

## Research Article

# Understanding a New Age of Physician Ratings via Sentiment Analysis: How are Patients Reviewing Shoulder and Elbow Surgeons Online?

Christopher A. White<sup>1a</sup>, Justin Tang<sup>1b</sup>, Akshar V. Patel<sup>2bc</sup>, Jordan D. Bernstein<sup>3d</sup>, Benjamin D. Gross<sup>e</sup>, Varun Arvind, PhD<sup>1f</sup>, Jun S. Kim, MD<sup>4g</sup>, Samuel K. Cho, MD<sup>3h</sup>, Paul J. Cagle, MD<sup>5i</sup>

---

a Christopher A. White is a current medical student at the Icahn School of Medicine at Mount Sinai who is hoping to pursue a career in orthopedic surgery.

[Conflicts of Interest Statement for Christopher A. White](#)

b Justin Tang is an MS3 at Mount Sinai interested in orthopedic surgery.

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

c [Conflicts of Interest Statement for Akshar V. Patel](#)

d Jordan Bernstein is a Clinical Research Coordinator in the Department of Orthopedics at Mount Sinai West. He received a BS in Neuroscience at the University of Wisconsin-Madison, and is passionate about contributing to research, health, and medicine.

[Conflicts of Interest Statement for Jordan Bernstein](#)

e [Conflicts of Interest Statement for Benjamin Gross](#)

f Dr. Varun Arvind, PhD is an MD-PhD student at the Icahn School of Medicine at Mount Sinai. He completed his undergraduate degree in biomedical engineering at Rutgers University where he studied the use of biomaterials for bone tissue engineering. Dr. Arvind participates in NIH funded research with a primary goal to further understand the cellular mechanisms of soft tissue injuries of tendons, ligaments, cartilage, and meniscus to improve care. Dr. Arvind's expertise is in the inflammation-stem cell cross talk that underpins regenerative vs fibrotic healing. In addition, he also leverages big data and artificial intelligence to utilize "big data" to further study and improve orthopedic care.

[Conflicts of Interest Statement for Varun Arvind](#)

g 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](#)

h Dr. Samuel Cho is a board-certified, fellowship-trained spine surgeon who specializes in the treatment of common and complex spinal disorders and deformities affecting adult and pediatric patients.

He currently serves as Chief of Spine Surgery at Mount Sinai West and the lead at the new Spine Center at Mount Sinai West. Dr. Cho is a professor of Adult and Pediatric Spine Surgery at the Leni & Peter W. May Department of Orthopaedic Surgery and the Department of Neurosurgery at the Icahn School of Medicine at Mount Sinai.

Dr. Samuel Cho has authored over 250 scientific articles dedicated to improving the quality of spine care and patient outcomes. He regularly appears at national and international meeting for spine surgeons worldwide. Dr. Cho is also a committee member of the leading spine societies in North America.

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

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

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

---

## Journal of Orthopaedic Experience & Innovation

Vol. 4, Issue 1, 2023

---

### Background

Physician review websites have a significant influence on patients with regard to provider selection. The goal of this study was to utilize machine learning to understand what influences patient ratings for shoulder and elbow surgeons online.

### Methods

All reviews and ratings were obtained from [Healthgrades.com](https://www.healthgrades.com). The “Valence Aware Dictionary and sEntiment Reasoner” (VADER) software was used to score the ratings on a scale between -1 and +1. Word frequency analysis was also performed to provide context to the words used to describe surgeons; both positive and negative word frequencies were analyzed. A multiple logistic regression identified the odds these words/phrases were associated with a sentiment score >0.50. Both star ratings (out of 5 points) and sentiment scores were collected.

### Results

493 surgeons (6,381 reviews) were included. There was no difference in ratings based on gender or geography. Age analysis indicated that star ratings and sentiment scores were lower for older providers ( $p < 0.01$ ). “Pain” was the most commonly used phrase in both the best and worst reviewed surgeon reports. When a comment included the phrases “knowledgeable”, “confident”, “listen[s]”, “recommend”, or “comfortable”, surgeons’ online ratings were 1.6x, 2.7x, 3.2x, 2.6x, and 3.8x more likely to be positive ( $p < 0.01$ ). Approximately 1 out of 5 reviews included mention of ancillary characteristics (e.g., “wait”, “front desk”, “office”).

### Conclusion

This unique code allows surgeons to analyze their field, from the individual to health system level, to see how they are being reviewed online. For shoulder and elbow surgeons, this study showed that more positive online reviews were seen for surgeons who are younger, have reduced office wait times, and have overall positive patient perceptions. Pain and pain management were the primary determinants of overall scores. As online ratings can influence a patient’s choice of provider, surgeons should consider implementing these findings to optimize their practice.

## INTRODUCTION

Current machine learning algorithms, such as sentiment analysis packages, provide users with the capability of assessing the positivity and negativity of online consumer comments. These powerful analytics can be applied to any consumer field in order to help providers improve their product. This becomes increasingly important in the field of medicine as patients continually turn to the internet to

find both medical information and healthcare physicians (Bujnowska-Fedak 2015; Bujnowska-Fedak, Waligóra, and Mastalerz-Migas 2019). As recently as 2019, multiple studies have shown that upwards of 70% of patients use the internet to search for health related information (Bujnowska-Fedak 2015; Bujnowska-Fedak, Waligóra, and Mastalerz-Migas 2019; Health Online 2013). With the world moving towards a progressively online form of communica-

---

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

i [Visit Dr. Cagle's Website](#)

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

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

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



©2022 Pacira Pharmaceuticals, Inc., a wholly owned subsidiary of Pacira Biosciences, Inc. All rights reserved. PP-NP-US-1328 05/22

\*Advertisement

[Click here to learn more about Pacira](#)

tion following the global COVID-19 pandemic, we can justifiably anticipate that this trend will continue.

A common platform by which patients acquire and deposit information about physicians' online is through physician review websites (PRWs). The comments and ratings on these websites, like any industry, have been shown to have a strong influence on a person's willingness to select a certain healthcare provider (Fanjiang et al. 2007; Li et al. 2015). In particular, multiple studies in the present orthopedic literature have addressed personal and ancillary characteristics of orthopedic surgical practices that drive both positive and negative online ratings. Of the twenty published articles on PRWs for orthopedic surgery, as summarized by Bernstein et al., 35%, 25%, and 15% cover spine, general orthopedics, and hand surgery respectively (Bernstein and Mesfin 2020). However, there has only been a single study focused on the content found on these websites as it pertains specifically to shoulder and elbow surgeons (Bernstein and Mesfin 2020; Syed, Acevedo, Narzikul, et al. 2019). In an analysis of 299 American Shoulder and Elbow Surgeons (ASES) members, Syed et al. showed that proper bedside manners and staff courtesy were key drivers of positive online ratings (Syed, Acevedo, Narzikul, et al. 2019). They further showed that an overwhelming majority of online ratings for ASES members were found to be positive (Syed, Acevedo, Narzikul, et al. 2019).

Moreover, in a recent 2018 survey of 1,000 patients, 74% of responders reported that they would be less willing to see a provider if they had negative online reviews or ratings (n.d.). This becomes even more important when we consider the growing number of annual orthopedic consultations and shoulder specific procedures being conducted (n.d.; Shukla et al. 2021; Best et al. 2021). With such a high number of patients turning to the internet, it is imperative that we fully understand the comments left by patients on PRWs for shoulder and elbow surgeons. Therefore, the goal of this study was to utilize machine learning to better understand the quantitative and qualitative ratings that drive positive and negative reviews for ASES members online, while also providing result-guided suggestions for practice improvement. To our knowledge the present study is the first to utilize sentiment analysis to analyze online written reviews for shoulder and elbow surgeons. In previous reports that utilized sentiment analysis to study orthopedic spine and hand surgeons, surgeon age, surgeon personal characteristics, ancillary practice characteristics, and pain management were all significant predictors of overall ratings and sentiment scores (Tang et al. 2022, 2021). We hypothesize that similar trends will be seen across the shoulder and elbow surgeon population.

## MATERIALS AND METHODS

### ASES SURGEON DATA

The surgeons analyzed here were obtained from those publicly listed as members on the ASES website. Those without online profiles on [Healthgrades.com](https://www.healthgrades.com), usually those who are residents or international members, or with fewer than three online written reviews were excluded from the study. Both the written reviews and star-ratings of surgeons were obtained from [Healthgrades.com](https://www.healthgrades.com). Healthgrades was chosen because it has publicly available data, permitting the acquisition of a large number of reviews and ratings. Further, many websites utilize firewalls to prevent the scrapping of large amounts of data from their interface. [Healthgrades.com](https://www.healthgrades.com) was selected because it permitted data acquisition without restrictions. Additionally, the ages and genders of the physicians listed on the websites were recorded. The states of the surgeons were manually searched on the ASES website and confirmed through online searches. For the regional analysis, the states, excluding Hawaii and Alaska, were separated into four geographic regions (west, northwest, north, and south) as previously defined by Kirkpatrick et al (Kirkpatrick, Abboudi, Kim, et al. 2017).

### SENTIMENT ANALYSIS CALCULATION

The "Valence Aware Dictionary and sEntiment Reasoner" (VADER) sentiment analysis was utilized to obtain sentiment analysis scores of written texts. VADER, developed by Hutto et al., utilizes a preformed dictionary that was curated by the authors for the software's package (Hutto and Gilbert 2014). This is a fully open-sourced software, as such there are no financial connections with any of the authors. The package is able to take sentences or paragraphs and output a score that represents the overall sentiment. This score is normalized to a scale of -1 to +1, where -1 represents a more negative sentiment and +1 represents a more positive sentiment. This sentiment scale is continuous and a zero on this scale represents a neutral sentiment. Although other sentiment analysis codes can accomplish similar analyses to VADER, it was selected based on its specificity for social media sentiments and its open-source nature, allowing us to cater the code for this specific analysis.

The score calculation is also able to interpret parts of speech and adjust sentiment calculations accordingly. It factors in the words that it recognizes from the library, punctuation, excessive capitalization, and adverbs, which can serve to change the meanings of words. Online reviews often have extensive punctuation (e.g., "!!!!") or excessive

capitalization (e.g., “AMAZING”). The VADER program has been trained to interpret these accordingly where an entirely capitalized “AMAZING” would contribute more to the overall score of a sentence than “amazing” since capitalization such as this is often used for emphasis. Adverbs such as “not” or “very” are also able to be accounted for. For example, “not friendly” changes the connotation of a previously positive adjective, “friendly”, to a negative one. As a result, VADER is able to flip the numerical contribution to a negative number when negating adverbs such as not is used, and the same is true for the other case of “very” and adverbs used to amplify the connotation of their following adjective. This means that “very rude” and “very helpful,” although both use the same adverb “very,” simply contribute more negative and more positive scores respectively.

A word frequency analysis was also performed. Before the analysis was conducted, words that were not clinically or behaviorally relevant were removed. For example, words such as “great” or “horrible” were removed in order to focus on characteristics that would aid in guiding physician behaviors. Although these words describe the patients’ experiences with their physicians, they do not aid in the analysis of what behavioral or ancillary characteristics are affecting sentiment scores. These words still contributed to the scores outputted by the program, but they were selectively removed from just the frequency analysis since it would not contribute to the understanding of the motives behind certain reviews. In order to provide more context to these individual words, a bigram frequency analysis was also conducted. A bigram analysis identifies the most commonly used word-pairs found in patient comments and outputs a sum for the overall word-pair frequency; bigrams for the remainder of this study will refer to two-word phrases.

#### MODEL VALIDATION AND DATA ANALYSIS

Linear regression analysis was performed in python version 3.8.2 in order to assess the average sentiment analysis score for each doctor to their average reported online five-point star score. This was conducted as a proof of concept for the sentiment analysis results, allowing the generated sentiment scores to be directly compared to publicly available star-ratings; quantitative star-ratings are considered the “gold standard” for online ratings. Sentiment scores provide additional qualitative insight to overall ratings. Student t-tests assessed the relationship between age and gender to the average sentiment analysis scores. Similar to the methods used in a recent study by Tang et al., surgeons were stratified into four cohorts (<40, 40-49, 50-59, and >60 years old) for an age sub-analysis (Tang et al. 2022, 2021). A one-way ANOVA was performed on the four geographical regions to assess potential differences in the average scores of different locations. Members who did not have their gender, age, or region of practice available on [Healthgrades.com](https://www.healthgrades.com) were removed only from these specific analyses; they were included in the whole cohort results. Finally, a multiple logistic regression was performed on specific words and phrases to analyze their effects on the overall review receiving a sentiment score >0.50.

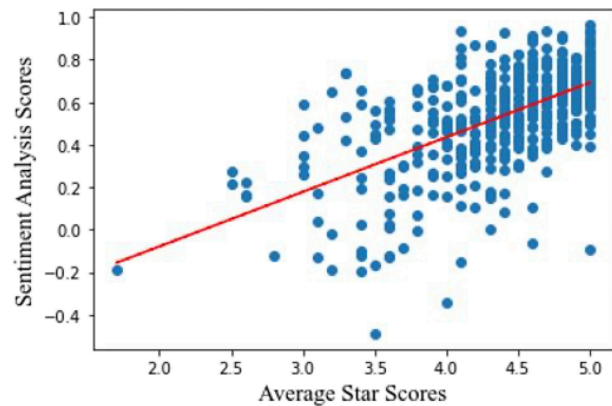


Figure 1.

## RESULTS

All data was collected from online records present as of July 2021. 760 shoulder and elbow surgeons were originally identified. 267 (35.1%) surgeons were removed because they had fewer than three reviews on [Healthgrades.com](https://www.healthgrades.com). 493 (64.9%) ASES member online profiles, consisting of 6,381 reviews, were included in the model validation, word frequency, and multiple logistic regression results. For sub-analyses on surgeon gender, age, and region of practice, 65 (13.1%), 113 (22.9%), and 18 (3.7%) of the included surgeons, respectively, were removed due to those specific variables not being listed on [Healthgrades.com](https://www.healthgrades.com).

#### MODEL VALIDATION BETWEEN RATINGS AND REVIEWS

The linear regression indicated a significant positive correlation between five-point star ratings and average sentiment analysis scores of written reviews ( $r^2 = 0.60$ ,  $p < 0.01$ ). This indicates that the calculated sentiment analysis scores correlate significantly with online reported star-value ratings (Figure 1).

#### ASES MEMBER DEMOGRAPHIC ANALYSIS

Gender analysis results indicated no significant differences between male ( $n=404$ ) and female ( $n=24$ ) physicians in their online reported star score or their written review sentiment analysis scores. Four cohorts were compared for the age analysis, <40, 40-49, 50-59, and >60 years old, consisting of 58, 145, 105, and 72 members respectively. 123, 82, 105, and 165 members made up the northeast, west, midwest, and south regions respectively. The age analysis indicated that in both reported star ratings and sentiment analysis scores, older physicians were rated significantly worse. Additionally, an ANOVA analysis was performed to assess the average sentiment analysis scores for each geographical region. For each region, the reviews and star-reviews of all surgeons who practiced in states belonging to a region were averaged. The average sentiment analysis scores for the northeast, west, midwest, and south were +0.57, +0.59, +0.55, and +0.58, respectively ( $p=0.50$ ), and the average star scores were 4.4, 4.5, 4.5, and 4.5, respectively ( $p=0.99$ ).

**Table 1. ASES Member Demographics**

Demographics	Scores
Gender (n=428)	
Male	404 (94.4%)
Female	24 (5.6%)
Age (n=380)	
<40 y/o	58 (15.3%)
40-49 y/o	145 (38.2%)
50-59 y/o	105 (27.6%)
>60 y/o	72 (18.9%)
Geographical Distribution (n=475)	
Northeast	123 surgeons, 1619 reviews
West	82 surgeons, 1171 reviews
Midwest	105 surgeons, 1289 reviews
South	165 surgeons, 2302 reviews

Complete demographic data and results can be found in [Table 1](#) and [Table 2](#).

**DESCRIPTIVE TERMS FREQUENCY AND MULTIPLE LOGISTIC ANALYSIS**

The most frequently used word, regardless of review positivity, was “pain”. Top words for the best surgeons included positive behavioral factors such as “kind” and “caring,” whereas the top words for the worst surgeons included “problem” and “rude.” A list of the top 50 positive and negative terms can be found in the supplemental tables section (Supplementary Tables 1 and 2). The most commonly used bigram in the best reviews is “pain free” and similarly in the worst reviews it was “no pain”. Upon further analysis of ancillary characteristics that could be affecting reviews, each review was checked for the presence of the words “wait,” “staff,” “office,” “nurse,” “front desk,” and “assistant.” We found that 1,286 of the 6,381 reviews (20.1%) included portions of these ancillary characteristics. [Table 3](#) indicates all

the phrases that were significantly affecting the odds of receiving a positive review. For example, if “wait” was included in a review, there was a reduced odds of receiving a positive score (odds ratio (OR) (95% confidence interval (CI)): 0.64 (0.46-0.89), p<0.01). However, if “pain free” was utilized there was a greater odds of receiving a positive review (OR (95% CI): 2.01 (1.24-3.26), p<0.01). For an extensive list of single phrase and bigrams identified, see the Supplemental Tables.

**DISCUSSION**

There is substantial literature to date that has analyzed the ratings left for orthopedic surgeons across multiple PRWs. In general, orthopedic surgeons who patients perceive to have proper bedside manner, are knowledgeable, and have easy means for scheduling have received significantly higher online ratings (Bernstein and Mesfin 2020; Bakhsh and Mesfin 2014). For shoulder and elbow surgeons in particular, physician and staff courtesy were associated with positive marks while age and region of practice had no impact on overall surgeon ratings (Syed, Acevedo, Narzikul, et al. 2019). Currently, Syed et al. has produced the only study that has evaluated the ratings for shoulder and elbow surgeons online using quantitative star-ratings (Syed, Acevedo, Narzikul, et al. 2019). However, as more patients continue to use the internet to find physicians, and with additional comments being deposited on PRWs daily, it is imperative that we analyze both quantitative and qualitative reviews found publicly online. Using machine learning, this study reports on the largest sample of shoulder and elbow surgeon (493 ASES members) internet reviews and ratings in the current literature, and it is the only study to score patient written comments using sentiment analysis for these surgeons (6,381 reviews). The sentiment analysis model was internally validated as the gold standard ratings for PRWs, star-ratings, directly correlated to our generated sentiment scores ( $r^2 = 0.60$ ,  $p<0.01$ ).

**DEMOGRAPHIC INFORMATION**

Multiple demographic factors have been shown to influence the ratings orthopedic surgeons receive online. Broadly, the majority of studies that have analyzed PRWs for orthope-

**Table 2. Influence of Surgeon Gender and Age on Star and Written Reviews**

	Male Average	Female Average	P-value		
Written Reviews	+0.54	+0.57	0.54		
Star Reviews	4.4	4.3	0.57		
	< 49 y/o Average	40-49 y/o Average	50-59 y/o Average	>60 y/o Average	P-value
Written Reviews	+0.64	+0.56	+0.50	+0.50	<0.01
Star Reviews	4.8	4.5	4.3	4.2	<0.01
	Northeast Average	West Average	Midwest Average	South Average	P-value
Written Reviews	+0.57	+0.59	+0.55	+0.58	0.50
Star Reviews	4.4	4.5	4.5	4.5	0.99

**Table 3. Multiple logistic regression analysis on specific clinically relevant words**

Written Term	2.5% CI	97.5% CI	OR	P-value
Wait	0.46	0.89	0.64	<0.01
Pain	0.41	0.58	0.49	<0.01
No pain	0.33	0.82	0.52	<0.01
Relief	0.64	3.24	1.44	0.37
Severe pain	0.09	0.77	0.26	0.02
Pain free	1.24	3.26	2.01	<0.01
Staff	1.67	2.45	2.02	<0.01
Accessible	0.12	11.07	1.15	0.90
Confident	1.63	4.44	2.69	<0.01
Listens	1.88	5.55	3.23	<0.01
Bedside manner	0.95	2.41	1.51	0.08
Knowledgeable	1.18	2.29	1.64	<0.01
Waiting room	0.15	1.24	0.43	0.12
Front desk	0.72	42.07	5.50	0.10
Parking	0.03	6.90	0.43	0.55
Nurse	0.50	1.47	0.86	0.59
Receptionist	0.09	0.86	0.28	0.03
Medication	0.07	0.68	0.23	0.01
Pain medication	0.58	95.72	7.45	0.12
Family	1.41	4.10	2.40	<0.01
Children	0.07	7.02	0.68	0.75
Recommend	2.19	3.02	2.57	<0.01
Refer	0.65	3.95	1.60	0.30
Return	0.62	1.67	1.02	0.94
Activities	0.55	2.40	1.15	0.70
Courtesy	0.20	17.43	1.84	0.59
Young	0.45	1.87	0.91	0.80
Old	0.35	0.53	0.43	<0.01
Conscientious	0.05	1.46	0.28	0.13
Organized	0.16	1.42	0.48	0.18
Comfortable	2.26	6.43	3.81	<0.01
Dress	0.82	3.39	1.66	0.16
Arthroscopy	0.12	12.8	1.25	0.85

dic surgeons have shown that younger surgeons received online reviews at higher rates and ones that were more positive than those for older surgeons (Bernstein and Mesfin 2020; Donnally, McCormick, Li, et al. 2018; Jack et al. 2017; Heimdal et al. 2021; Runge et al. 2020; Damodar et al. 2019). This study agrees with these findings as we showed that ASES members who were younger received significantly higher star-ratings and sentiment scores. Further, we also found that comments containing the term “old” resulted in surgeons having more than a 50% lower chance of receiving a positive written evaluation online (OR (95% CI): 0.43 (0.35-0.53),  $p < 0.01$ ). Yet, in their analysis of 299 shoulder and elbow surgeons, Syed et al. indicated that surgeon age had no impact on the ratings they received (Syed, Acevedo, Narzikul, et al. 2019). Thus, to date, a discrepancy exists between our findings and those of the only

other analysis of PRWs for shoulder and elbow surgeons; increased research is still needed on this subset of surgeons. Further, there is currently no literature that correlates differences in star- and or sentiment-scores to clinical outcomes or patient experiences. While the present study can analyze statistical differences overall, we are unable to determine what level of difference a patient considers important. Thus, while the age of the surgeons was found to be a statistically significant predictor of overall score, a difference of 0.07 in sentiment analysis and 0.30 in star rating scores may not be detectable by patients reading these reviews online.

In part, however, these results may be attributed to the fact that younger surgeons receive ratings and reviews at a higher rate online compared to more senior surgeons, providing room for increased review variability and dynamic-

**BE A PART OF BETTER ORTHOPAEDIC CARE.  
CLICK HERE TO LEARN HOW.**



**THE ORTHOPAEDIC IMPLANT COMPANY**

\*Advertisement

ity (Bernstein and Mesfin 2020; Bakhsh and Mesfin 2014). Donnally et al. and Damodar et al. partially attribute these increased rates to the greater use of social media by younger physicians, comparatively (Hutto and Gilbert 2014; Heimdal et al. 2021). However, it must be acknowledged that the rate at which one obtains reviews does not necessarily correlate with review positivity (Hutto and Gilbert 2014). Nevertheless, given that our study demonstrates a negative correlation between review positivity and age, all while using the most up to date and publicly available reviews, it is important that the practicing shoulder and elbow surgeon is aware of the current PRW landscape that their patients are using.

Further, our present study shows that female surgeons and male surgeons achieved comparable ratings (female: 4.3/5.0 vs. male: 4.4/5.0;  $p=0.57$ ) and sentiment scores (female: +0.57 vs. male: +0.54;  $p=0.54$ ). Individual analyses of multiple orthopedic subspecialties (i.e., spine, hand, hip and knee) support a similar notion (Kirkpatrick, Abboudi, Kim, et al. 2017; Runge et al. 2020; Melone et al. 2020). However, other studies indicate that female orthopedic surgeons may receive higher ratings compared to their male counterparts (Heimdal et al. 2021; Nwachukwu et al. 2016). Alternatively, Syed et al. also showed that female ( $n=11$ ) ASES members received lower ratings than male ( $n=288$ ) members (Syed, Acevedo, Narzikul, et al. 2019). Given these current discrepancies in the orthopedic literature, and the limited literature on shoulder and elbow surgeons, we propose more research needs to be conducted to clearly elucidate the impact surgeon gender plays on online remarks. This is especially true for our cohort as only 5.6% of the included ASES members were female, making it difficult to draw definitive conclusions as to the impact surgeon gender has on overall ratings.

In the present orthopedic PRW literature, surgeon region of practice was not related to overall rating scores. For general orthopedic surgeons, Frost et al. noted no differences in ratings by regional stratification, however they did find that the majority of reviews were written on surgeons practicing in southern states (Frost and Mesfin 2015). Similarly, Haglin et al. and Zhang et al. noted no differences in regional scores in their analyses of 282 and 209 spine surgeons respectively (Haglin et al. 2018; Zhang, Omar, and Mesfin 2018). While shoulder and elbow surgeons were found to have more reviews in the Northeast, a factor that the Syed et al. contributed to potential population densities or internet use patterns, there was no difference in ratings

between surgeons practicing in the northeast (8.5/10.0), west (8.3/10.0), midwest (8.4/10.0), or south (8.4/10.0) (Syed, Acevedo, Narzikul, et al. 2019). Similarly, we report that both star-ratings and sentiment scores do not significantly differ between surgeons practicing in the northeastern (4.4/5.0, +0.57), western (4.5/5.0, +0.59), midwestern (4.5/5.0, +0.55), or southern (4.5/5.0, +0.58) United States. These findings are despite the U.S. News and World Report consistently ranking more northeastern, midwestern, and western hospitals in their "Top 20 Best Hospitals for Orthopedics" compared to southern hospitals; in 2021 seven, seven, five, and one northeastern, midwestern, western, and southern based hospitals were in the top twenty respectively (U. S. News and World Report 2020).. Similar to what was seen in Canadian analyses, this could have led to variations in ratings by region. In Canada, physicians in the central and western provinces were found to have received worse ratings than those practicing in the eastern provinces where the majority of the "Best Hospitals Canada - 2021" are located (Liu, Matelski, and Bell 2018; "Best Hospitals 2021 - Canada" 2021). Nevertheless, as local differences in ratings were not noted in our analysis or in the aforementioned orthopedic PRW studies, we conclude that orthopedic surgeons are being rated similarly across the United States regardless of their geographic location of practice.

#### SURGEON PERSONAL AND ANCILLARY CHARACTERISTICS

Previous reports have also noted multiple surgeon specific characteristics that are associated with overall positive online ratings. In orthopedics in particular, surgeons who exemplify proper bedside manner, listening and communicative skills, knowledge, and trustworthiness are more likely to receive higher online ratings (Bakhsh and Mesfin 2014; Kalagara et al. 2019; Velasco et al. 2019). These same qualities were stressed in the sole PRW analysis of shoulder and elbow surgeons where higher remarks on proper bedside manner correlated with higher overall scores (Syed, Acevedo, Narzikul, et al. 2019). In our analysis, when a patient's comment included the terms "knowledgeable", "confident", "listen[s]", or "comfortable", surgeons' online ratings were significantly more likely to be positive. Further, we saw that surgeons who conveyed proper "bedside manner[s]" also trended towards receiving more positive reviews (OR (95% CI): 1.51 (0.95-2.41),  $p=0.08$ ). However, Syed et al. noted that in a survey of 100 shoulder surgeons, 74% reported infrequently checking their own PRW profiles

while another 94% reported making no changes to their practice based on these reviews (Syed, Acevedo, Narzikul, et al. 2019). Ultimately the culmination of these findings suggests that while known personal characteristics can drive review sentiment, surgeons may not be integrating these remarks into their medical practice. However, as Li et al. found that patients were more willing to select a provider if the physician had reviews highlighting positive social and technical skills, surgeons may want to consider following these reviews more carefully (Li, Lee-Won, and McKnight 2018). Moreover, as we demonstrate that reviews containing the term “recommend” were more likely to obtain a positive review, it is possible that stressing these important interpersonal skills will ultimately improve care and help physicians attract future patients.

Further, research suggests that the aesthetic aspects of a practice, such as physician’s dress or office design, can significantly influence a patient’s satisfaction with and perception of a provider (Kamata et al. 2020; LaVela et al. 2016; Rehman et al. 2005). We note multiple office and staff-based prompts significantly impacting review sentiment. Interestingly, reviews mentioning practice “staff” doubled the chances of receiving a positive review while those mentioning “receptionist[s]” or “wait[s]” resulted in significantly reduced odds of having a positive review. Previous PRW studies also found that long wait times, scheduling difficulties, unfriendly staff, and worse office environments (e.g. cleanliness) resulted in more negative online reviews (Bujnowska-Fedak 2015; n.d.; Zhang, Omar, and Mesfin 2018; Rehman et al. 2005), while ratings that stressed helpful and friendly staff significantly increased the odds of surgeons receiving positive ratings (Kalagara et al. 2019; Yu et al. 2020). These points become increasingly important when considering how comments and ratings on ancillary characteristics directly impact a surgeon’s personal PRW profile. In support of this point, Gao et al. showed that staff and physician online quality ratings were seen to be positively correlated ( $r=0.73$ ,  $p<0.01$ ) (Gao et al. 2012). Thus, ancillary characteristics of a practice can influence how a provider is reviewed even if what is being commented on is not related personally to the physician. In the limited ancillary terms that we scrubbed for, we showed that one in five reviews commented on aspects related to the office environment and staff. These findings should encourage surgeons to stress to their supporting staff the importance of patient centered care, and to review how their practice is being run, in order to optimize their online rating scores.

The word and bigram frequency analysis of the best and worst reviewed surgeons revealed that pain is a substantial contributor to a patient’s perception of the skills and satisfaction with a provider. This is also exemplified in the multiple logistic regression as “pain” decreased the odds of a positive score whereas “pain free” improved these odds. In their study of postoperative pain in lumbar spine surgery, Mancuso et al. indicated that patients who expected greater pain improvement, had documented depression, or had another surgery yielded less pain improvement (Mancuso et al. 2017). They concluded, similarly to this paper, that because pain is unpredictable and likely common despite

surgery, effective pain expectation management should become an emphasis for physicians prior to treatment. The present study supports these findings and reinforces the importance of patient preoperative and postoperative education on what is expected after surgery. Willingham et al. also indicate that patient attitudes have an association with their health outcomes, specifically endorsing negative expectations as being associated with lasting pain postoperatively (Willingham et al. 2021). Our study serves to add to this growing body of literature surrounding pain expectations following surgery. Through screening the bigrams of the worst reviewed surgeons and the results of the multiple logistic regression here, it is clear that pain and its severity populate these worst reviews. Thus, lasting pain is greatly influencing patient’s perceptions of their physicians. By preemptively addressing patient expectations of pain and clarifying any misconception that surgical intervention guarantees complete pain resolution, physicians may be able to prevent unnecessary patient dissatisfaction. Of note, this study was unable to determine whether patients were reviewing surgeons based on operations or in office visits and we were only able to collect data from a single PRW.

#### FUTURE DIRECTIONS

Overall, patients need to be cautious of relying on PRWs as well as the reviews provided on them as profiles are occasionally automatically generated and may not reflect updated data on the provider if they are unaware that this is happening. Further, the anonymity that the sites guarantee allows for patients to write potentially falsified reviews or accounts as there is no way for websites to ensure that they actually met with the provider (Murphy, Radadia, and Breyer 2019; Burkle and Keegan 2015; Ramkumar et al. 2017; Strech 2011). This same notion can be applied to the physician and potential marketing teams.

Ultimately, however, this multifaceted code allows surgeons to analyze not only their field, but their hospital system, department, and person to see how patients are reviewing and rating them online and in real time. Going beyond star-ratings, this quantitative approach to analyzing a previously qualitative metric allows ASES surgeons to adapt to an ever-changing field of orthopedic medicine through their patient’s written perspective. Further, we are able to screen through specialty and subspecialty specific phrases in order to see how strongly they impact a certain field’s review positivity. For example, while the term arthroscopy was not a significant predictor of overall review sentiment here, it acts to highlight the significance of this algorithm. Hypothetically, this sentiment analysis algorithm allows practices and individuals to determine if terms they stress in their care process are being stressed by the patients who review them online. Future directions for this project are seeking to develop an application or website that would allow surgeons to individually see statistics about their own PRW reviews in order to make decisions for adjusting care practices.

## CONCLUSIONS

The present study is the largest cohort analysis to date looking at internet ratings and written comments for shoulder and elbow surgeons, and it is the first sentiment analysis that has been conducted on this subgroup of orthopedic surgeons. Sentiment analysis provides physicians the ability to improve upon their practice, and to curtail their medical doctrine, using the specific comments of the patients

whom they treat. Ultimately, younger shoulder surgeons who patients perceive to be knowledgeable, confident, and attentive, while providing reasonable pain relief, are the most likely physicians to receive positive online reviews.

Submitted: November 25, 2022 EDT, Accepted: January 21, 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

- Bakhsh, W., and A. Mesfin. 2014. "Online Ratings of Orthopedic Surgeons: Analysis of 2185 Reviews." *Am J Orthop* 43. <https://pubmed.ncbi.nlm.nih.gov/25136868/>.
- Bernstein, David N., and Addisu Mesfin. 2020. "Physician-Review Websites in Orthopaedic Surgery." *JBJS Reviews* 8 (3): e0158. <https://doi.org/10.2106/jbjs.rvw.19.00158>.
- "Best Hospitals 2021 - Canada." 2021. 2021. <https://www.newsweek.com/best-hospitals-2021/canada>.
- Best, Matthew J., Keith T. Aziz, John H. Wilckens, Edward G. McFarland, and Uma Srikumaran. 2021. "Increasing Incidence of Primary Reverse and Anatomic Total Shoulder Arthroplasty in the United States." *Journal of Shoulder and Elbow Surgery* 30 (5): 1159–66. <https://doi.org/10.1016/j.jse.2020.08.010>.
- Bujnowska-Fedak, Maria M. 2015. "Trends in the Use of the Internet for Health Purposes in Poland." *BMC Public Health* 15 (1): 194. <https://doi.org/10.1186/s12889-015-1473-3>.
- Bujnowska-Fedak, Maria M., Joanna Waligóra, and Agnieszka Mastalerz-Migas. 2019. "The Internet as a Source of Health Information and Services." *Advances in Experimental Medicine and Biology*, 1–16. [https://doi.org/10.1007/5584\\_2019\\_396](https://doi.org/10.1007/5584_2019_396).
- Burkle, Christopher M., and Mark T. Keegan. 2015. "Popularity of Internet Physician Rating Sites and Their Apparent Influence on Patients' Choices of Physicians." *BMC Health Services Research* 15 (1): 416. <https://doi.org/10.1186/s12913-015-1099-2>.
- Damodar, Dhanur, Chester J. III Donnally, Johnathon R. McCormick, Deborah J. Li, Giuseppe V. Ingrassi, Martin W. Roche, Rushabh M. Vakharia, Tsun Y. Law, and Victor H. Hernandez. 2019. "How Wait-Times, Social Media, and Surgeon Demographics Influence Online Reviews on Leading Review Websites for Joint Replacement Surgeons." *Journal of Clinical Orthopaedics and Trauma* 10 (4): 761–67. <https://doi.org/10.1016/j.jcot.2019.01.021>.
- Donnally, C.J., J.R. McCormick, D.J. Li, et al. 2018. "How Do Physician Demographics, Training, Social Media Usage, Online Presence, and Wait Times Influence Online Physician Review Scores for Spine Surgeons?" *J Neurosurg Spine* 30:279–88.
- Fanjiang, Gary, Ted von Glahn, Hong Chang, William H. Rogers, and Dana Gelb Safran. 2007. "Providing Patients Web-Based Data to Inform Physician Choice: If You Build It, Will They Come?" *Journal of General Internal Medicine* 22 (10): 1463–66. <https://doi.org/10.1007/s11606-007-0278-1>.
- Frost, Chelsea, and Addisu Mesfin. 2015. "Online Reviews of Orthopedic Surgeons: An Emerging Trend." *Orthopedics* 38 (4): 257–62. <https://doi.org/10.3928/01477447-20150402-52>.
- Gao, Guodong Gordon, Jeffrey S McCullough, Ritu Agarwal, and Ashish K Jha. 2012. "A Changing Landscape of Physician Quality Reporting: Analysis of Patients' Online Ratings of Their Physicians Over a 5-Year Period." *Journal of Medical Internet Research* 14 (1): e38. <https://doi.org/10.2196/jmir.2003>.
- Haglin, Jack M., Adam E. M. Eltorai, Saisanjana Kalagara, Brandon Kingrey, Wessley M. Durand, Jessica P. Aidlen, and Alan H. Daniels. 2018. "Patient-Rated Trust of Spine Surgeons: Influencing Factors." *Global Spine Journal* 8 (7): 728–32. <https://doi.org/10.1177/2192568218767385>.
- Health Online. 2013. 2013. <https://www.pewresearch.org/internet/2013/01/15/health-online-2013/>.
- 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). <https://doi.org/10.3928/01477447-20201210-07>.
- 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>.
- Jack, R. A. II, M. B. Burn, P. C. McCulloch, S. R. Liberman, K. E. Varner, and J. D. Harris. 2017. "Does Experience Matter? A Meta-Analysis of Physician Rating Websites of Orthopaedic Surgeons." *Musculoskeletal Surgery*, 102 (1): 63–71. <https://doi.org/10.1007/s12306-017-0500-1>.
- 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>.
- Kamata, Kazuhiro, Akira Kuriyama, Vineet Chopra, Sanjay Saint, Nathan Houchens, Christopher M Petrilli, Latoya Kuhn, et al. 2020. "Patient Preferences for Physician Attire: A Multicenter Study in Japan." *Journal of Hospital Medicine* 15 (4): 204–10. <https://doi.org/10.12788/jhm.3350>.
- Kirkpatrick, W., J. Abboudi, N. Kim, et al. 2017. "An Assessment of Online Reviews of Hand Surgeons." *Arch Bone Jt Surg* 5:139–44.
- LaVela, Sherri L., Bella Etingen, Jennifer N. Hill, and Scott Miskevics. 2016. "Patient Perceptions of the Environment of Care in Which Their Healthcare Is Delivered." *HERD* 9 (3): 31–46. <https://doi.org/10.1177/1937586715610577>.
- Li, Siyue, Bo Feng, Meng Chen, and Robert A. Bell. 2015. "Physician Review Websites: Effects of the Proportion and Position of Negative Reviews on Readers' Willingness to Choose the Doctor." *Journal of Health Communication* 20 (4): 453–61. <https://doi.org/10.1080/10810730.2014.977467>.

- Li, Siyue, Roselyn J. Lee-Won, and Jessica McKnight. 2018. "Effects of Online Physician Reviews and Physician Gender on Perceptions of Physician Skills and Primary Care Physician (PCP) Selection." *Health Communication* 34 (11): 1250–58. <https://doi.org/10.1080/10410236.2018.1475192>.
- Liu, Jessica Janine, John Justin Matelski, and Chaim M Bell. 2018. "Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study." *Journal of Medical Internet Research* 20 (3): e76. <https://doi.org/10.2196/jmir.7475>.
- Mancuso, Carol A., M.C. Reid, Roland Duculan, and Federico P. Girardi. 2017. "Improvement in Pain After Lumbar Spine Surgery: The Role of Preoperative Expectations of Pain Relief." *The Clinical Journal of Pain* 33 (2): 93–98. <https://doi.org/10.1097/ajp.0000000000000383>.
- Melone, Giresse, James Jr. Brodell, Cesar Hernandez, Emmanuel Menga, Rishi Balkissoon, Ximing Liu, Jun Zhang, and Addisu Mesfin. 2020. "Online Ratings of Spinal Deformity Surgeons: Analysis of 634 Surgeons." *Spine Deformity* 8 (1): 17–24. <https://doi.org/10.1007/s43390-019-00012-4>.
- Murphy, Gregory P, Kushan D Radadia, and Benjamin N Breyer. 2019. "Online Physician Reviews: Is There a Place for Them?" *Risk Management and Healthcare Policy* Volume 12 (May):85–89. <https://doi.org/10.2147/rmhp.s170381>.
- Nwachukwu, Benedict U., Joshua Adjei, Samir K. Trehan, Brenda Chang, Kelms Amoo-Achampong, Joseph T. Nguyen, Samuel A. Taylor, Frank McCormick, and Anil S. Ranawat. 2016. "Rating a Sports Medicine Surgeon's 'Quality' in the Modern Era: An Analysis of Popular Physician Online Rating Websites." *HSS Journal*® 12 (3): 272–77. <https://doi.org/10.1007/s11420-016-9520-x>.
- 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>.
- Rehman, Shakaib U., Paul J. Nietert, Dennis W. Cope, and Anne Osborne Kilpatrick. 2005. "What to Wear Today? Effect of Doctor's Attire on the Trust and Confidence of Patients." *The American Journal of Medicine* 118 (11): 1279–86. <https://doi.org/10.1016/j.amjmed.2005.04.026>.
- Runge, Nicholas E., Jordan H. Jay, Franz H. Vergara, and Julius K. Oni. 2020. "An Analysis of Online Ratings of Hip and Knee Surgeons." *The Journal of Arthroplasty* 35 (5): 1432–36. <https://doi.org/10.1016/j.arth.2019.12.004>.
- Shukla, Dhairya, Sharvil Patel, Lesley Clack, Tyler B. Smith, and Michael S. Shuler. 2021. "Retrospective Analysis of Trends in Surgery Volumes between 2016 and 2019 and Impact of the Insurance Deductible: Cross-Sectional Study." *Annals of Medicine and Surgery* 63 (March):102176. <https://doi.org/10.1016/j.amsu.2021.02.022>.
- Strech, Daniel. 2011. "Ethical Principles for Physician Rating Sites." *Journal of Medical Internet Research* 13 (4): e113. <https://doi.org/10.2196/jmir.1899>.
- Syed, U.A., D. Acevedo, A.C. Narzikul, et al. 2019. "Physician Rating Websites: An Analysis of Physician Evaluation and Physician Perception." *Arch Bone Jt Surg* 7:136–42.
- Tang, Justin E., Varun Arvind, Christopher A. White, Calista Dominy, Jun S. Kim, Samuel K. Cho, and Amanda Walsh. 2021. "Using Sentiment Analysis to Understand What Patients Are Saying About Hand Surgeons Online." *Hand*, December, 155894472110604. <https://doi.org/10.1177/15589447211060439>.
- Tang, Justin E., Christopher A White, Varun Arvind, Samuel Cho, Jun S Kim, and Jeremy Steinberger. 2022. "What Are Patients Saying About Minimally Invasive Spine Surgeons Online: A Sentiment Analysis of 2,235 Physician Review Website Reviews." *Cureus* 14 (April):24113. <https://doi.org/10.7759/cureus.24113>.
- U. S. News and World Report. 2020. "Best Hospitals 2021." *Best Hospitals*, 2020.
- Velasco, Brian T., Bonnie Chien, John Y. Kwon, and Christopher P. Miller. 2019. "Online Ratings and Reviews of American Orthopaedic Foot and Ankle Surgeons." *Foot & Ankle Specialist* 13 (1): 43–49. <https://doi.org/10.1177/1938640019832363>.
- Willingham, Mark D., Molly R. Vila, Arbi Ben Abdallah, Michael S. Avidan, and Simon Haroutounian. 2021. "Factors Contributing to Lingering Pain after Surgery: The Role of Patient Expectations." *Anesthesiology* 134 (6): 915–24. <https://doi.org/10.1097/aln.0000000000003754>.
- Yu, Jessica, Linsen T. Samuel, Sercan Yalçin, Assem A. Sultan, and Atul F. Kamath. 2020. "Patient-Recorded Physician Ratings: What Can We Learn From 11,527 Online Reviews of Orthopedic Surgeons?" *The Journal of Arthroplasty* 35 (6): S364–67. <https://doi.org/10.1016/j.arth.2019.11.021>.
- Zhang, Jun, Adan Omar, and Addisu Mesfin. 2018. "Online Ratings of Spine Surgeons: Analysis of 208 Surgeons." *Spine* 43 (12): E722–26. <https://doi.org/10.1097/brs.0000000000002488>.
- N.d. Accessed August 13, 2021. [https://cms.doctor.com/wp-content/uploads/2018/05/patienttrends\\_report\\_v1.4.pdf](https://cms.doctor.com/wp-content/uploads/2018/05/patienttrends_report_v1.4.pdf).

## SUPPLEMENTARY MATERIALS

### **Supplemental materials**

Download: <https://journaloei.scholasticahq.com/article/68118-understanding-a-new-age-of-physician-ratings-via-sentiment-analysis-how-are-patients-reviewing-shoulder-and-elbow-surgeons-online/attachment/135595.zip>

---