the impact of multiple user-generated factors on HIV advice perceptions. Health Commun 2018;33(1):57-67 [FREE Full text] [doi: 10.1080/10410236.2016.1242036] [Medline: 27911096]

7.   López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]

8.   Liang B, Scammon DL. Incidence of online health information search: a useful proxy for public health risk perception. J Med Internet Res 2013 Jun 17;15(6):e114 [FREE Full text] [doi: 10.2196/jmir.2401] [Medline: 23773974]

9.   Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. J Am Med Assoc 2014 Feb 19;311(7):734-735 [FREE Full text] [doi: 10.1001/jama.2013.283194] [Medline: 24549555]

10. Li S, Feng B, Chen M, Bell RA. Physician review websites: effects of the proportion and position of negative reviews on readers' willingness to choose the doctor. J Health Commun 2015 Apr;20(4):453-461 [FREE Full text] [doi: 10.1080/10810730.2014.977467] [Medline: 25749406]

11. McLennan S, Strech D, Reimann S. Developments in the frequency of ratings and evaluation tendencies: a review of German physician rating websites. J Med Internet Res 2017;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]

12. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]

13. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]

14. Rau J. Kaiser Health News. 2013. Nearly 1,500 Hospitals Penalized Under Medicare Program Rating Quality URL: https://khn.org/news/value-based-purchasing-medicare/ [WebCite Cache ID 6yS5LxbFr]

15. Medical Group Management Association. 2014. MGMA Survey: Practices Report Gradual Shifts Toward Provider Pay Tied to Quality and Patient Satisfaction Metrics URL: https://www.healthitoutcomes.com/doc/mgma-survey-shifts-quality-patient-satisfaction-metrics-0001 [WebCite Cache ID 6yRpRfx4r]

16. Berry LL, Bendapudi N. Health care: a fertile field for service research. J Serv Res 2007;10(2):111-122 [FREE Full text] [doi: 10.1177/1094670507306682]

17. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011 Dec 6;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]

18. Adams JG, Biros MH. The elusive nature of quality. Acad Emerg Med 2002 Nov;9(11):1067-1070 [FREE Full text] [doi: 10.1197/aemj.9.11.1067] [Medline: 12414452]

19. Healthgrades. 2014. Doctors, hospitals and health plans: how consumers choose URL: http://www.multivu.com/assets/63942/documents/63942-healthgrades-infographic-original.pdf [WebCite Cache ID 6yRrR9ZBb]

20. Büyüközkan G, Çifçi G, Güleryüz S. Strategic analysis of healthcare service quality using fuzzy AHP methodology. Expert Syst Appl 2011 Aug;38(8):9407-9424 [FREE Full text] [doi: 10.1197/aemj.9.11.1067]

21. McColl-Kennedy JR, Vargo SL, Dagger TS, Sweeney JC, van Kasteren Y. Health care customer value cocreation practice styles. J Serv Res 2012 May;15(4):370-389 [FREE Full text] [doi: 10.1177/1094670512442806]

22. Gray B, Vandergrift J, Lipner RS, Gao G, McCullough J. Evaluating Website Ratings: How Well do Patient Ratings of Physicians Predict Their quality of Care? In: Annual Research Meeting. 2014 Presented at: AcademyHealth; June 2014; Washington, DC.

23. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]

24. Aiken KD, Boush DM. Trustmarks, objective-source ratings, and implied investments in advertising: investigating online trust and the context-specific nature of internet signals. J Acad Mark Sci 2006 Jul 1;34(3):308-323 [FREE Full text] [doi: 10.1177/0092070304271004]

25. Chevalier J, Mayzlin D. The effect of word of mouth on sales: online book reviews. J Market Res 2018 Oct 10;43(3):345-354 [FREE Full text] [doi: 10.1509/jmkr.43.3.345]

26. Lagu T, Hannon N, Rothberg M, Lindenauer P. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]

27. Zimmermann M, Jucks R. How experts' use of medical technical jargon in different types of online health forums affects perceived information credibility: randomized experiment with laypersons. J Med Internet Res 2018 Dec 23;20(1):e30 [FREE Full text] [doi: 10.2196/jmir.8346] [Medline: 29362212]

28. Flanagin AJ, Metzger MJ. Trusting expert- versus user-generated ratings online: the role of information volume, valence, and consumer characteristics. Comput Human Behav 2013;29(4):1626-1634 [FREE Full text] [doi: 10.1016/j.chb.2013.02.001]

29. Khare A, Labrecque LI, Asare AK. The assimilative and contrastive effects of word-of-mouth volume: an experimental examination of online consumer ratings. J. Retail 2011 Mar;87(1):111-126 [FREE Full text] [doi: 10.1016/j.jretai.2011.01.005]

30. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]

31. Gershoff AD, Broniarczyk SM, West PM. Recommendation or evaluation? task sensitivity in information source selection. J Consum Res 2001 Dec;28(3):418-438 [FREE Full text] [doi: 10.1086/323730]

32. Gilly MC, Graham JL, Wolfinbarger MF, Yale LJ. A dyadic study of interpersonal information search. J Acad Mark Sci 1998;26(2):83-100 [FREE Full text] [doi: 10.1177/0092070398262001]

33. Ketelaar PE, Willemsen LM, Sleven L, Kerkhof P. The good, the bad, and the expert: how consumer expertise affects review valence effects on purchase intentions in online product reviews. J Comput Mediat Commun 2015 Nov 27;20(6):649-666 [FREE Full text] [doi: 10.1111/jcc4.12139]

34. Brock TC. Communicator-recipient similarity and decision change. J Pers Soc Psychol 1965;1(6):650-654 [FREE Full text] [doi: 10.1037/h0022081]

35. Simons HW, Berkowitz NN, Moyer RJ. Similarity, credibility, and attitude change: a review and a theory. Psychol Bull 1970;73(1):1-16 [FREE Full text] [doi: 10.1037/h0028429]

36. Ruef M, Aldrich HE, Carter NM. The structure of founding teams: homophily, strong ties, and isolation among U.S. entrepreneurs. Am Sociol Rev 2003 Apr;68(2):195-222 [FREE Full text] [doi: 10.2307/1519766]

37. Price LL, Feick LF. The role of interpersonal sources in external search: an informational perspective. Adv Consum Res 1984;11:250-255 [FREE Full text]

38. Wang Z, Walther JB, Pingree S, Hawkins RP. Health information, credibility, homophily, and influence via the internet: web sites versus discussion groups. Health Commun 2008 Jul;23(4):358-368 [FREE Full text] [doi: 10.1080/10410230802229738] [Medline: 18702000]

39. Goldstein NJ, Cialdini RB, Griskevicius V. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. J Consum Res 2008;35(3):472-482 [FREE Full text] [doi: 10.1086/586910]

40. Alba JW, Hutchinson JW. Dimensions of consumer expertise. J Consum Res 1987 Mar;13(4):411-454 [FREE Full text] [doi: 10.1086/209080]

41. Gotlieb JB, Sarel D. Comparative advertising effectiveness: the role of involvement and source credibility. J Advert 1991;20(1):38-45 [FREE Full text] [doi: 10.1080/00913367.1991.10673205]

42. Brown J, Broderick AJ, Lee N. Word of mouth communication within online communities: conceptualizing the online social network. J Interact Mark 2007;21(3):2-20 [FREE Full text] [doi: 10.1002/dir.20082]

43. Eastin MS. Credibility assessments of online health information: the effects of source expertise and knowledge of content. J Comput Mediat Commun 2001;6(4):- [FREE Full text] [doi: 10.1111/j.1083-6101.2001.tb00126.x]

44. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]

45. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Communic Res 2010 Nov 25;37(1):105-132 [FREE Full text] [doi: 10.1177/0093650209351512]

46. Mitchell AA, Dacin PA. The assessment of alternative measures of consumer expertise. J Consum Res 1996;23(3):219-239. [doi: 10.1086/209479]

47. Bansal HS, Voyer PA. Word-of-mouth processes within a services purchase decision context. J Serv Res 2000;3(2):166-177 [FREE Full text] [doi: 10.1177/109467050032005]

48. Schweiger S, Oeberst A, Cress U. Confirmation bias in web-based search: a randomized online study on the effects of expert information and social tags on information search and evaluation. J Med Internet Res 2014 Mar 26;16(3):e94 [FREE Full text] [doi: 10.2196/jmir.3044] [Medline: 24670677]

49. Feng B, MacGeorge EL. The influences of message and source factors on advice outcomes. Commun Res 2010 Jun 09;37(4):553-575 [FREE Full text] [doi: 10.1177/0093650210368258]

50. Zhang Z, Ye Q, Law R, Li Y. The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. Int J Hosp Manag 2010;29(4):694-700 [FREE Full text] [doi: 10.1016/j.ijhm.2010.02.002]

51. Feldman SP, Spencer MC. The effect of personal influence in the selection of consumer services. In: Holloway RJ, Mittelstaedt RA, Venkatesan M, editors. Consumer Behavior: Contemporary Research in Action. Boston: Houghton Mifflin; 1971:247-257.

52. Woodside AG, Davenport Jr JW. The effect of salesman similarity and expertise on consumer purchasing behavior. J Mark Res 1974 May;11(2):198-202 [FREE Full text] [doi: 10.2307/3150562]

53. Floyd K, Freling R, Alhoqail S, Cho HY, Freling T. How online product reviews affect retail sales: a meta-analysis. J Retail 2014;90(2):217-232 [FREE Full text] [doi: 10.1016/j.jretai.2014.04.004]

54. Lim YS, van der Heide B. Evaluating the wisdom of strangers: the perceived credibility of online consumer reviews on Yelp. J Comput Mediat Comm 2015;20(1):67-82 [FREE Full text] [doi: 10.1111/jcc4.12093]

55. van der Heide B, Lim YS. On the conditional cueing of credibility heuristics. Communic Res 2016;43(5):672-693 [FREE Full text] [doi: 10.1177/0093650214565915]

56. Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annu Rev Psychol 2004;55:591-621 [FREE Full text] [doi: 10.1146/annurev.psych.55.090902.142015] [Medline: 14744228]

57. Tanford S, Penrod S. Social influence model: a formal integration of research on majority and minority influence processes. Psychol Bull 1984;95(2):189-225 [FREE Full text] [doi: 10.1037/0033-2909.95.2.189]

58. López-Pintado D, Watts DJ. Social influence, binary decisions and collective dynamics. Ration Soc 2008 Nov;20(4):399-443 [FREE Full text] [doi: 10.1177/1043463108096787]

59. Watts DJ, Dodds PS. Influentials, networks, and public opinion formation. J Consum Res 2007 Dec;34(4):441-458 [FREE Full text] [doi: 10.1086/518527]

60. Hasher L, Zacks RT. Automatic processing of fundamental information: the case of frequency of occurrence. Am Psychol 1984;39(12):1372-1388 [FREE Full text] [doi: 10.1037/0003-066X.39.12.1372]

61. Baker SM, Petty RE. Majority and minority influence: source-position imbalance as a determinant of message scrutiny. J Pers Soc Psychol 1994;67(1):5-19 [FREE Full text] [doi: 10.1037/0022-3514.67.1.5]

62. Latané B. The psychology of social impact. Am Psychol 1981;36(4):343-356 [FREE Full text] [doi: 10.1037/0003-066X.36.4.343]

63. Dholakia UM, Basuroy S, Soltysinski K. Auction or agent (or both)? A study of moderators of the herding bias in digital auctions. Int J Res Market 2002 Jun;19(2):115-130 [FREE Full text] [doi: 10.1016/S0167-8116(02)00064-2]

64. Liu Y. Word of mouth for movies: its dynamics and impact on box office revenue. J Mark 2006 Jul;70(3):74-89 [FREE Full text] [doi: 10.1509/jmkg.70.3.74]

65. Salganik MJ, Dodds PS, Watts DJ. Experimental study of inequality and unpredictability in an artificial cultural market. Science 2006 Feb 10;311(5762):854-856 [FREE Full text] [doi: 10.1126/science.1121066] [Medline: 16469928]

66. Prendergast C, Stole L. Impetuous youngsters and jaded old-timers: acquiring a reputation for learning. J Polit Econ 1996 Dec;104(6):1105-1134 [FREE Full text] [doi: 10.1086/262055]

67. Ye Q, Law R, Gu B, Chen W. The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Comput Hum Behav 2011;27(2):634-639 [FREE Full text] [doi: 10.1016/j.chb.2010.04.014]

68. Greene WH. Econometric Analysis 5th Edition. Upper Saddle River, NJ: Prentice Hall; 2003.

69. Becker-Olsen KL. And now, a word from our sponsor--a look at the effects of sponsored content and banner advertising. J Advert 2003;32(2):17-32 [FREE Full text] [doi: 10.1080/00913367.2003.10639130]

70. Rodgers S. The effects of sponsor relevance on consumer reactions to internet sponsorships. J Advert 2003;32(4):67-76. [doi: 10.1080/00913367.2003.10639141]

71. Chandran S, Morwitz VG. Effects of participative pricing on consumers' cognitions and actions: a goal theoretic perspective. J Consumer Res 2005 Sep;32(2):249-259 [FREE Full text] [doi: 10.1086/432234]

72. Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990 Oct;19(3):39-52 [FREE Full text] [doi: 10.1080/00913367.1990.10673191]

73. Newell SJ, Goldsmith RE. The development of a scale to measure perceived corporate credibility. J Bus Res 2001 Jun;52(3):235-247 [FREE Full text] [doi: 10.1016/S0148-2963(99)00104-6]

74. Jaccard J, Turrisi R. Interaction Effects in Multiple Regression. 2nd Edition. Thousand Oaks, CA: Sage Publications; 2003.

75. Willemsen L, Neijens P, Bronner F. The ironic effect of source identification on the perceived credibility of online product reviewers. J Comput-Mediat Comm 2012 Oct 10;18(1):16-31 [FREE Full text] [doi: 10.1111/j.1083-6101.2012.01598.x]

76. King RA, Racherla P, Bush VD. What we know and don't know about online word-of-mouth: a review and synthesis of the literature. J Interact Mark 2014;28(3):167-183 [FREE Full text] [doi: 10.1016/j.intmar.2014.02.001]

77. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]

78. Hoerger TJ, Howard LZ. Search behavior and choice of physician in the market for prenatal care. Med Care 1995;33(4):332-349. [Medline: 7731276]

79. Darby MR, Karni E. Free competition and the optimal amount of fraud. J Law Econ 1973;16(1):67-88 [FREE Full text] [doi: 10.1086/466756]