

Research Article

Which Behaviors Generate The Best Reviews? A Sentiment Analysis of Online Reviews on AOSSM Surgeons

Justin E Tang^a, Ting Cong^b, Arielle Hall^c, Jun Kim, MD^d, James Gladstone, MD^e

Keywords: Sports Medicine Reviews, Online Reviews, Sentiment Analysis

<https://doi.org/10.60118/001c.87964>

Journal of Orthopaedic Experience & Innovation

Vol. 5, Issue 1, 2024

Background

Online surgeon reviews can significantly influence a patient's selection of a provider, and are important in the movement towards quality-based physician compensation models. Written reviews, however, are subjective and are thus difficult to quantitatively analyze. Sentiment analysis using artificial intelligence (AI) provides the ability to quantitatively assess surgeon reviews to provide actionable feedback.

Purpose

The objective of this study is to quantitatively analyze the online written reviews of AOSSM surgeons utilizing sentiment analysis and report trends in the most frequently used words in the best and worst reviews.

Study Design

Cross-sectional study using publicly-available online reviews

-
- a Justin Tang is a current fourth year medical student at Mount Sinai, working closely with several attendings and residents of the Orthopaedic department. His work is largely focused on large clinical database studies and Artificial Intelligence (AI) and machine learning studies.

[Conflicts of Interest Statement for Justin Tang](#)

- b Orthopaedic surgery fellow with academic interest in the cellular and metabolic underpinnings of the osteochondral injury – post-traumatic osteoarthritis axis, their roles in disease progression, their targets for early surgical intervention, and clinical-translational innovations thereof.

[Connect with Ting Cong on LinkedIn](#)

[Conflicts of Interest Statement for Ting Cong](#)

- c [Conflicts of Interest Statement for Arielle Hall](#)

- d Dr. Jun S. Kim is a fellowship trained spine surgeon at The Mount Sinai Hospital and an Assistant Professor in the Department of Orthopedic Surgery. Dr. Kim specializes in cervical spine and scoliosis surgery, using a combination of minimally invasive, open, and microsurgical techniques to treat disorders of the cervical, thoracic, and lumbar spine. Dr. Kim attended Cornell University and graduated with Bachelor of Science in Chemical and Biomolecular Engineering.

He received his MD from the Lewis Katz School of Medicine at Temple University, and completed orthopaedic surgery residency at Mount Sinai Medical Center. After his residency at Mount Sinai, Dr. Kim completed a spine fellowship in adult and pediatric scoliosis and deformity surgery at Columbia University Medical Center where he served as an assistant attending physician and a clinical instructor.

[Visit Dr. Kim's Website](#)

[Connect with Dr. Kim on LinkedIn](#)

[Conflicts of Interest Statement for Dr. Kim](#)

- e Chief of the Sports Medicine Division and Associate Professor of Orthopaedic Surgery. Special interests in ACL reconstruction and repair as well as the management of patellofemoral disorders from chondromalacia to instability to arthroplasty. Team physician for the USA Davis Cup Tennis Team and US Ski and Snowboard Team. Orthopaedic consultant to the US Open Tennis Championships.

[Visit the Open Payments Data Page for Dr. Gladstone](#)

[Conflicts of Interest Statement for Dr. Gladstone](#)

Methods

Online reviews and star-ratings of AOSSM surgeons were obtained from healthgrades.com and zocdoc.com. A sentiment analysis algorithm was used to compute sentiment analysis scores of each written review. Sentiment scores were validated against star-ratings. Positive and negative word and word-pair frequency analysis was performed to identify common items associated with high and low scores. A multiple logistic regression was run on clinically relevant phrases.

Results

Following the inclusion and exclusion criteria, 18,386 AOSSM surgeon reviews were analyzed for 2071 surgeons. There was no significant difference in sentiment scores by provider gender. Surgeons who are younger than 50 years old had more positive reviews (mean sentiment = +0.536 versus +0.458, $p < 0.01$). The most frequently used and meaningful bi-grams used to describe top-rated surgeons are words correlating with kindness, caring personalities, and efficiency in pain management; whereas, those with the worst reviews are often characterized as unable to relieve the pain of their patients. The multiple logistic regression was significant for several clinically relevant words that confer greater or less odds of an improved score. Pain is significantly correlated with a decreased odds of receiving a positive review and positive behavioral factors confer a greater odds of receiving a positive review.

Conclusion

Sentiment analysis provides a means of quantifying written reviews of surgeons, and analysis of the reviews. This study provides insight into factors contributing to positive reviews, especially surgeon confidence, staff friendliness, warm disposition, and pain relief.

Clinical Relevance

This study delineates factors that impact the public reviews on AOSSM providers.

This article, like all JOEI articles, is available for continuing education

Reflect Now!

[Click here to start your pathway to continuing education!](#)

INTRODUCTION

With progressive efforts to transition orthopaedic surgical practice in the United States from the traditional fee-for-service model to value-based (pay-for-performance) compensation models, there has been a push at the provider-level to understand, and ultimately quantify, what constitutes “value”, “performance”, and “quality” in orthopaedic surgery. With this trend, surgeons have become more attentive to, and perhaps sensitive to the topic of, the tactics and data that comprise these measures. As a provider, taking a step back and reflecting on what fundamentally constitutes good patient care and the teachings of the Hippocratic Oath, one can arrive at the conclusion that determination of “value”, “performance” and “quality”, relates in great part to the patient experience.

Patient experience, in one of its simplest and purest forms, exists as patient ratings of orthopaedic surgeons.

Over the course of the past decade, online rating platforms, such as healthgrades® and ZocDoc®, have taken off as a prevalent source of information for prospective patients to make informed decisions about the doctors they choose. This democratization of the patient-doctor relationship can, to a measurable degree, influence patient physician preference (Noble et al. 2006), and can lead to changes at a surgical practice level.

Quantifying the patient experience is not a novel concept. Prior studies have performed analyses correlating positive patient reviews to factors such as a surgeon’s social media presence (Donnally, Li, et al. 2018), academic rather than non-academic practice, and mid- rather than early- or late-career practice (Frost and Mesfin 2015). These studies were generally performed manually, by tabulating numerical “star ratings” and correlating this to qualities of the surgeon and their practice (Arthur, Etzioni, and Schwartz 2019; Ramkumar et al. 2017). A minority of studies ana-

lyzed the anecdotal accounts of online surgeon reviews in order to identify modifiable risk factors for poor patient experience (Liu et al. 2019). Furthermore, the majority of these studies were descriptive, and so provide less value to practice modification (Hong et al. 2019a). To date, there is little data in the literature summarizing, in a comprehensive and quantifiable manner, the important qualities of an orthopaedic sports medicine surgeon and their practice that lead to a good patient experience.

The “Valence Aware Dictionary and sEntiment Reasoner” (VADER) Sentiment analysis is a widely used and accepted python computer programming package used to obtain compound sentiment analysis scores of written texts (Hutto and Gilbert 2014). Written text is given a numerical score based on its positivity and negativity of language, performed against an established library of natural language. Using this powerful tool, we can distill online written patient accounts down to a number, providing measurable value to all existing online reviews of orthopaedic sports medicine surgeons. In this study, we sought to use VADER sentiment analysis against online reviews of American Orthopaedic Society for Sports Medicine (AOSSM)-member sports medicine surgeons, in order to identify the most important patient experiential factors that lead to higher surgeon ratings. We hope this information can aid in surgeon practice modification to better the patient experience.

METHODS

STUDY DESIGN

The names of the physicians in the AOSSM Society were extracted manually from the AOSSM directory. These names were correlated on Healthgrades® and Zocdoc®, and a web-scraper code was used to obtain the online written reviews and star-rating reviews for each surgeon. Some websites such as Google® reviews, Wellness®, Vitals®, and rateMD®, were excluded because they contained firewalls that prevented the mass web-scraping of the publicly-available data. Healthgrades® and Zocdoc® are two popularly used websites for reviews that permitted this data extraction. For the following portions of this study, star-rating reviews will refer to the reported ratings out of five stars given to surgeons on these websites. This data is publicly available, and star-rating reviews provide an overall average rating acting as a standardized reference value. In order to validate our sentiment analysis model, we correlated our automated quantitative sentiment analysis scores against the surgeon’s overall star ratings by linear regression. Inclusion criteria included all surgeons who were listed in the directory of the AOSSM website who also had online profiles on [healthgrades.com](https://www.healthgrades.com). Exclusion criteria included those surgeons who had no online star ratings, and those who had no written reviews.

The outcomes of the study are:

1. Primary outcome: Validation of our sentiment analysis model of anecdotal reviews by comparing against star ratings. This would permit use of the sentiment

analysis model for analysis of online reviews in order to identify patient experiential factors that contribute to good reviews.

2. Secondary outcomes: Compare sentiment analysis scores against other surgeon qualities including gender, age, time in practice, practice type, and practice geography. Also, additional review words and phrases that may affect these scores were determined *a priori* by author opinion and correlated to sentiment analysis.

SURGEON DEMOGRAPHICS

All of the demographic characteristics of the surgeons were also extracted from [healthgrades.com](https://www.healthgrades.com) and [zocdoc.com](https://www.zocdoc.com). The physician’s age and gender identity were pulled directly from what was reported online (Table 1). Physician state of practice, MD/DO status, and practice type (categorized as academic/hospital or health system employed, private practice, or military/VA) was pulled manually from online websites. Those without these characteristics listed were excluded from the analysis. The states of practice were then separated into six geographic regions (West, Midwest, Northeast, South, Other, and International). The practice location also separated into metropolitan areas vs. non-metropolitan areas. The four regions were defined based on categories previously established in literature (Kirkpatrick et al. 2017), however certain US states and territories were excluded, so an “other” category was made as well as an international one.

SENTIMENT ANALYSIS

The “Valence Aware Dictionary and sEntiment Reasoner” (VADER) package is able to take writing and assign numerical scores to the “sentiment” of each individual sentence and then output an overall average sentiment for a given paragraph. VADER was used to obtain scores of each written review for every surgeon on the [healthgrades.com](https://www.healthgrades.com) and [zocdoc.com](https://www.zocdoc.com) databases. It is able to take sentences as inputs in order to output a compound score based on positivity or negativity of specific words with equations to account for punctuation, capitalization, and modulators.

VADER SCORE CALCULATION

VADER relies on a previously defined dictionary of specific and common words that retain inherent positive or negative qualities. The dictionary was developed through ten independent human raters. These raters assigned scores ranging from -4 to +4 to this dictionary of words. On this scale, a score of 0 represented a neutral sentiment (Hutto and Gilbert 2014). The algorithm then takes the inputted sentences, scans for these specific words, and sums and normalizes the scores to between -1 to +1, where -1 indicates a worse sentiment and +1 indicates a positive sentiment.

VADER calculation also factors in modifiers to words. A positive or negative empirically derived mean is factored into the calculation when either an emphasizing adverb

Table 1. Demographic Data on AOSSM Surgeons Analyzed

Demographics		Counts
Gender	Male	1703 (92.8%)
	Female	132 (7.2%)
Age	<50 y/o	773 (47.6%)
	>= 50 y/o	850 (52.4%)
Geographic Location	West	351 Surgeons, 3459 Reviews
	Midwest	425 Surgeons, 4154 Reviews
	South	670 Surgeons, 7761 Reviews
	Northeast	401 Surgeons, 4213 Reviews
	Other**	9 Surgeons, 64 Reviews
	International***	17 Surgeons, 139 Reviews
Degree	MD	1798 (95.3%)
	DO/DPM	88 (4.7%)
Type of Practice	Private	1099 (58.6%)
	Academic/Hospital Employed	745 (39.8%)
	Military/VA	10 (0.5%)
	Other	20(1.1%)

*some physicians did not have certain demographics listed, as such not included in respective analyses

**other states/US territories include: Hawaii, Alaska, Virgin Islands, & Puerto Rico

***International countries include: Canada, Argentina, Brazil, Columbia, Germany, Israel, Japan, Santiago, and Taiwan

or negating adverb is used respectively. This means that phrases such as “very informative” would be given a higher score than just “informative.” Further, negation is also factored in when it precedes words within the VADER dictionary by reversing the sign during calculation. Therefore, positive sentiment words would be calculated as negative and vice versa.

MODEL VALIDATION

Linear regression analysis was performed comparing the average sentiment analysis score for every doctor to their average star score in order to show a correlation between calculated scores of this study and the online rating.

WORD FREQUENCY ANALYSIS

Frequencies of most used words recognized by NLTK are also reported. In order to focus this analysis mainly on words that reflect behaviors and practices of surgeons, some words that generally describe a patient’s experience but that were not clinically relevant were removed. For example, “great” and “worst” may be used to describe a patient’s visit with a provider, but it does not indicate what factored into those experiences.

MULTIPLE LOGISTIC REGRESSION

Finally, a multiple logistic regression was performed on clinically relevant keywords to determine the odds of affecting a sentiment score (Table 3). Specific clinically relevant words, when included in a review, increased or decreased the odds of receiving a sentiment analysis score >0.5, indicating a largely positive review.

STATISTICAL ANALYSIS

Student t-tests were performed to determine the relationship between demographic variables (age, gender, degree type) and average sentiment score of written reviews. A one-way ANOVA was used to test for potential differences among average scores with respect to geographic regions as well as practice type.

RESULTS

Out of 2623 AOSSM surgeons identified, following application of the inclusion and exclusion criteria, a subset of 2080 AOSSM surgeons were analyzed consisting of 19,664 online reviews.

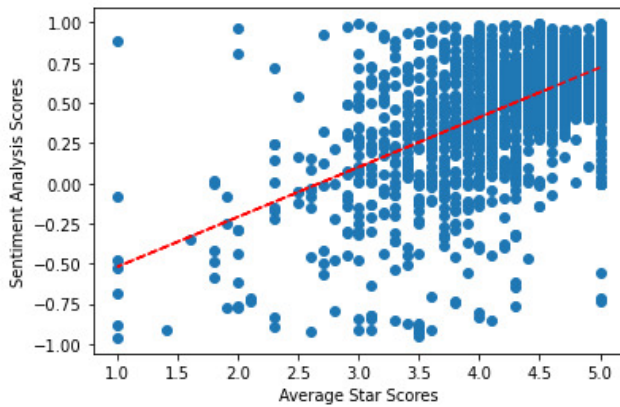


Figure 1. Linear regression analysis of average online reported star score compared to calculated sentiment analysis score

MODEL VALIDATION: LINEAR REGRESSION

The linear regression analysis of average sentiment analysis scores to average star scores showed a positive correlation between the scores (Figure 1, $r^2 = 0.586$, p -value < 0.01), indicating good concordance between sentiment scores and reported overall star-reviews.

DEMOGRAPHIC ANALYSIS

A Student T-test was run to check for a significant difference between the means of the sentiment analysis scores given to male and female surgeons. The test indicated an insignificant correlation between gender and greater or lower sentiment analysis scores (mean sentiments: male = +0.496, female = +0.546; $p = 0.132$). The student T-test performed on the average star scores compared to gender also returned insignificant (average stars: male = 4.29, female = 4.34; $p = 0.391$). These results are summarized in Table 2.

Further T-tests were conducted to check for a significant difference between means of sentiment analysis scores given to surgeons older and younger than the age of 50. The age cutoff of 50 years old was selected in order to establish as equal cohorts of surgeons under and above the cutoff as possible. The test indicated a significant correlation between older surgeons and lower sentiment analysis scores (mean sentiments: $< 50y = +0.536$, $> 50y = +0.460$; $p = < 0.01$). This also returned significant (average stars: $< 50y = 4.42$, $> 50y = 4.14$; $p = < 0.01$). This indicates that there is

a significant decrease in written sentiment about surgeons older than 50. These results are summarized in Table 2.

The average sentiment analysis scores for geographic regions were: West: 0.58, Midwest: 0.59, South: 0.59, Northeast: 0.57, Other: 0.54, International: 0.54 ($p=0.81$) indicated an insignificant difference among average sentiment analysis scores for all surgeons practicing in these regions. The average score of doctors practicing in metropolitan areas was 0.57 and the average score for non-metropolitan areas was 0.59 ($p = 0.08$), also showing no statistical difference. The student t-test for degree type also indicated an insignificant difference between how MDs and DOs are written about (MD Mean=0.59, DO Mean = 0.59; $p=0.97$). Finally, the ANOVA for practice type also returned insignificant (Private: 0.58, Academic/Hospital: 0.60, Military: 0.65; $p = 0.16$)

WORD FREQUENCY ANALYSIS

Frequencies of most used words recognized by NLTK are also reported. In order to focus this analysis mainly on words that reflect behaviors and practices of surgeons, some words that generally describe a patient's experience but that were not clinically relevant were removed. For example, "great" and "worst" may be used to describe a patient's visit with a provider, but it does not indicate what factored into those experiences.

MULTIPLE LOGISTIC REGRESSION

Finally, a multiple logistic regression was performed on clinically relevant keywords to determine the odds of affecting a sentiment score (Table 4).

DISCUSSION

In this study we used a computer algorithm to analyze, in a comprehensive and consistent manner, online review sentiments that correlate with good patient experience for AOSSM surgeons. As patients increasingly rely on online sentiments in choosing a practitioner for their treatment, this democratization of the patient-surgeon relationship should be well-understood by surgeons and be factored strategically into practice behavior.

According to Nielsen – a global performance management company – friends and family recommendation remains the most credible form of advertising, followed

Table 2. Student T-test comparing Star and Written Reviews to Gender and Age

	Male Average	Female Average	P Val
Written Reviews*	+0.496	+0.546	0.132
Star Reviews**	4.29	4.34	0.391
	>50 y/o Average	<50 y/o Average	P Val
Written Reviews*	+0.460	+0.536	<0.01
Star Reviews**	4.14	4.42	<0.01

*A more positive sentiment analysis number means a more positive review (scale of -1 to 1)

**On a scale for 1-5 stars, a greater number means a better review

Table 3. Clinically-Relevant Single Word Frequency Analysis of Best and Worst Reviews

Best Reviews		Worst Reviews	
Word	Frequency	Word	Frequency
Care	2395	Pain	1575
Caring	1381	Injury	332
Friendly	1191	Rude	256
Pain	1182	Care	249
Kind	970	Unprofessional	99
Bigram	Frequency	Bigram	Frequency
Pain free	380	No pain	238
Kind caring	230	Severe pain	92
Feel comfortable	228	Knee pain	91
Cares patients	222	Shoulder pain	90
No pain	208	Pain free	82

Table 4. Multiple logistic regression analysis on clinically relevant keywords

	2.5% CI	97.5% CI	OR	P val
Long Wait	0.172	0.718	0.351	<0.01
Pain	0.270	0.333	0.300	<0.01
No pain	0.667	1.321	0.939	0.718
Severe pain	0.174	0.743	0.359	<0.01
Pain free	2.537	5.202	3.633	<0.01
Friendly staff	1.267	23.422	5.448	0.02
Warm	1.866	7.562	3.756	<0.01
Confident	3.462	8.030	5.273	<0.01
Listens	1.533	2.429	1.930	<0.01
Bedside manner	1.300	2.239	1.706	<0.01
Knowledgeable	1.221	1.743	1.459	<0.01
Recover	0.947	4.326	2.02	0.06
Accessible	0.500	4.547	1.509	0.46
Front desk	0.757	2.199	1.290	0.34
Parking	0.323	7.738	1.582	0.57
Nurse	0.510	0.922	0.686	0.01
Receptionist	0.253	0.693	0.419	<0.01

closely by online reviews (“Audience Is Everything®,” n.d.). There is evidence that word of mouth reputation drives a significant portion of doctoral referrals in the United States (Tu and Lauer 2008). Specifically for orthopedic surgeons, prior literature has suggested that female sex, fewer years in practice or mid-career practice, and greater number of reviews are correlated with higher website physician ratings (Frost and Mesfin 2015; Heimdal et al. 2021). Though our data did not demonstrate sex differences in online ratings, we did find that younger surgeon age, less than 50, is correlated with higher sentiment scores. This may reflect cultural progression in the field of orthopaedic surgery as well as advancements in modern surgical training, which may make surgeons more likeable to patients, though there is little data to support this in the literature (Blasier 2009). One study demonstrated that academic practice is associ-

ated with better surgeon reviews (Frost and Mesfin 2015), which we were not able to replicate using our algorithm. Geographic and metropolitan versus non-metropolitan practice location did not significantly correlate to patient satisfaction. We also found no difference in DO and MD surgeon sentiment scores, though the small DO pool is a limiting factor.

Based on our word frequency analysis (Table 3), the current study agrees with previous, where phrases such as warm, confident, knowledgeable, and have proper bedside manner are associated with significantly higher overall review scores (Kalagara et al. 2019). As a result, this study and previous literature should encourage surgeons to continue to practice these interpersonal characteristics and place emphasis on these traits. Additionally, multiple logistic regression demonstrated that many positive behavioral char-

acteristics such as “warm,” “confident,” “listens,” and “bed-side manner” significantly improved the odds of receiving a positive review. It’s clear that patients are still cognizant and appreciative of these behavioral characteristics when exemplified by their surgeons. As a result, it is prudent for sports medicine doctors to be cognizant of these influences and continue to incorporate these facets into their medical doctrine.

Our multiple regression analyses of common review phrases revealed a number of points for discussion (table 4). Studies have indicated that the reviews written assigned to specific surgeons are often less reflective of the skills or about the surgeons themselves but rather factors such as office environment or the other ancillary staff (Donnally, McCormick, et al. 2018). Hong et al. in their study indicated that office efficiency, long wait times, staff interactions, and other non-physician related factors have significant contributions to online review (Hong et al. 2019b). This is also further reflected by the present study since the multivariate analysis indicated that when the phrase “long wait” was used in a review, it was a third as likely to receive a largely positive review from the patient. Additionally, comments that included mention of a “nurse” or “receptionist” also decreased the odds of receiving a positive review. This is likely variable, not to say that nurses and receptionists will always decrease the odds of a positive review, but rather indicating that the staff of a facility and not the physician is having an influence on the reviews attributed to the physician. To that end, reviews that included “friendly staff” were five times more likely to receive positive reviews. Both of these are in accordance with the results found in Hong et al.’s study where a positive interaction between patients and office personnel was often noted in reviews as well as negative non-physician specific related factors such as wait times (Hong et al. 2019b). As such, surgeons should be cognizant of the fact that despite their best efforts toward their own behavior, patients are often assessing their visits holistically. As a result, reviews may not solely be focused on the surgeon but could be a reflection of the office as a whole, thus indicating the need to emphasize the importance of these findings with all those involved in a patient visit.

The bigram and word frequency analysis indicates that one of the largest contributing factors driving both positive and negative reviews of sports medicine surgeons is in fact pain and pain management. Feucht et al. in their study on patient expectations prior to primary and revision anterior cruciate ligament surgery indicated that patients generally have very high post-op expectations in terms of returning to normal function and pain (Feucht et al. 2016). Additionally, throughout orthopedics it has been noted that unrealistic preoperative expectations that go unfulfilled after surgery results in significant dissatisfaction (Culliton et al. 2012; Iversen et al. 1998; Noble et al. 2006). This study highlights the importance and necessity of establish-

ing proper pain expectations prior to surgery and bolsters this growing body of literature surrounding pain expectations in orthopedics. This is highlighted by both the word frequency analyses and the multiple logistic regression as the top-rated surgeons, although are described with behavioral factors, are also being lauded for pain resolution, the most used bigram being “pain free.” The reviews reflect the fact that patients are dissatisfied most with lasting pain post-operatively, where if “pain” was utilized in any review, it had 0.30 odds of receiving a positive review. It is prudent, therefore, for surgeons to practice proper pain expectation management during pre-operative visits. By elucidating any misconceptions and addressing potentially unrealistic pain resolution expectations before surgery, physicians may be able to quell patient anxieties preemptively and prevent future dissatisfaction. Pain resolution is highly unpredictable and making this clear to patients early may improve the negative feedback being communicated online.

This study is not without its limitations. Some physicians, knowing the importance of social media, may have artificially altered their online profiles. As a result, their reviews displayed may not be representative of all of their patients. However, these reviews are those that future patients would be seeing and making educated decisions with, so the results of this study are still factors that should be considered in order to try to improve one’s own profile moving forward. Additionally, many surgeons may encourage their “successful” cases to leave them a review and neglect to ask others, contributing further to this artificial bias. Finally, there is a possibility that the reviews being written are not reflective of the actions of the doctor. Because pain is inherently difficult to treat and predict, a surgeon may have followed proper protocol to treating pain, but the patient may just not have seen complete resolution and written a review because of that experience. At the current time, there is no way to discern among patients who had lasting pain due to lack of care and those who have pain despite proper management.

CONCLUSION

In this study we used computerized sentiment analysis to calculate numerical scores for online written reviews for AOSSM surgeons, in order to identify qualities which improve online surgeon reviews. As expected, compassionate interpersonal skills and efficacy in controlling pain were associated with the best online review scores. These potentially modifiable factors can be incorporated into efforts to improve patient satisfaction and should be critically incorporated into surgeon practice.

Submitted: August 07, 2023 EDT, Accepted: September 14, 2023 EDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-ND-4.0). View this license’s legal deed at <https://creativecommons.org/licenses/by-nc-nd/4.0> and legal code at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> for more information.

REFERENCES

- Arthur, Jaymeson R., David Etzioni, and Adam J. Schwartz. 2019. "Characterizing Extremely Negative Reviews of Total Joint Arthroplasty Practices and Surgeons on Yelp.Com." *Arthroplasty Today* 5 (2): 216–20. <https://doi.org/10.1016/j.artd.2019.02.009>.
- "Audience Is Everything®." n.d. Nielsen. <https://www.nielsen.com/id/en/press-releases/2015/recommendations-from-friends-remain-most-credible-form-of-advertising/>.
- Blasier, Ralph B. 2009. "The Problem of the Aging Surgeon: When Surgeon Age Becomes a Surgical Risk Factor." *Clinical Orthopaedics & Related Research* 467 (2): 402–11. <https://doi.org/10.1007/s11999-008-0587-7>.
- Culliton, Sharon E., Dianne M. Bryant, Tom J. Overend, Steven J. MacDonald, and Bert M. Chesworth. 2012. "The Relationship between Expectations and Satisfaction in Patients Undergoing Primary Total Knee Arthroplasty." *The Journal of Arthroplasty* 27 (3): 490–92. <https://doi.org/10.1016/j.arth.2011.10.005>.
- Donnally, Chester J., III, Deborah J. Li, James A. Maguire Jr., Eric S. Roth, Grant P. Barker, Johnathon R. McCormick, Augustus J. Rush III, and Nathan H. Lebowitz. 2018. "How Social Media, Training, and Demographics Influence Online Reviews across Three Leading Review Websites for Spine Surgeons." *The Spine Journal* 18 (11): 2081–90. <https://doi.org/10.1016/j.spinee.2018.04.023>.
- Donnally, Chester J., Johnathon R. McCormick, Deborah J. Li, James A. Maguire, Grant P. Barker, Augustus J. Rush, and Michael Y. Wang. 2018. "How Do Physician Demographics, Training, Social Media Usage, Online Presence, and Wait Times Influence Online Physician Review Scores for Spine Surgeons?" *Journal of Neurosurgery: Spine* 30 (2): 279–88. <https://doi.org/10.3171/2018.8.spine.18553>.
- Feucht, Matthias J., Matthias Cotic, Tim Saier, Philipp Minzlaff, Johannes E. Plath, Andreas B. Imhoff, and Stefan Hinterwimmer. 2016. "Patient Expectations of Primary and Revision Anterior Cruciate Ligament Reconstruction." *Knee Surgery, Sports Traumatology, Arthroscopy* 24 (1): 201–7. <https://doi.org/10.1007/s00167-014-3364-z>.
- Frost, Chelsea, and Addisu Mesfin. 2015. "Online Reviews of Orthopedic Surgeons: An Emerging Trend." *Orthopedics* 38 (4): e257–62. <https://doi.org/10.3928/01477447-20150402-52>.
- Heimdal, Tyler R., Stephanie S. Gardner, Ujalashah M. Dhanani, Joshua D. Harris, Shari R. Liberman, and Patrick C. McCulloch. 2021. "Factors Affecting Orthopedic Sports Medicine Surgeons' Online Reputation." *Orthopedics* 44 (2): e281–86. <https://doi.org/10.3928/01477447-20201210-07>.
- Hong, Y Alicia, Chen Liang, Tiffany A Radcliff, Lisa T Wigfall, and Richard L Street. 2019a. "Metadata and Table Caption Correction: What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews." *Journal of Medical Internet Research* 21 (7): e14823. <https://doi.org/10.2196/14823>.
- . 2019b. "What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews." *Journal of Medical Internet Research* 21 (4): e12521. <https://doi.org/10.2196/12521>.
- Hutto, C., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 216–25. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- Iversen, Maura D, Lawren H Daltroy, Anne H Fossel, and Jeffrey N Katz. 1998. "The Prognostic Importance of Patient Pre-Operative Expectations of Surgery for Lumbar Spinal Stenosis." *Patient Education and Counseling* 34 (2): 169–78. [https://doi.org/10.1016/s0738-3991\(97\)00109-2](https://doi.org/10.1016/s0738-3991(97)00109-2).
- Kalagara, Saisanjana, Adam E.M. Eltorai, J. Mason DePasse, and Alan H. Daniels. 2019. "Predictive Factors of Positive Online Patient Ratings of Spine Surgeons." *The Spine Journal* 19 (1): 182–85. <https://doi.org/10.1016/j.spinee.2018.07.024>.
- Kirkpatrick, William, Jack Abboudi, Nayoung Kim, Juana Medina, Mitchell Maltenfort, Daniel Seigerman, Kevin Lutsky, and Pedro K. Beredjiklian. 2017. "An Assessment of Online Reviews of Hand Surgeons." *The Archives of Bone and Joint Surgery* 5 (3): 139–44.
- Liu, Cynthia, Meka Uffenheimer, Yosef Nasser, Jason Cohen, and Joshua Ellenhorn. 2019. "'But His Yelp Reviews Are Awful!': Analysis of General Surgeons' Yelp Reviews." *Journal of Medical Internet Research* 21 (4): e11646. <https://doi.org/10.2196/11646>.
- Noble, Philip C, Michael A Conditt, Karon F Cook, and Kenneth B Mathis. 2006. "The John Insall Award: Patient Expectations Affect Satisfaction with Total Knee Arthroplasty." *Clinical Orthopaedics and Related Research* 452 (November):35–43. <https://doi.org/10.1097/01.blo.0000238825.63648.1e>.
- Ramkumar, Prem N., Sergio M. Navarro, Morad Chughtai, Ton Jr. La, Evan Fisch, and Michael A. Mont. 2017. "The Patient Experience: An Analysis of Orthopedic Surgeon Quality on Physician-Rating Sites." *The Journal of Arthroplasty* 32 (9): 2905–10. <https://doi.org/10.1016/j.arth.2017.03.053>.
- Tu, H T, and J R Lauer. 2008. "Word of Mouth and Physician Referrals Still Drive Health Care Provider Choice." *Res Brief* 9:1–8.