Online Patient Ratings Are Not Correlated with Total Knee Replacement Surgeon-Specific Outcomes

Samir K. Trehan, MD · Joseph T. Nguyen, MPH · Robert Marx, MD · Michael B. Cross, MD · Ting J. Pan, MPH · Aaron Daluiski, MD · Stephen Lyman, PhD

Received: 9 August 2017/Accepted: 1 December 2017/Published online: 10 January 2018
© Hospital for Special Surgery 2018

Abstract *Background:* Despite potential concerns regarding their validity, physician-rating websites continue to grow in number and utilization and feature prominently on major search engines, potentially affecting patient decision-making regarding physician selection. *Questions/Purposes:* We sought to determine whether patient ratings on public physician-rating websites correlate with surgeon-specific outcomes for high-volume total knee replacement (TKR) surgeons in New York State (NYS) from 2010 to 2012. *Methods:* Online patient ratings were compared to surgeon-specific outcomes from the Statewide Planning and Research Cooperative System (SPARCS) database from the NYS Department of Health. For each surgeon, we determined the infection rate, re-admission rate, and revision surgery rate within the study period, as well as the mean inpatient length of stay, for TKR from the SPARCS database. Online ratings were collected from two physician-rating websites (Vitals.com and HealthGrades.com). *Results:* One hundred seventy-four high-volume TKR surgeons were identified in NYS from 2010 to 2012. The mean rates of in-hospital infection, 90-day infection, 30-day re-admission, 90-day re-admission, and revision surgery were 0.25, 1.00, 4.89, 8.43, and 1.31%, respectively. The mean number of

ratings for individual surgeons on HealthGrades.com and Vitals.com were 24.0 (range: 0 to 109) and 19.3 (range: 0 to 114), respectively, and mean overall ratings were 4.2 and 4.1 (out of 5) stars, respectively. As with online patient ratings of individual surgeons, variability was observed in the total adverse event rate distribution for individual surgeons. Despite sufficient variability in both online patient rating and surgeon-specific outcomes for high-volume TKR surgeons in NYS, no correlation was observed. *Conclusion:* There was no correlation between surgeon-specific TKR outcome measures and online patient ratings. We therefore advise that patients exert caution when interpreting ratings on these websites.

Keywords online patient ratings · physician-review websites · social media · total knee replacement

Introduction

Increased focus on patient satisfaction, consumer-driver healthcare, and the rise of social media have contributed to a growing industry surrounding quality-of-care metrics, hospital rankings, and online physician ratings. While several studies have attempted to correlate oncologic, orthopedic, and cardiovascular surgical outcomes with publicly available hospital rankings, the potential correlation between online physician ratings and physician-specific outcomes remains unknown [1, 9, 10]. Despite potential concerns regarding the validity of physician-rating websites, they continue to grow in number and utilization and feature prominently on major search engines, potentially affecting patient decision-making regarding physician selection [6, 7]. Twenty-five to 28% of patients visit these websites and 11% actively post ratings and/or comments [3, 6]. Thus, determining whether online patient ratings on publicly available websites correlate with physician-specific outcomes would be of great value to patients and physicians.

Level of Evidence: Level III: Retrospective, Cohort Study

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11420-017-9600-6>) contains supplementary material, which is available to authorized users.

S. K. Trehan, MD · R. Marx, MD · M. B. Cross, MD ·
A. Daluiski, MD
Department of Orthopaedic Surgery, Hospital for Special Surgery,
New York, NY, USA

J. T. Nguyen, MPH · T. J. Pan, MPH · S. Lyman, PhD (✉)
Healthcare Research Institute, Hospital for Special Surgery,
535 East 70th Street,
New York, NY 10021, USA
e-mail: LymanS@hss.edu

The purpose of this study was to determine whether patient ratings on public physician-rating websites correlate with surgeon-specific outcomes. Total knee replacement (TKR) was chosen given the high procedure volume, uniform procedural coding, and relative incidence of relevant outcome measures.

Materials and Methods

The Statewide Planning and Research Cooperative System (SPARCS) database from the New York State (NYS) Department of Health was queried to identify all surgeons who had performed TKR in NYS from 2010 to 2012. The SPARCS database is a census of all non-federal acute care hospital admissions and ambulatory surgical procedures in NYS. TKR procedures were identified by ICD-9-CM (International Classification of Diseases, 9th Revision, Clinical Modification) code 81.54. Only high-volume TKR surgeons, defined as having performed 180 procedures or more during the study period, were included to maximize the likelihood that online ratings reflected patient experience as related to TKR. One hundred seventy-four surgeons were identified using these criteria.

The SPARCS database allows identification of patient demographic and admission information with an individual physician's NYS Physician License Number, as well as unique anonymous patient identifiers to track outcomes after discharge. As a result, surgeon-specific patient outcomes can be obtained by querying all subsequent records after the index admission for TKR. Specifically, for each surgeon, we calculated the infection rate (in-hospital and within 90 days), re-admission rate (within 30 and 90 days), and revision surgery rate within the study period. These values were analyzed individually, as well as collectively (i.e., total adverse event rate), for each surgeon. In addition, we calculated the mean inpatient TKR length of stay for each surgeon. ICD-9 CM codes were used to identify surgical infections (996.66, 996.67, 998.3x, 998.5x, 998.6, and 998.83) and revision surgeries (81.55, 00.80, 00.81, 00.82, 00.83, and 00.84).

Individual surgeon profiles were then reviewed on the two most commonly visited, publicly available physician-rating websites: HealthGrades.com and Vitals.com [8]. All online data were collected between August 25, 2014, and February 24, 2015. Information recorded from HealthGrades.com and Vitals.com included physician age, years in practice, gender, number of ratings, and overall rating (out of five stars). Surgeons with no online ratings on Vitals.com (4) and HealthGrades.com (1) were excluded from the respective analyses, yielding 170 surgeons with Vitals.com and 173 with HealthGrades.com data.

Assessment of normality for the surgeon-specific patient outcomes from the SPARCS database were evaluated using Shapiro-Wilk test, results of which showed that all outcome data were not from a normally distributed population. Spearman's rank correlation coefficients (r_s) were calculated to determine whether surgeon-specific outcomes were correlated with mean overall surgeon ratings from each website. Mean overall surgeon ratings were then converted to ordinal categories for each website. Non-parametric Kruskal-Wallis

tests were conducted to evaluate differences in outcomes between the ordered categories. Statistical significance was defined as $p < 0.05$, and all analyses were performed using SPSS version 22.0 (IBM Corporation, Armonk, NY, USA).

Results

One hundred seventy-four high-volume surgeons each performed 180 or more TKRs from 2010 to 2012 in NYS. The median number of TKRs performed by these surgeons from 2010 to 2012 was 295 (interquartile range: 228 to 471). The mean number of years in practice was 24.6 (range: 6 to 50). The mean numbers of ratings for individual surgeons on HealthGrades.com and Vitals.com were 24.0 (range: 0 to 109) and 19.3 (range: 0 to 114), respectively. The mean overall ratings for individual surgeons on HealthGrades.com and Vitals.com were 4.2 and 4.1 (out of five) stars, respectively. Figure 1a demonstrates the distribution of mean overall ratings for individual surgeons on Vitals.com and HealthGrades.com.

Surgeon-specific TKR outcome measures calculated from the SPARCS database included the rate of infection (in-hospital and 90 days), re-admission (30 and 90 days), and revision surgery from 2010 to 2012 (Fig. 1b). The mean in-hospital infection, 90-day infection, 30-day re-admission, 90-day re-admission, and revision surgery rates from 2010 to 2012 were 0.25, 1.00, 4.89, 8.43, and 1.31%, respectively. Collectively, these outcome measures contributed to the total adverse event rate for individual surgeons from 2010 to 2012 (Fig. 1c). As with online patient ratings of individual surgeons, variability is observed in the total adverse event rate distribution for individual surgeons. The mean inpatient length of stay for TKR surgery was 3.6 (range: 2.3 to 6.2) days.

There was no significant correlation between individual surgeon online ratings on Vitals.com or HealthGrades.com with surgeon-specific total adverse event rate, inpatient length of stay, or any individual TKR outcome measure, except for the revision surgery rate from 2011 to 2012 with Vitals.com scores ($p = 0.049$). It should be noted that the surgeon-specific revision surgery rates during the entire study period (2010 to 2012) and/or from 2010 to 2011 were not correlated with Vitals.com scores.

Discussion

We found no correlation between online physician ratings and TKR surgeon-specific outcomes. There was a significant correlation between the surgeon-specific revision surgery rate from 2011 to 2012 with Vitals.com scores, but this correlation was not observed from 2010 to 2011 and/or during the study period, as a whole. These findings suggest that patients should use caution when interpreting physician-rating websites, thus supporting other, previously cited limitations and concerns regarding these websites. For example, there is no mechanism to confirm that online users are patients of the physician whom they have reviewed, and physician ratings are frequently based on a very small number of reviews. In addition, previous authors have noted that negative online written reviews more frequently cite factors

Fig. 1. **a** Distribution of mean overall ratings for individual surgeons on physician-rating websites. **b** Distribution of surgeon-specific total knee replacement outcome measures, 2010 to 2012, from the Statewide Planning and Research Cooperative System database: rates of infection (in-hospital and within 90 days), re-admission (within 30 and 90 days), and revision surgery. **c** Distribution of total adverse event rate for high-volume total knee replacement surgeons in New York State.

independent of physician competence (e.g., wait times, cost, office staff), thus highlighting differences in how patients and physicians may assess care quality [11]. Presumably, physician assessment of care quality is more closely linked to clinical outcome, while patient assessment can be impacted by other factors such as wait time, interaction with office staff, and subjective qualities of the interpersonal interaction with the physician.

Our study has several important limitations. First, the generalizability of our findings is limited by the narrow scope—i.e., physicians from one medical specialty in one state over a three-year time period. With regard to limitations in the geographic and chronologic scope, it should be noted that the TKR outcome measures presented—i.e., 30-day re-admission rate (4.89%) and length of stay (3.6 days)—are consistent with recent literature [2]. Furthermore, we intentionally studied a specific subspecialty with objective, clinically relevant short-term outcome measures. A second study limitation is that there is no mechanism to confirm that the users posting online ratings actually were patients who underwent TKRs. We attempted to minimize this variable by limiting our study population to only high-volume TKR surgeons (i.e., minimizing the chance that another procedure was performed). A third study limitation is that the SPARCS database does not contain important surgical

outcome measures such as pain, range of motion, and/or validated functional outcome scores.

Previous studies have attempted to correlate online physician ratings with various surrogates of care quality. Gao et al. examined over 386,000 online ratings and determined that patients gave “slightly higher” (but statistically significant) online ratings to physicians who were board-certified and graduated from a highly ranked medical school (top 50 according to *U.S. News & World Report*) but found no association with malpractice claims [4]. However, none of these variables have been associated with the quality of care provided by an individual physician. Greaves et al. examined over 16,000 family physician ratings in the UK on the English National Health Service Choices website and determined that the association with clinical outcome measures documented by the National Health Service was weak [5].

A unique advantage of this study is that we isolated a homogenous group of physicians (high-volume TKR surgeons) performing an elective surgical procedure with relevant short-term outcomes and have access to a robust database (SPARCS), which tracks patients longitudinally and allows assessment of surgeon-specific outcomes. Interestingly, Cram et al. concluded that there were no significant differences in outcomes (i.e., postoperative complication rate, re-admission rate, or inpatient length of stay) in Medicare

patients undergoing TKR at “top ranked” versus “non-top-ranked” hospitals according to *U.S. News & World Report* [1].

In conclusion, this study provides evidence, at a provider level, that online patient ratings do not correlate with physician-specific outcomes. The strength of this study’s conclusions is rooted in the relative homogeneity of our study population, clinically relevant short-term outcome measures following TKR, and robust database that tracks individual patients and physicians. Given these conclusions, patients should exert a high degree of caution when interpreting data from physician-rating websites.

Compliance with Ethical Standards

Conflict of Interest: Samir K. Trehan, MD, Ting J. Pan, MPH, and Aaron Daluiski, MD, declare that they have no conflicts of interest. Robert Marx, MD, reports the following relationships outside this work: research support from DePuy Synthes; book royalties from Springer and Demos Health; and editorial board memberships with *Journal of Bone and Joint Surgery*, *Evidence-Based Orthopedics*, and *HSS Journal*; he also is chairperson of the newsletter committee of the International Society of Arthroscopy, Knee Surgery and Orthopaedic Sports Medicine. Michael B. Cross, MD, reports the following relationships outside this work: Acelity (paid consultant, surgical advisory board), Exactech (paid consultant), Intellijoint Surgical (paid consultant, stock options), Link Orthopaedics (paid consultant), Smith and Nephew (paid consultant, research support), Theravance Biopharma (paid consultant), Zimmer Biomet (paid consultant); he also reports editorial board membership with *Bone and Joint 360*, *Journal of Orthopaedics and Traumatology*, and *Techniques in Orthopaedics*. Joseph T. Nguyen, MPH, and Stephen Lyman, PhD, report receiving funds from the Clinical Translational Science Center (CTSC), National Center for Advancing Translational Sciences (NCATS), grant no. UL1-RR024996. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCATS, based in Rockville, MD.

Human/Animal Rights: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2013.

Informed Consent: Informed consent was waived from all patients for being included in this study.

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

References

1. Cram P, Cai X, Lu X, Vaughan-Sarrazin MS, Miller BJ. Total knee arthroplasty outcomes in top-ranked and non-top-ranked orthopedic hospitals: an analysis of Medicare administrative data. *Mayo Clin Proc.* 2012;87(4): 341–348.
2. Cram P, Lu X, Kates SL, Singh JA, Li Y, Wolf BR. Total knee arthroplasty volume, utilization, and outcomes among Medicare beneficiaries, 1991–2010. *JAMA.* 2012;308(12):1227–1236.
3. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. *J Med Internet Res.* 2013; 15(8): e187.
4. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients’ online ratings of their physicians over a 5-year period. *J Med Internet Res.* 2012;14(1):e38.
5. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, Millett C. Patients’ ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. *J Med Internet Res.* 2012;14(5):e146.
6. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Parental awareness and use of online physician rating sites. *Pediatrics.* 2014;134(4):e966–75.
7. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. *JAMA.* 2014;311(7):734–735.
8. 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;13(4):e95.
9. Osborne NH, Nicholas LH, Ghaferi AA, Upchurch GR, Jr, Dimick JB. Do popular media and internet-based hospital quality ratings identify hospitals with better cardiovascular surgery outcomes? *J Am Coll Surg.* 2010;210(1):87–92.
10. Osborne NH, Ghaferi AA, Nicholas LH, Dimick JB, Mph M. Evaluating popular media and internet-based hospital quality ratings for cancer surgery. *Arch Surg.* 2011;146(5):600–604.
11. Trehan SK, DeFrancesco CJ, Nguyen JT, Charalel RA, Daluiski A. Online patient ratings of hand surgeons. *J Hand Surg Am.* 2016;41(1):98–103.