Rating a Sports Medicine Surgeon's "Quality" in the Modern Era: an Analysis of Popular Physician Online Rating Websites

Benedict U. Nwachukwu, MD, MBA · Joshua Adjei, BA · Samir K. Trehan, MD · Brenda Chang, MS, MPH · Kelms Amoo-Achampong, BA · Joseph T. Nguyen, MPH · Samuel A. Taylor, MD · Frank McCormick, MD · Anil S. Ranawat, MD

Received: 8 May 2016/Accepted: 15 July 2016/Published online: 17 August 2016
© Hospital for Special Surgery 2016

Abstract *Background:* Consumer-driven healthcare and an increasing emphasis on quality metrics have encouraged patient engagement in the rating of healthcare. As such, online physician rating websites have become mainstream and may play a potential role in future healthcare policy. *Questions/Purposes:* The purpose of this study was to evaluate online patient ratings for US sports medicine surgeons, determine predictors of positive ratings and analyze for inter-website scoring correlation. *Methods:* The American Orthopedic Society for Sports Medicine (AOSSM) member directory was sampled. Surgeon demographic and rating data were searched on three online physicians rating websites: HealthGrades.com (HG), RateMDs.com (RM) and Vitals.com (V). Written rating comments were categorized as relating to the following: surgeon competence, surgeon affability and process of care. Bivariate linear regression, Pearson correlation and multivariable analyses were used to determine factors associated with positive ratings. *Results:* Two hundred seventy-five sports medicine surgeons were included. Two hundred seventy-one (99%) had ratings on at least one of the three websites. Sports surgeons were rated highly across all three websites (mean >4.0/5); however, there was only a low to moderate degree

of correlation among websites. On HG, female surgeons and surgeons in academia were more likely to receive higher overall ratings. Across all three websites, increased number of years in practice inversely correlated with ratings; this relationship neared significance for HG and was significant for RM. A surgeon's online presence or geographic location was not associated with higher ratings. In multivariable regression analysis for ratings on HG, female sex was the only significant predictor of higher ratings. Two thousand three hundred forty-one written comments were analyzed: perceived surgeon competence and communication influenced the direction of ratings for the top and bottom tier surgeons. *Conclusion:* There was a low degree of correlation among online websites for surgeon ratings. Female surgeons and those with fewer years in practice appear to have higher ratings on these websites; comment content analysis suggests that high and low ratings are influenced by perceived surgeon competence and affability.

Keywords quality · online surgeon ratings · healthcare policy · sports medicine

Introduction

Healthcare has transitioned into a value-based era [12, 14] with resulting attention to patient-focused delivery metrics [13]. Along with social media, online physician rating websites have become a popular medium for patients to express their level of satisfaction and have been shown to influence physician selection. As evidence, cross-sectional studies from the USA and Germany demonstrate that 28% of patients have visited physician rating websites and 11% had posted ratings or written comments [5, 7]. Furthermore, Hanauer et al. noted in their survey of 3563 US adults that physician review websites were considered very important (19%), somewhat important (40%) and not important (41%) when selecting a physician [8]. Although it is debatable

Electronic supplementary material The online version of this article (doi:10.1007/s11420-016-9520-x) contains supplementary material, which is available to authorized users.

B. U. Nwachukwu, MD, MBA (✉) · J. Adjei, BA ·
S. K. Trehan, MD · B. Chang, MS, MPH · J. T. Nguyen, MPH ·
S. A. Taylor, MD · A. S. Ranawat, MD
Hospital for Special Surgery,
535 East 70th Street,
New York, NY 10021, USA
e-mail: nwachukwu@hss.edu

K. Amoo-Achampong, BA · F. McCormick, MD
Sports Medicine Department, SOAR Institute,
Miami, FL 32901, USA

whether online ratings websites in their current form truly capture patient satisfaction and objectively evaluate the delivery of care, they represent a potential tool for both payors and healthcare systems to gauge how surgeons are assessed by their patients. With such potential for online physician rating websites to play a role in guiding healthcare policy, a critical analysis and interpretation of these websites' output is warranted.

From the clinical surgeon's perspective, the demonstrable staying power and high rate of patient adoption of these websites should provide an exhortation to understand, at least on a general level, the drivers of positive and negative ratings. Improved appreciation for these factors may be useful for the surgeon in better guiding patient encounters, overseeing his/her practice and potentially managing online reputation. Historically, three key qualities—affability, availability and ability (“3 As”)—have been suggested to promote a successful surgical career and patient interaction [2, 10]. It is unclear however whether these principles are still valid in the modern Internet era.

While content analysis of online physician rating has been performed across multiple medical and surgical specialties [3, 7, 16], the few studies [1, 6, 19] that looked at orthopaedic surgery were limited in their scope and analysis. Sports medicine is a high volume, elective field within orthopaedics and may provide the opportunity to analyze and validate the content of online physician rating websites. The purpose of this study was to evaluate online patient ratings for US sports medicine surgeons, determine predictors of positive ratings and determine inter-website scoring correlation.

Materials and Methods

The American Orthopaedic Society for Sports Medicine (AOSSM) online member directory was queried to identify the surgeons included as part of this study. The directory was accessed on May 20, 2015, yielding 2813 active sports medicine surgeon entries. “Active” membership designation within the AOSSM organization is differentiated from other tiers of membership in that “Active” membership is associated with greater clinical requirement. Every tenth surgeon profile was selected on a continuous basis for a total 10% database sampling. Seven of the identified patients were excluded as they did not have an MD degree and were assumed to be non-clinical; the final surgeon in the registry was added for a summation of 275 included surgeons. Surgeon name, geographic location and practice type (private vs academic) were recorded for each included surgeon. For surgeons with no practice type recorded on the AOSSM directory, their name was secondarily searched on Google to determine current institution of practice. Academic institutions were considered those that had a University affiliation and/or had an associated orthopaedic residency program. Each surgeon's state was categorized into geographic regions that were in accordance with the Bureau of Economic Analysis' definitions (i.e. New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain and Far West).

Physician ratings on online ratings websites are typically derived from overall patient ratings for the surgeon and depending on the website are also derived from various surgeon and care delivery attributes. Profiles for individual surgeons were then reviewed on three of the most commonly visited physician review websites: HealthGrades.com (HG), Vitals.com (V) and RateMDs.com (RM). HG and V were selected because they are the two most visited physician review websites and RM was selected due to the availability of written patient comments [1, 9, 16]. Data collected from these three various online websites were utilized to develop a dataset of online rating metrics available for each included surgeon. Data collected from HG included the following: physician age, “Care Philosophy” (i.e. personal profile established by the physician on the website), number of ratings and overall rating (scored out of 5). Data collected from V included years in practice, “Personal Statement” (i.e. personal profile established by the physician on the website), Castle Connolly status (a designation as a regional Top Doctor based on the Castle Connolly peer nomination and review process—<https://www.castleconnolly.com/about/nomprocess.cfm>), number of ratings, overall rating (scored out of 5) and number of written comments. Data collected from RM included number of ratings, overall rating (scored out of 5) and number of written comments. Written comments on V and RM were individually reviewed and categorized into one of three categories: (1) comments related to surgeon professional competence, ability and/or clinical outcome (“surgeon competence”); (2) comments related to surgeon communication, affability or availability (“surgeon affability”) and (3) comments related to process of care and with minimal relation to the surgeon (“process of care”), i.e. cost considerations, office staff and institutional/facility considerations. Classification of written comments on physician review websites has been previously described by Emmert et al. [4].

Secondary searches for each surgeon were conducted beyond the search on HG, V and RM. Surgeons were searched on Google (“first name” + “last name” + “, MD”) and the number of physician rating websites within the top ten search results was recorded as a proxy for the visibility of these websites when searching for surgeons. It was also determined whether each surgeon had a professional website, Facebook page and/or Twitter profile. Surgeon “online presence” was derived based on the following five criteria—personal professional website, Facebook page, Twitter page and personal profiles on HG and/or V [19].

All online data was collected between May 24, 2015 and September 10, 2015. Bivariate correlations using a simple linear regression analysis was used to generate Spearman's rank correlation coefficients (*R*) to determine the correlation among online rating websites based on ratings across the three websites for each surgeon. Spearman correlation coefficients were also used to determine the relationship between the number of total ratings and the mean overall rating for surgeons on each respective website. To interpret Pearson correlation coefficients, we followed a published guideline suggesting that 0.0–0.3 is a minor correlation, 0.3–0.5 a low correlation, 0.5–0.7 a

moderate correlation, 0.7–0.9 a high correlation and 0.9–1.0 a very high correlation [11]. Bivariate analysis was also used to compare continuous factors (e.g. year in practice) against mean overall surgeon rating for each website. Independent samples *t* tests were used to compare mean overall ratings between different discrete variables potentially associated with online ratings (e.g. Castle Connolly status, personal website, online social media profiles, etc.). A multivariable regression analysis was performed to see which factors were predictors for the online ratings. Additionally, after the written comments had been categorized as relating to surgeon competence, surgeon affability or process of care—a regression analysis was performed to evaluate whether the comment types left for the surgeon on RM influenced their overall rating on RM as reported by the patient on a 1–5 scale. Specifically it was assessed whether a certain tier of surgeon was more likely to be associated with certain categories of comments. Comments were categorized as positive or negative by one of the authors (JA) based on overall assessment of the tone of the comment. Surgeons were tiered into top tier (mean rating score, 4.0–5.0), middle tier (mean rating score, 2.01–3.99) and bottom tier (0.0–2.0). Statistical significance was set at a *p* value less than 0.05 for all analyses. Analyses were performed using SAS Software version 9.3 (SAS Institute).

Source of Funding

This study did not receive any external funding.

Results

Surgeon/Rating Website Characteristics

Two hundred and seventy-five sports medicine surgeons were included as part of this study. Of these, 262 (95.3%) were male and 76 (27.6%) were in an academic practice setting. Forty-six (16.7%) surgeons had their own personal website, 42 (15.3%) had a Facebook page and 32 (11.6%) had a Twitter account (Table 1). On average, 3 of the first 10 (mean, 3.07) hits when a surgeon's name is input to Google are related to an online rating website.

Among the included surgeons, 271 (98.5%) had ratings on at least one of the three websites—V (271, 98.2%), HG (270, 98.2%) and RM (173, 62.9%). The mean number of ratings available for each surgeon was 18.9, 16.7 and 4.5 on HG, V and RM, respectively. The mean score for each surgeon was 4.16 (SD $\pm$ 0.68), 4.05 (SD $\pm$ 0.75) and 4.01 (SD $\pm$ 1.0) on HG, V and RM, respectively. There was no significant correlation between the number of ratings and likelihood for higher ratings—HG ($R = -0.06$, $p = 0.32$), V ($R = 0.07$, $p = 0.25$) and RM ($R = -0.07$, $p = 0.31$). Scores on HG demonstrated a statistically significant low correlation with scores on V ($R = 0.32$, $p < 0.0001$) and scores on RM ($R = 0.308$, $p < 0.0001$) while V and RM scores demonstrated a statistically significant moderate correlation ($R = 0.51$, $p < 0.0001$) (Fig. 1)

Table 1 General demographics of included surgeons

Total number of surgeons	<i>N</i> = 275 (%)
Online rating site	
Healthgrades	270 (98.2)
RateMDs	173 (62.9)
Vitals	271 (98.5)
Sex	
Male	262 (95.3)
Female	13 (4.7)
Practice setting	
Academic	76 (27.6)
Private	199 (72.4)
Geography	
Far West	31 (11.3)
Great Lakes	48 (17.5)
Midwest	56 (20.4)
New England	22 (8.0)
Plains	19 (6.9)
Rocky Mountain	9 (3.3)
Southeast	54 (19.6)
Southwest	26 (9.5)
Missing	10 (3.6)
Social media	
Facebook	42 (15.3)
Twitter	32 (11.6)
Personal website	46 (16.7)

Factors Associated with Higher Ratings

Female sports medicine surgeons demonstrated significantly higher ratings on HG compared to their male counterparts (female, 4.54 ± 0.54 ; male: 4.14 ± 0.68 , $p = 0.04$); no significant difference was observed on V ($p = 0.42$) and RM ($p = 0.70$) based on sex however. Increased years in practice was associated with a statistically significant minor negative correlation with increased scores on HG ($R = -0.11$, $p = 0.072$) and RM ($R = -0.18$, $p = 0.021$); no significant correlation was demonstrated on V ($R = -0.08$, $p = 0.18$). Surgeons with an academic affiliation had a significantly higher rating score on HG than non-academic surgeons (academic, 4.30 ± 0.64 ; non-academic, 4.11 ± 0.69 , $p = 0.03$); no significant difference was found on V and RM regarding surgeon affiliation. No significant association was found for surgeon rating and geography across any of the websites.

HG and V allow for physicians to personalize their ratings pages through Care Philosophy (HG) and a Personal Statement (V). On HG, surgeons who had a Care Philosophy ($N = 63$; mean 24.6 reviews) had a significantly higher number of reviews compared to those surgeons who did not ($N = 210$; mean 17.14 reviews) ($p = 0.003$). Having a Care Philosophy however did not influence how highly surgeons were rated; mean, 4.15 (no Care Philosophy) vs 4.17 (Care Philosophy); $p = 0.805$. On V, having a personal statement neither influenced the quantity ($p = 0.872$) or positivity ($p = 0.805$) of ratings. Additionally surgeons attaining Castle Connolly Status did not have a significantly higher mean overall rating across all three websites.

Surgeon presence on social media including Facebook, Twitter and possession of a personalized website did not influence the strength of ratings on any of the three websites.

Fig. 1. Box plots comparing sports surgeon scores reported on HG vs RM and V vs RM. A higher degree of correlation is shown between V and RM. However in both comparisons, surgeons with 4.0–5.0 ratings on one website are likely to receive a similar rating on the comparator website. There is a higher score distribution in mean scores ranging between 2.5 and 3.5.

Online presence was similarly not significantly correlated with overall rating on any of the websites.

Multivariable analysis was possible only for HG. In multivariable analysis on HG, female sex was the only significant factor associated with higher ratings ($p = 0.04$). The number of years in practice ($p = 0.10$) and Castle Connolly Status ($p = 0.13$) approached but did not reach significance.

Two thousand three hundred forty-one comments were reviewed and categorized as part of a detailed comment content analysis (V, 1532 comments; RM, 809 comments). Certain types of comments were more significantly associated with certain tiers of surgeon rating. On V, bottom tier surgeons

were significantly associated with negative comments in regard to surgeon competence and surgeon affability while top tier surgeons were significantly associated with positive comments relating to surgeon affability ($p < 0.0001$ for all). Middle tier surgeons however were not significantly associated with any comment type. Similarly on RM, bottom tier surgeons were significantly associated with negative comments regarding surgeon competence and surgeon affability ($p < 0.0001$ for both); top tier surgeons were significantly associated with positive comments regarding surgeon competence ($p = 0.02$) and middle tier surgeons were significantly associated with process of care comments ($p = 0.04$).

Discussion

Physician online rating websites are increasingly utilized and have the potential to play a role in healthcare policy and the assessment of physician quality of care. Sports medicine is a high volume, elective field within orthopaedics and may provide the opportunity to analyse and validate the content of online physician rating websites. In this study, we found that there was only a low to moderate degree of correlation among online rating websites for surgeon ratings. Only one of the websites (HG) afforded the opportunity for multivariable analysis and it was noted that female sex was a significant predictor of higher rating, while fewer years in practice and Castle Connolly status were near significant predictors of higher ratings. The top and bottom tier surgeons (based on tier surgeon ratings) were significantly more likely to receive comments about their competence or affability; however, middle tier surgeon did not demonstrate any association with surgeon specific comments.

This study has certain limitations. The surgeons included in this study were part of a limited dataset from the AOSSM member registry and as such represents a limited sampling of sports medicine surgeons. However, the entire database was sampled using a consecutive number sequence therefore providing a fairly balanced sample. Additionally, this study is limited by the quality of the data itself—online surgeon data may not represent an accurate appraisal of surgeon quality and may not even be provided by real patients [6]. However, this study is a narrative description and analysis of this publically available data and it is beyond the scope of this analysis to specifically comment on the validity or veracity of the content.

Online physician ratings have been previously studied to a limited extent within orthopaedic surgery. Bakhsh and Mesfin [1] evaluated the online ratings for orthopaedic surgeons across various online rating websites; the authors identified process of care measures relating to positive online ratings but their study was restricted to surgeons in St. Louis, MO, and thus limited in generalizability. Frost and Mesfin [6] extended this prior analysis to assess representation of orthopaedic surgeons across 30 US states on online rating websites. The authors found, based on analysis with a limited set of uncontrolled variables, that academic surgeons and those in practice between 6 and 10 years had higher ratings. In the present study, academic affiliation was similarly associated with

higher ratings; however, this association lost significance in the multivariable model. Surgeon experience was evaluated based on the number of years in practice and we used regression analysis with years in practice as a continuous variable as opposed to categorical grouping of number of years in practice; even with this methodology, we similarly found a bias toward surgeons with fewer years in practice. Most recently, Trehan et al. [19] investigated the predictors of higher online ratings among hand surgeons. The authors found that positive comments were more related to perceived surgeon ability while negative comments were independent of perceived surgeon ability. The current analysis suggests however that both surgeon affability and competence traits drive the highest and the lowest ratings without a significant impact on mediocre ratings. Additionally, Trehan et al. found that online presence and Castle Connolly Status were correlated with higher ratings for hand surgeons. Online presence was not a significant predictor among sports medicine surgeons; however, interestingly Castle Connolly status was.

A low degree of correlation for individual surgeon ratings across different online ratings websites is an important finding that has not been previously demonstrated. Online rating websites are for-profit business enterprises, which at this point demonstrate significant growth potential. However, the low degree of correlation between online ratings websites is concerning and questions the collective utility of these websites and also potentially demonstrates the individually capricious nature of physician online reviews. Low inter-website correlation may be explained by an insufficiency of reviews by which to appropriately grade a surgeon. Prior authors have expressed concern with regard to an overall relative paucity of reviews constituting a physician's online "grade" [15, 17]. Thus, the lack of correlation among websites may indeed be related to a rapid proliferation of these websites and a subsequent dilution of target patients across competing websites. The critical mass number of reviewers needed per site to improve reliability is unclear and may warrant further investigation. Improved correlation and reliability could potentially be achieved through either consolidation of websites or an institution of regulatory oversight to potentially amalgamate reviews and improve reliability. The former scenario however would require a subversion of business and profit incentives while the latter would likely require legislative and physician-driven advocacy.

Additionally, the implicit goal of online physician rating websites is to capture the quality of care delivered by the physician under rating. Assessment of physician quality using the current methods employed in the websites included in this review is likely flawed and subjective in nature. Orthopaedic interventions aim to restore patients to pre-injury function and/or relieve pain. As such surgeon quality may be best assessed by reporting functional measures and/or patient-reported outcome data on these websites as well. As part of this system for example, a knee surgeon performing anterior cruciate ligament reconstruction would have available for public review mean Tegner, IKDC or other knee functional outcome score for patients providing grades. There is an opportunity to refine methods for online

physician grading to become more focused on actual outcome and thus quality of care.

It was demonstrated in this study that written comments on both top and bottom tier surgeons were significantly more likely to be related to perceived surgeon competence or affability. This finding was all the more striking because middle tier surgeons only had an association with process of care comments and conversely process of care comments were not significantly associated with surgeons at either end of the rating scale. This finding suggests issues related to surgeon ability and affability can have polarizing effects for the patient prompting the patient to either review the surgeon very highly or very lowly. As such, it appears that even in the modern era and with the adoption of online rating mechanisms, the traditional 3 As of availability, affability and ability hold sway. In light of this finding, it is also possible that issues relating to affinity and personal factors may contributed to the finding that women and surgeons with fewer years in practice have higher ratings. Although we did not examine the demographics of patients providing reviews, prior evidence suggests that the demographic profile of a patient more likely to post a review is that of a young, female and well-educated individual [5, 18]. This subset of patients may feel greater affinity for younger surgeons or those of the same sex and thus may provide higher reviews to these individuals. It is also worth noting that different patient demographics may have different expectations for the outcome of their care and this may impact the overall surgeon rating. Ratings websites may benefit from providing a demographic breakdown of patients completing reviews.

In conclusion, sports medicine surgeons are highly represented on online physician rating websites. There was a low degree of correlation among the three reviewed websites for surgeon ratings. Female sports medicine surgeons and those with fewer years in practice appear to have higher ratings on these websites; comment content analysis suggests that high and low ratings are influenced by perceived surgeon competence and affability. An understanding of the factors relating to online physician rating may have important implications for the sports medicine surgeon. More attention should be paid to improving the validity of online ratings for assessing quality and the outcome of care delivered.

Compliance with Ethical Standards

Conflict of Interest: Benedict U. Nwachukwu, MD, MBA, Joshua Adjei, BA, Samir K. Trehan, MD, Brenda Chang, MS, MPH, Kelms Amoo-Achampong BA, Joseph T. Nguyen, MPH, Samuel A. Taylor, MD and Frank McCormick, MD, have declared that they have no conflict of interest. Anil S. Ranawat, MD reports personal fees from Stryker-MAKO Surgical Corp, outside the work; receives royalties for a patent with Elsevier Inc and Editor in Chief at Springer.

Human/Animal Rights: This article does not contain any studies with human or animal subjects performed by the any of the authors.

Informed Consent: N/A

Required Author Forms Disclosure forms provided by the authors are available with the online version of this article.

References

1. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: analysis of 2185 reviews. *American journal of orthopedics (Belle Mead, NJ)*. 2014; 43: 359-363.
2. Chow A, Mayer EK, Darzi AW, et al. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery*. 2009; 146: 435-443.
3. Ellimoottil C, Hart A, Greco K, et al. Online reviews of 500 urologists. *The Journal of urology*. 2013; 189: 2269-2273.
4. Emmert M, Meier F, Heider AK, et al. What do patients say about their physicians? An analysis of 3000 narrative comments posted on a German physician rating website. *Health policy (Amsterdam, Netherlands)*. 2014; 118: 66-73.
5. Emmert M, Meier F, Pisch F, et al. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. *Journal of medical Internet research*. 2013; 15, e187.
6. Frost C, Mesfin A. Online reviews of orthopedic surgeons: an emerging trend. *Orthopedics*. 2015; 38: e257-262.
7. Hanauer DA, Zheng K, Singer DC, et al. Parental awareness and use of online physician rating sites. *Pediatrics*. 2014; 134: e966-975.
8. Hanauer DA, Zheng K, Singer DC, et al. Public awareness, perception, and use of online physician rating sites. *Jama*. 2014; 311: 734-735.
9. Kadry B, Chu LF, Kadry B, et al. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of medical Internet research*. 2011; 13, e95.
10. Lindsey RW. The role of the department chair in promoting diversity. *The Journal of the American Academy of Orthopaedic Surgeons*. 2007; 15(Suppl 1): S65-69.
11. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi*. 2012; 24: 69-71.
12. Nwachukwu BU, Bozic KJ, Schairer WW, et al. Current status of cost utility analyses in total joint arthroplasty: a systematic review. *Clin Orthop Relat Res*. 2015; 473: 1815-1827.
13. Nwachukwu BU, Hamid KS, Bozic KJ. Measuring Value in Orthopaedic Surgery. *JBJS Reviews*. 2013;1.
14. Nwachukwu BU, Schairer WW, Bernstein JL, et al. Cost-effectiveness analyses in orthopaedic sports medicine: a systematic review. *The American journal of sports medicine*. 2015; 43: 1530-1537.
15. Sabin JE. Physician-rating websites. *The virtual mentor : VM*. 2013; 15: 932-936.
16. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? *JAMA otolaryngology– head & neck surgery*. 2014; 140: 635-638.
17. Tanne JH. Doctor-rating websites base their reports on only a few patient reviews. *BMJ (Clinical research ed)*. 2013; 346: f295.
18. Terlutter R, Bidmon S, Rottl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. *Journal of medical Internet research*. 2014; 16, e97.
19. Trehan SK, DeFrancesco CJ, Nguyen JT, et al. Online Patient Ratings of Hand Surgeons. *The Journal of hand surgery*. 2016; 41: 98-103.