

Original research

Characterizing extremely negative reviews of total joint arthroplasty practices and surgeons on yelp.com

Jaymeson R. Arthur, MD ^a, David Etzioni, MD ^b, Adam J. Schwartz, MD, MBA ^{a,*}^a Department of Orthopaedic Surgery, Mayo Clinic Arizona, Phoenix, AZ, USA^b Division of Colon and Rectal Surgery, Mayo Clinic Arizona, Phoenix, AZ, USA

ARTICLE INFO

Article history:

Received 9 January 2019

Received in revised form

19 February 2019

Accepted 23 February 2019

Available online 10 April 2019

Keywords:

Yelp

Online review

Total joint replacement

Total hip replacement

Total knee replacement

ABSTRACT

Background: Although physicians tend to prefer data-driven quality metrics, emerging evidence suggests that patients prefer crowd-sourced information containing patient narrative descriptions of the care experience. Currently, [yelp.com](http://www.yelp.com) is the most commonly accessed Web resource among patients who use online information to choose a surgeon. The purpose of this study is to characterize extremely negative reviews of total joint arthroplasty surgeons and practices on [yelp.com](http://www.yelp.com).

Methods: We searched [yelp.com](http://www.yelp.com) for one-star (out of 5) reviews of total joint providers and practices in 8 major US metropolitan areas. These reviews were then classified into categories based on content: clinical, nonclinical, or both. Reviews were further subcategorized as “surgical” and “nonsurgical” representing reviews of a nonsurgical experience (eg, initial office visit).

Results: A higher proportion of reviews came from patients who did not report prior surgery by the surgeon or practice named in the review than from those who reported surgery (240 reviews, 75.0%, 95% confidence interval: 70.0%–79.4% vs 80 reviews, 25.0%, 95% confidence interval: 20.6%–30.0%, $P < .0001$). Compared with surgical reviews, nonsurgical reviews were more likely to contain nonclinical complaints (92.1% vs 53.8%, $P < .0001$) and less likely to contain clinical complaints (21.3% vs 78.7%, $P < .0001$).

Conclusions: The vast majority of extremely negative reviews of total joint arthroplasty surgeons and practices were related to nonclinical concerns posted by patients who did not report prior surgery by the surgeon or practice being reviewed. The results of this study may help explain the wide disparity commonly observed between conventional quality metrics and crowd-sourced online reviews.

© 2019 The Authors. Published by Elsevier Inc. on behalf of The American Association of Hip and Knee Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Total joint arthroplasty (TJA) is a successful treatment for end-stage arthritis of the hip and knee [1,2]. With this success has come a high demand for these high-cost, yet high-value, procedures, resulting in a massive economic burden to the health-care system [3]. Given the extraordinary costs associated with TJA, it is not surprising that recent years have witnessed heightened interest in maximizing the value of the episode of care through public

disclosure of cost and outcome data. While provider quality data have been publically available for decades, with passage of the Affordable Care Act in 2010 and the subsequent mandate of an open-data policy in all federal departments, the disclosure of previously confidential Medicare utilization and payment data now aims to provide more transparent comparison of providers, with a focus on patient-centered outcomes.

On April 9, 2014, the Centers for Medicare and Medicaid Services (CMS) released detailed information containing utilization information related to over 10 million records accounting for more than \$77 billion in Medicare payments [4]. This unprecedented step was followed by the release of the Consensus Core of Orthopedic Measures in 2016, a set of 7 quality measures designed to improve value-based payment and purchasing and facilitate patient decision-making [5]. Since this release, online performance reviews and ratings by various government and for-profit websites have become increasingly relied on by payers, institutions, and patients to compare providers.

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.artd.2019.02.009>.

* Corresponding author. Department of Orthopaedic Surgery, Mayo Clinic Arizona, 5777 E Mayo Blvd, Phoenix, AZ 85054, USA. Tel.: +1 480 342 2762.

E-mail address: schwartz.adam@mayo.edu

<https://doi.org/10.1016/j.artd.2019.02.009>

2352-3441/© 2019 The Authors. Published by Elsevier Inc. on behalf of The American Association of Hip and Knee Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

While physicians are trained in the scientific method and tend to prefer disease-specific, data-driven outcome metrics, satisfaction with the overall care experience is now widely recognized as an important component of health-care quality. This recognition is exemplified by the Hospital Consumer Assessment of Healthcare Providers and Systems, a survey instrument and data collection methodology for measuring patients' perceptions of their hospital experience. This survey is currently included as part of the CMS quarterly publication of overall institutional ratings on the Hospital Compare website [6]. In addition to government-sponsored online ratings systems, many other public and private websites are available for patients to obtain information regarding a prospective provider. One study found that [yelp.com](https://www.yelp.com) was the most frequently accessed site among patients who used online comparison data, whereas CMS's Hospital Compare site was ranked eighth [7].

The vast majority of reviews posted to crowd-sourced websites such as Yelp are of the extreme variety, either positive or negative [8]. Previous studies have analyzed the factors associated with higher rankings among orthopedic providers [9–12]; however, we are unaware of any study to examine the factors associated with extremely negative online reviews of TJA surgeons. The purpose of this study was to characterize extremely negative reviews of TJA practices and surgeons.

Material and methods

We performed a search in [yelp.com](https://www.yelp.com) using the keywords “total hip replacement” and “total knee replacement” for 8 major metropolitan areas including New York, Boston, San Francisco, Los

Angeles, Dallas, Phoenix, Washington D.C., and Seattle. These locations were chosen because of the high density of high-volume joint arthroplasty centers of excellence [13]. The practice or group named in the review was confirmed as one that performs TJA via secondary internet search using physician websites, Healthgrades, Doximity, Angie's List, and [vitals.com](https://www.vitals.com). When no physician was named in the review in a group practice that included other subspecialties (such as sports medicine, foot and ankle, trauma, and so forth), these reviews were excluded in an attempt to isolate only confirmed TJA reviews.

Only single-star (out of a possible 5 stars) reviews were included in the analysis (Fig. 1). Reviews were classified as directed toward a specific physician, a specific practice, or both. Reviews were also classified as clinical or nonclinical. Clinical reviews included comments related to treatment complications, readmissions, reoperations, uncontrolled pain, perceived misdiagnosis, delay in communicating treatment plan or care, unmet expectations, and other (eg, “Hip surgery gone very wrong. One year after receiving a hip resurfacing from Dr. X, the peroneal nerve damage that occurred during the surgery has not healed at all, leaving me partly crippled”). Nonclinical reviews included physician bedside manner or professionalism, midlevel (eg, physician assistant, nurse practitioner) bedside manner or professionalism, office staff bedside manner or unprofessionalism, wait time, cost or insurance issues, facilities, scheduling, not enough time spent with the provider, and other (eg, “Unfortunately, I never got to see Dr. X. I arrived early, filled out my paper work for a consultation and waited 55 minutes more ... Finally [sic] told the receptionist, Who [sic] was very pleasant! That [sic] I was no longer going to wait, At [sic] that time

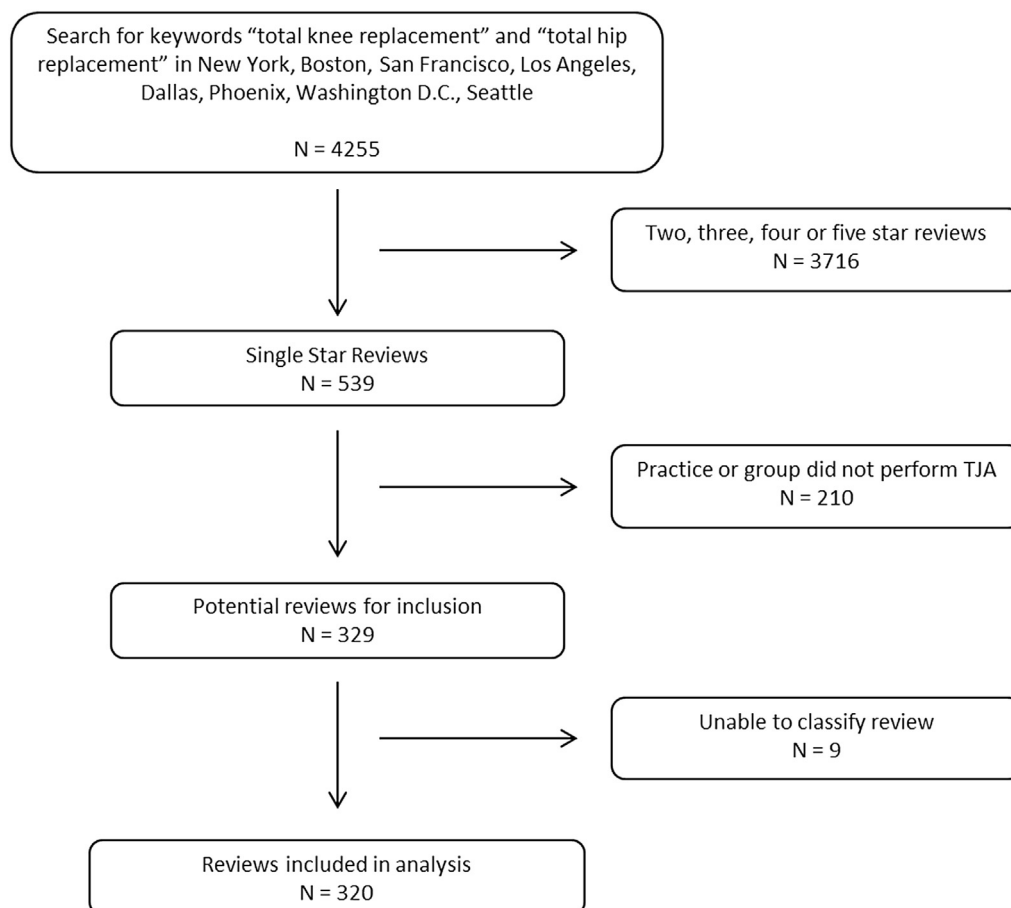


Figure 1. Flow diagram demonstrating inclusion and exclusion criteria for reviews.

she apologized, Said [sic] the Dr. [sic] was running behind—why don't they tell you this when you get there?—and asked If I'd like to reschedule, [sic] I declined and decided to go with another Dr. [sic] for my surgery.") If a surgical episode of care was specifically referenced, the review was classified as surgical (eg, "Currently facing 3 more knee surgeries to repair the original injury and Dr. X's surgery"); if no reference was made to surgical care, the review was labeled as nonsurgical (eg, "Horrendous wait times. Out of my 4 visits I've had to wait no less than 20 minutes over my appointment time. There is always a waiting room packed full of patients. I will be taking my knee issue elsewhere"). Reviews were excluded if an accurate classification was unable to be determined (eg, "Never going back! What a jerk!").

Univariate analysis was performed to determine means and 95% confidence intervals. Continuous variables were compared using the student's t-test, and categorical variables were compared using the chi-square test with alpha set to 0.05. The rate of reviews in each category was determined for surgical and nonsurgical reviews, and the rate ratio (the ratio of the rate for nonsurgical divided by surgical reviews) for each category was determined. Statistical analysis was performed using a commercially available software package (JMP Pro v.13; SAS Institute, Inc.).

Results

A total of 320 single-star reviews were included for study. Comments were directed toward the physician directly in 156 (48.8%), toward the practice directly in 110 (34.4%), and toward both the physician and the practice in 54 (16.9%) (Table 1). A significantly greater number of reviews were related to nonclinical concerns (264 reviews, 82.5%, 95% confidence interval [CI]: 78.0%–86.3%) than to clinically related issues (114 reviews, 35.6%, 95% CI: 30.6%–41.1%, $P < .0001$; numbers add to greater than 320 as 58 reviews contained both clinical and nonclinical concerns).

Table 1
Characteristics of extremely negative reviews of total joint arthroplasty surgeons on yelp.com.

Focus of single star review	Number of responses	Percent of total respondents (N = 320)	Percent of responses within category ^a
Physician	156	48.8%	
Practice	110	34.4%	
Both	54	16.9%	
Clinical	114	35.6%	
Uncontrolled pain	63	19.7%	55.3%
Misdiagnosis	25	7.8%	21.9%
Complication	20	6.3%	17.5%
Reoperation	20	6.3%	17.5%
Unclear treatment plan	17	5.3%	14.9%
Delay in care	14	4.4%	12.3%
Readmission	2	0.6%	1.8%
Other	1	0.3%	0.9%
Nonclinical	264	82.5%	
Physician bedside manner	124	38.8%	47.0%
Wait time	99	30.9%	37.5%
Not enough time spent with provider	78	24.4%	29.5%
Office staff	52	16.3%	19.7%
Cost or insurance	43	13.4%	16.3%
Scheduling issues	41	12.8%	15.5%
Facilities	15	4.7%	5.7%
Midlevel bedside manner	12	3.8%	4.5%
Other	0	0%	0.0%

^a Categories include clinical and nonclinical reviews.

Clinical factors most commonly addressed included uncontrolled pain (63 reviews, 19.7%), perceived misdiagnosis (25 reviews, 7.8%), need for reoperation (20 reviews, 6.3%), complications related to surgery (20 reviews, 6.3%), unmet treatment expectations (17 reviews, 5.3%), and delay in care (14 reviews 4.4%). The most common nonclinical factor cited was physician bedside manner (124 reviews, 38.8%), followed by wait time (99 reviews, 30.9%), inadequate explanation of the medical condition (78 reviews, 24.4%), rude or unprofessional office staff (52 reviews, 16.3%), cost or insurance issues (43 reviews, 13.4%), scheduling issues (41 reviews, 12.8%), poor facilities (15 reviews, 4.7%), and rude or unprofessional midlevel provider (12 reviews, 3.8%).

A significantly higher proportion of reviews came from patients who did not report prior surgery by the surgeon or practice named in the review than from those who reported surgery (240 reviews, 75.0%, 95% CI: 70.0%–79.4% vs 80 reviews, 25.0%, 95% CI: 20.6%–30.0%, $P < .0001$) (Table 2). Compared with surgical reviews, nonsurgical reviews were more likely to contain nonclinical complaints (92.1% vs 53.8%; rate ratio, 1.7; $P < .0001$) and less likely to contain clinical complaints (21.3% vs 78.7%; rate ratio, 0.3; $P < .0001$). Insufficient time spent with the provider (rate ratio, 3.4; $P < .0001$), scheduling issues (rate ratio, 3.1; $P < .05$), wait time (rate ratio, 3.0; $P < .0001$), rude office staff (rate ratio, 2.1; $P < .05$), and physician bedside manner (rate ratio, 1.5; $P < .05$) were more common for nonsurgical reviews than for surgical reviews. Uncontrolled pain was the most frequently identified clinical issue among surgical reviews (62.5%, 95% CI: 51.5%–72.3%).

Discussion

Although TJA is a reliable treatment for end-stage arthritis, rising demand for this procedure coupled with increasing use and dissemination of various quality metrics create new challenges for orthopedic surgeons. To overcome these challenges, it is important to understand how patients evaluate and ultimately choose their surgeon. Crowd-sourced websites such as yelp.com have become increasingly used by patients to guide surgeon selection [14,15]. Although prior studies have examined factors associated with positive reviews on crowd-sourced sites [9–13], the purpose of our study was to characterize the factors leading to extremely negative reviews of TJA practices and surgeons. We found that in this sample of single-star reviews, nonclinical factors such as physician bedside manner, long wait times, and an inadequate explanation of the medical condition were over twice as common than clinical factors such as uncontrolled pain, need for reoperation, and complications related to surgery. Furthermore, only 25% of negative reviews reported any experience with a surgical procedure.

Over the past decade, both government and nongovernment websites have increased reporting of data-driven outcomes. In an effort to shift from volume-based to value-based health-care delivery, CMS now routinely publishes various quality metrics for patients to compare their health-care providers [16]. The availability of such measures, coupled with the lowered barrier to data access, has led to the widespread popularization of online public performance reviews of physicians [17,18]. It is estimated that while currently approximately 30% of patients use online rating systems to choose a surgeon, this number will continue to grow in future years [19,20]. However, it is less clear exactly how this information is processed by patients when choosing a provider. Current evidence suggests that in contrast to conventional, highly numeric data familiar to medical personnel, patients prefer a simple presentation that does not require high cognitive burden [15]. In a recent study analyzing patients' use of common online health-care ratings to select a surgeon, the CMS Physician Compare website was ranked only eighth out of 13 total sites included [7]. The authors

Table 2
Characteristics by surgical vs nonsurgical patients.

Focus of single star review	Surgical patients (n = 80)	%	Nonsurgical patients (n = 240)	%	Rate ratio	P value ^a
Target of comment						
Physician	60	75.0	96	40.0	0.5	<.0001
Practice	8	10.0	102	42.5	4.3	<.0001
Both	12	15.0	42	17.5	1.2	.601
Total	80	100.0	240	100.0	1.0	<.0001
Clinical	63	78.8	51	21.3	0.3	<.0001
Uncontrolled pain	50	62.5	13	5.4	0.1	<.0001
Misdiagnosis	4	5.0	21	8.8	1.8	.258
Complication	20	25.0	0	0.0	0.0	<.0001
Reoperation	20	25.0	0	0.0	0.0	<.0001
Unclear treatment plan	4	5.0	13	5.4	1.1	.885
Delay in care	2	2.5	12	5.0	2.0	.316
Readmission	2	2.5	0	0.0	0.0	<.05
Other	0	0.0	1	0.4	N/A	.448
Nonclinical	43	53.8	221	92.1	1.7	<.0001
Physician bedside manner	23	28.8	101	42.1	1.5	<.05
Wait time	10	12.5	89	37.1	3.0	<.0001
Not enough time spent with provider	7	8.8	71	29.6	3.4	<.0001
Office staff	7	8.8	45	18.8	2.1	<.05
Cost or insurance	10	12.5	33	13.8	1.1	.775
Scheduling issues	4	5.0	37	15.4	3.1	<.05
Facilities	3	3.8	12	5.0	1.3	.639
Midlevel bedside manner	3	3.8	9	3.8	1.0	.999
Other	0	0.0	0	0.0	N/A	N/A

^a P values <.05 were considered statistically significant

point out that one potential explanation was respondents preference for patient comments compared with conventional quality metrics.

Several publicly available websites, including [yelp.com](https://www.yelp.com), now report crowd-sourced outcome information regarding patient-centered satisfaction with the overall care experience. Although some evidence exists that conventional, data-driven quality ratings may correlate with surgical outcomes [21], it is less clear if crowd-sourced websites, driven largely by the patient experience, accurately reflect the quality of medical care. Ramkumar et al. [9] performed a study of 5 of the most popular physician-rating websites with user-generated data and found that none of the sites contained information related to all 7 Consensus Core of Orthopedics Measures. Similarly, Johari et al. [22] compared quality ratings for nursing homes between the CMS-based site called Nursing Home Compare (NHC) and Yelp. The authors found poor correlation between NHC and Yelp ratings, with the latter demonstrating consistently lower staff and quality measures and higher inspection ratings than the NHC rating. In addition, as patients tend to prefer less information presented more concisely in online reviews [15], it is more likely that user-generated data will contain more extreme information than more objective, data-driven sites. Frost and Mesfin [8] provide evidence for this contention as they found in their review of 557 orthopedic surgeons rated on 7 websites that extremely positive or extremely negative reviews accounted for the majority (64.6%) of online comments.

Data-driven and crowd-sourced outcomes are vastly different methods of reporting health-care quality, and the availability of both may paint a more heterogeneous picture of the overall quality of care provided than either metric alone. Previous studies across various specialties have attempted to compare provider quality scores between more conventional data-driven metrics and crowd-sourced information. In the field of dermatology, Smith and Lipoff [23] categorized patient ratings on Yelp and ZocDoc into high-scoring and low-scoring groups. High-scoring reviews were associated with physician temperament, knowledge and competency, physical examination, communication abilities, and mindfulness of cost. Low-scoring reviews were commonly associated with considerations outside of the physician-patient interaction such as

negative interactions with staff, difficulty with scheduling, practice cleanliness, and insurance issues. Similarly, Trehan et al. [11] evaluated the factors associated with positive reviews of hand surgeons on 3 different crowd-sourced websites. They did not include Yelp in their analysis but did find that positive comments were most often related to perceived surgeon competence. Similar to our findings, they also found that negative comments were most often independent of perceived surgeon competence and more often related to communication and office staff or practice characteristics.

This study is not without significant limitations. First, we included only reviews of surgeons and practices in 8 large metropolitan areas that have a high proportion of high-volume TJA centers of excellence. The findings of this study may not be generalizable to smaller rural populations or even other highly populated cities that are not included here. Second, our classification of reviews may be inaccurate. We attempted to mitigate this problem by having 2 separate, independent reviewers classify reviews. One reviewer was a second-year orthopedic surgery resident, and the other was a practicing fellowship-trained adult reconstruction surgeon. In addition, we excluded reviews that were unable to be confidently classified according to the system described. Second, complaints regarding midlevel providers office staff may not have reflected on the arthroplasty portion of the practice as many websites did not specify which supervising physician correlated to a specific midlevel provider. Finally, this is a limited sample of the entire online review universe, and it is possible that the findings seen on [yelp.com](https://www.yelp.com) may not be applicable to other online review sites.

Conclusions

Patient decision-making is a complex process, likely influenced by referral patterns, insurance coverage, out-of-pocket costs, proximity, and many other factors [14,24]. To our knowledge, this is the first study that specifically examines the factors associated with extremely negative reviews of TJA practices and surgeons. We found that the vast majority of extremely negative reviews were related to nonclinical concerns and that 75% of these single-star reviews were posted by patients who did not report prior surgery

by the surgeon or practice being reviewed. We found that nonsurgical patients were over 4 times as likely to complain about the practice, whereas surgical patients were almost twice as likely to complain about the surgeon. In addition, nonsurgical patients were 3 times as likely to complain about wait times, not enough time spent with the provider, and scheduling issues.

Although understanding and focusing on conventional quality metrics remains critically important, patient satisfaction is becoming an increasingly important part of health-care quality measures. TJA practices and surgeons should be cognizant of the factors that are most likely to drive patients to post extremely negative reviews on these increasingly popular websites.

Acknowledgments

Source of Funding: This research was made possible in part by the Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, United States.

References

- [1] Koenig L, Zhang Q, Austin MS, et al. Estimating the societal benefits of THA after accounting for work status and productivity: a Markov Model Approach. *Clin Orthop Relat Res* 2016;474(12):2645.
- [2] Skou ST, Roos EM, Laursen MB, et al. A randomized, controlled trial of total knee replacement. *N Engl J Med* 2015;373(17):1597.
- [3] Centers for Medicare and Medicaid Website. <https://innovation.cms.gov/initiatives/cjr>. [Accessed 10 January 2018].
- [4] Brennan N, Conway PH, Tacenner M, Marilyn RN. The Medicare physician-data release - context and rationale. *N Engl J Med* 2014;317(2):99.
- [5] Centers for Medicare and Medicaid website. <https://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2016-Press-releases-items/2016-02-16.html>; 2016. [Accessed 10 January 2018].
- [6] Centers for Medicare and Medicaid services hospital consumer assessment of healthcare providers and systems website. <https://www.hcahpsonline.org/home.aspx>. [Accessed 10 January 2018].
- [7] Ziemba JB, Allaf ME, Haldeman D. Consumer preferences and online comparison tools used to select a surgeon. *JAMA Surg* 2017;152(4):410.
- [8] Frost C, Mesfin A. Online reviews of orthopedic surgeons: an emerging trend. *Orthopedics* 2015;38(4):e257.
- [9] Ramkumar PN, Navarro SM, Chughtai M, et al. The patient experience: an analysis of orthopedic surgeon quality on physician-rating sites. *J Arthroplasty* 2017;32(9):2905.
- [10] Ramkumar PN, Navarro SM, Haeberle HS, et al. Social media and total joint arthroplasty: an analysis of patient utilization on instagram. *J Arthroplasty* 2017;32(9):2694.
- [11] Trehan SK, DeFrancesco CJ, Nguyen JT, Charalel RA, Daluiski A. Online patient ratings of hand surgeons. *J Hand Surg* 2016;41(1):98.
- [12] Zhang J, Omar A, Mesfin A. Online ratings of spine surgeons: analysis of 208 surgeons. *Spine* 2018;43(12):E722.
- [13] Laucis NC, Chowdhury M, Dasgupta A, Bhattacharyya T. Trend toward high-volume hospitals and the influence on complications in knee and hip arthroplasty. *J Bone Joint Surg Am* 2016;98(9):707.
- [14] Bozic KJ, Kaufman D, Chan VC, Caminiti S, Lewis C. Factors that influence provider selection for elective total joint arthroplasty. *Clin Orthop Relat Res* 2013;471(6):1865.
- [15] Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. *Med Care Res Rev* 2007;64(2):169.
- [16] Krause A, Saleh J, Padela MT, Saleh KJ. The CMS overall hospital quality star rating system. 2018. <https://www.aaos.org/AAOSNow/2018/Jan/Advocacy/advocacy06/?ssopc=1>. Accessed June 21, 2018.
- [17] Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. *JAMA* 2014;311(7):734.
- [18] Kassutto SM, Shah RJ. Patient satisfaction: why and how patients grade you and your pulmonary practice. *Chest* 2015;148(3):833.
- [19] Cohen J. How do patients select a new physician? 77% say online reviews. <https://www.beckershospitalreview.com/healthcare-information-technology/how-do-patients-select-a-new-physician-77-say-online-reviews.html>. Accessed June 21, 2018.
- [20] Sulco PK, McLawhorn AS, Fehring KA, De Martino I. The future of social media in orthopedic surgery. *Curr Rev Musculoskelet Med* 2017;10(2):278.
- [21] Kaye DR, Norton EC, Ellimoottil C, et al. Understanding the relationship between the Centers for Medicare and Medicaid Services' Hospital Compare star rating, surgical case volume, and short-term outcomes after major cancer surgery. *Cancer* 2017;123(21):4259.
- [22] Johari K, Kellogg C, Vazquez K, et al. Ratings game: an analysis of nursing home compare and Yelp ratings. *BMJ Qual Saf* 2018;27:619.
- [23] Smith RJ, Lipoff JB. Evaluation of Dermatology practice online reviews: lessons from qualitative analysis. *JAMA Dermatol* 2016;152(2):153.
- [24] Finlayson SR, Birkmeyer JD, Tosteson AN, Nease Jr RF. Patient preferences for location of care: implications for regionalization. *Med Care* 1999;37(2):204.