AO SPINE

How Are Patients Reviewing Spine Surgeons Online? A Sentiment Analysis of Physician Review Website Written Comments

Justin E. Tang BS[1], Varun Arvind PhD[1], Calista Dominy BS[1] , Christopher A. White MA[1] , Samuel K. Cho MD[1] , and Jun S. Kim MD[1]

Abstract

Study Design: A Sentiment Analysis of online reviews of spine surgeons.

Objectives: Physician review websites have significant impact on a patient's provider selection. Written reviews are subjective, but sentiment analysis through machine learning can quantitatively analyze these reviews. This study analyzes online written reviews of spine surgeons and reports biases associated with demographic factors and trends in words utilized.

Methods: Online written and star-reviews of spine surgeons were obtained from healthgrades.com. A sentiment analysis package was used to analyze the written reviews. The relationship of demographic variables to these scores was analyzed with t-tests and word and bigram frequency analyses were performed. Additionally, a multiple regression analysis was performed on key terms.

Results: 8357 reviews of 480 surgeons were analyzed. There was a significant difference between the means of sentiment analysis scores and star scores for both gender and age. Younger, male surgeons were rated more highly on average ($P < .01$). Word frequency analysis indicated that behavioral factors and pain were the main contributing factors to both the best and worst reviewed surgeons. Additionally, several clinically relevant words, when included in a review, affected the odds of a positive review.

Conclusions: The best reviews laud surgeons for their ability to manage pain and for exhibiting positive bedside manner. However, the worst reviews primarily focus on pain and its management, as exhibited by the frequency and multivariate analysis. Pain is a clear contributing factor to reviews, thus emphasizing the importance of establishing proper pain expectations prior to any intervention.

Keywords

sentiment analysis, machine learning, patient reviews, patient satisfaction

Introduction

The field of orthopedic and neurosurgical spine surgery is one that sees constant development. Over the last two decades, there has been a continual increase in the volume of spinal procedures performed.[1-3] During this same time period, major technological advancements in the field, including minimally invasive surgery, computer assisted navigation, and robotic guidance, have changed the manner in which these procedures are conducted.[4,5] With these constant developments, it is necessary that we stay up to date with what patients are saying about the surgeons who conduct these practices online as their reviews are never stagnant.

[1]Department of Orthopedics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Corresponding Author:
Jun S. Kim, MD, The Orthopaedic Center, Icahn School of Medicine at Mount Sinai, Mount Sinai West 425 West 59th Street, 5th Floor, New York, NY 10019, USA.
Email: jun.kim@mountsinai.org

Unsurprisingly, orthopedic patients are frequently turning to the internet to both find and review medical information, clinics, and physicians.[6-8] Physician review websites in particular (e.g., Vitals.com, Healthgrades.com, Google.com, RateMDs.com) have emerged as popular platforms for patients to disseminate their opinions on the providers they have seen; in short, these review websites allow prospective patients to explore former patient experiences and reviews when choosing their future care provider. Presently, studies have begun to evaluate physician review websites across orthopedics in general and the majority of orthopedic subspecialties (i.e., Spine, Hand, Foot and Ankle, Arthroplasty, Sports, Shoulder and Elbow).[8] Given that information found on the internet, including physician review websites, has the ability to impact a person's thought process, it is vital that the orthopedic community understands what patients are saying about spine surgeons online.[9,10]

Multiple studies in the literature have analyzed patient reviews for spine surgeons on popular physician review websites. In terms of demographics, both younger surgeons and those who work in an academic setting were found to have more positive online reviews, while surgeon sex and region of practice had no impact on overall ratings.[11-15] Surgeon personal characteristics, including surgeon trustworthiness, punctuality, confidence, character, and bedside manner were associated with positive reviews and ratings.[12,13] However, the overall number of a surgeon's scientific publications were found to have no impact on patient ratings of surgeon characteristics such as trustworthiness.[16] Yet, practice-related characteristics such as ancillary staff friendliness, ease of scheduling, and office environments were found to be associated with worse overall ratings.[11,12,17,18] Further, to highlight the changing times, it was recently shown that surgeons with a social media presence are significantly more likely to have higher ratings across physician review websites;[13] a trend not seen in previous years.[12] This last point should serve as evidence that we need to continually evaluate the comments being deposited on these dynamic websites.

Of the aforementioned physician review website studies for spine surgeons, only the work of Donnally et al. and Kalagara et al. have analyzed written reviews,[13,17] while the rest assessed granular star-reviews or rating scores; the most recent study focused on physician reviews from 2019.[12] While there are currently eight publications that have analyzed physician review websites for spine surgeons, none have utilized a natural language processing program to analyze written reviews.[11-18] With new patient comments being left daily, and as the field of spine surgery continues to change, it is imperative that these findings are continually updated. The goal of this study was to use sentiment analysis to quantitatively analyze patient's comments and ratings of spine surgeons to determine what factors are most associated with both positive and negative reviews. Using written reviews up through 2021, we hypothesized that surgeons who are younger, provide sufficient pain relief, and who have positive personal characteristics (e.g., punctual, confident, etc.) will have the most favorable written reviews and ratings.

Methods

Data Acquisition

The website healthgrades.com was used to collect publicly available written and start-rating reviews of spine surgeons. Healthgrades was chosen as when generally searching for providers, this review website was one of the first few websites suggested as well as due to ease of ability to web scrape large amounts of data from the websites without restriction. Although other physician-rating websites exist, many have firewalls in place that prevent the use of code to extract large amounts of data. A web scraper code was developed in order to obtain the demographic data. This code "scrapes" data from websites after being given the URLs for each provider's healthgrade.com link. It is then able to parse through the HTML code in order to extract key demographic information needed for the study, such as provider age, gender, and reviews. Star-rating reviews refer to the reported ratings out of five stars given to surgeons from these websites. This data is publicly available, and the star-rating reviews provide an overall average star rating for each surgeon. Inclusion criteria included surgeons who were listed on the "The Physician Payments Sunshine Act" as "Spine Surgeons."[19] This list of surgeons was then also cross-checked online review websites to confirm that they were listed as spine surgeons within their online profiles as well. Exclusion criteria included those surgeons who had no online ratings or less than seven written reviews. Several linear regressions were performed to confirm the relationship between our calculated sentiment analysis scores and online reported star scores. These regressions were performed on the entire cohort, with no surgeons excluded, all the way to excluding surgeons with <20 reviews. Out of all iterations, surgeons with at least seven reviews provided the greatest fit out of all regressions, and as a result, this cutoff was selected in order to include as many providers as possible while accurately calculating average scores.

Sentiment Analysis

The "Valence Aware Dictionary and sEntiment Reasoner" (VADER) sentiment analysis is a widely used python package used to obtain sentiment analysis scores of written text. This package is built into the Natural Language Toolkit (NLTK) library.[20] The input for VADER is a written excerpt and the output is a score that represents the "sentiment" of the sentence. Sentiment is defined as how positive or negative a sentence is based on the prose utilized, the connotations of the words, punctuation, capitalization, and word modulators such as "very" or "not. VADER was used to obtain scores of each written review for every spine surgeon. Currently, VADER has seen use in the analysis of various modes of social media and has recently seen application into healthcare related social media.[21-23] It has been used for the analysis of Twitter tweets and other forms of social media in order to obtain insights into how the general public feels about particular subjects. As a result, VADER was selected for

this study as it has been utilized in many studies focusing on online prose, and thus lends itself to be an ideal program for the present study.

VADER Score Calculation

VADER relies on a word dictionary that was developed by ten independent human raters who were trained and assessed for inter-rater reliability. The raters assigned scores ranging from -4 to $+4$, with 0 representing a neutral sentiment, to each word in their dictionary.[20] VADER works by taking the inputted sentences, scanning for words included in the dictionary, and summing and normalizing the matching scores to between -1 to $+1$, where -1 represents negative sentiment and $+1$ represents positive sentiment. Scores are also affected by punctuation marks as well as capitalization, both of which are used for emphasis in online reviews.

Additionally, the calculation that VADER performs factors in potential modifiers to words. A negating or emphasizing adverb, such as "not" or "very," will impact the score of an overall text excerpt. This means that "not helpful" will contribute more negatively to overall sentiment while "very helpful" contributes more positively.

Model Validation

Linear regression analysis was implemented to compare the average sentiment analysis score for each doctor to their average star score in order to show an association between calculated sentiment scores of the written reviews and the online star rating.

Data Analysis

Student t-tests were completed when assessing how between demographic variables (age, gender) related to average sentiment scores of written reviews.

Word frequency analysis was performed to assess the most common content included in the surgeon reviews. Specifically, the most positive reviews (sentiment score $>.75$) and the most negative reviews (sentiment score <0) were independently assessed with word frequency analysis to determine what content was present in highly positive and highly negative surgeon reviews in particular. Further context was provided to the word frequency analysis by looking at the most commonly used word-pairs, or bigrams. This means that we searched every review for the frequencies of all two-word sequences.

Finally, we performed a multiple logistic regression on key words and word-pairs to analyze their association with a sentiment score $>.5$.

Results

Surgeon Demographics

Following the inclusion and exclusion criteria, 480 surgeons were included, leading to 8357 spine surgeon reviews being analyzed. All of the demographic characteristics of the surgeons were also extracted from healthgrades.com. The physician's age and gender identity were pulled directly from what was reported online (Table 1). The number of reviews per provider ranged from 8 reviews to 20 reviews, but a greater number of reviews was not shown to correlate with higher average sentiment scores.

Model Validation: Linear Regression

The linear regression analysis of average sentiment analysis scores to average star scores showed a positive association between the scores (Figure 1, $r^2 = .711$, P-value $< .01$), indicating good concordance between sentiment scores and reported overall star-reviews.

Model Validation and Demographic Analysis: Student T-Tests

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. There was a significant difference between gender and greater or lower sentiment analysis scores (mean sentiments: male $= +.437$, female $= +.353$; $P = .04$).

Table 1. Demographic Data on Spine Surgeons Analyzed

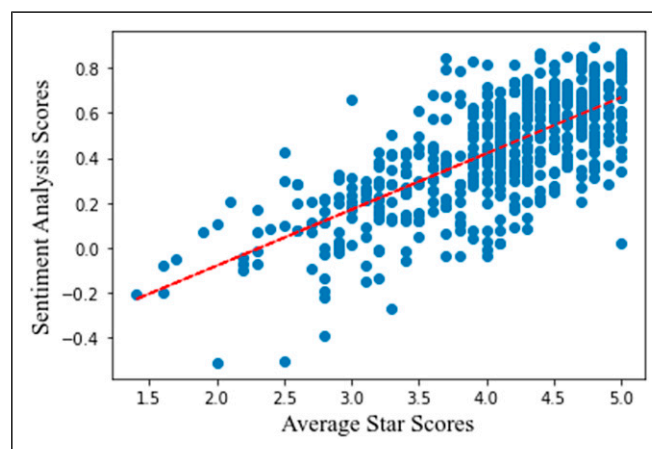| Demographics	Counts
Gender	
Male	440 (91.7%)
Female	40 (8.3%)
Age	
<50 y/o	163 (37.8%)
≥ 50 y/o	268 (62.2%)

Figure 1. Linear regression of average calculated sentiment analysis scores of each surgeon compared to their reported online star ratings.

This indicates that on average, male spine surgeons received higher sentiment analysis scores in their written reviews than their female colleagues. The average star scores compared to gender also returned significant (average stars: male = 3.97, female = 3.82; P = .03). 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. There was a significant difference between older surgeons and younger ones as, on average, older surgeons received lower sentiment analysis scores (mean sentiments: < 50y = +.486, >50y = +.396; p = **<.01**). The average star scores compared to gender also returned significant (average stars: < 50y = 4.59, >50y = +3.99; p = **<.01**). This indicates that there is a significant decrease in written sentiment about surgeons older than 50. These results are summarized in Table 2.

Word Frequency Analysis

Frequencies of most used words recognized by NLTK are also reported. The most frequently used and meaningful words used to describe top-rated surgeons are words associated with care, compassion, and comfort; whereas, those with the worst reviews are often characterized as rude, arrogant, and unable to relieve the pain of their patients. Words that were high frequency but not clinically or behaviorally relevant were removed in order to focus on characteristics that would be helpful in determining what factors affect patient reviews. For example, words such as "great" and "horrible" were removed because although they describe generally the experience the patient had with a physician, it does not aid in our analysis of what behavioral or practice characteristics are associated with these reviews.

After this initial filtration, we were able to identify the most frequently used words used in the best and worst reviewed surgeons. For surgeons who were most positively reviewed, their single-word descriptors mainly focused on qualitative and behavioral attributes such as "caring," "kind," "friendly," and "compassionate" (Table 3).

When reviewing the single-word descriptors, we were initially surprised by the high frequency of the word "pain" in the positively reviewed surgeons. In order to better understand why it was so frequently used, we performed a bigram analysis looking for the most commonly used word-pairs. The goal of this was to determine whether the surrounding words were affecting the context surrounding the words and thus affecting the score, pushing it to a positive contribution. In the bigram analysis of the most positive review, the most frequently used two-word sequence that was clinically/behaviorally relevant was "pain free," thus indicating the importance of pain management to patients. "No pain" was also the fifth most used and relevant bigram (Table 4). Both of these findings confirm the suspicions that the words preceding and following pain were affecting their connotation and thus appropriately being scored as positive attributes for the providers.

For the most negatively reviewed surgeons, their descriptors centered around levels of pain as well as inefficiency in pain management. Of the reviews used in this analysis, those reviews which had a negative sentiment analysis score, pain was used 1063 times, and the next relevant word was rude at 241 (Table 3). This shows that the clear factor driving negative reviews of spine surgeons is in fact pain and pain management. Although there are behavioral words in the most used words for the negative reviews, they are used significantly less frequently than pain is. Bigram analysis of negative reviews also supports this claim as all of the top 5 clinically relevant, highest frequency bigrams were about pain, descriptors of pain, or regions of pain (Table 4).

Multiple Logistic Regression

Finally, a multiple logistic regression was performed on clinically relevant keywords. This regression was performed to identify the likelihood that selected high frequency, clinically relevant words and bigrams would be included in a review with an overall sentiment analysis score of >.5

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

	Male average	Female average	P val
Written reviews	+.437	+.353	**.04**
Star reviews	3.97	3.82	**.03**
	>50 y/o average	**<50 y/o average**	P val
Written reviews	+.396	+.486	**<.01**
Star reviews	3.99	4.59	**<.01**

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

Best reviews		Worst reviews	
Word	Frequency	Word	Frequency
Care	967	Pain	1063
Pain	778	Rude	241
Caring	654	Care	196
Kind	426	Problem	164
Friendly	393	Unprofessional	76

Table 4. Clinically Relevant Bigram Frequency Analysis of Best and Worst Reviews.

Best reviews		Worst reviews	
Bigram	Frequency	Bigram	Frequency
Pain free	260	Pain free	116
Kind caring	96	Back pain	101
Feel comfortable	94	No pain	100
Staff friendly	89	Severe pain	56
No pain	82	Excruciating pain	52

(Table 5). A review with a score >.5 indicates that the review is largely positive for a particular surgeon.

The results of this regression showed us that words defining positive surgeon behaviors, such as "listens", "knowledgeable," "warm," and "confident", were positively associated with reviews that had positive sentiment scores. The more positive behaviors that were described by patients, the more likely she or he is to get a better review. This is shown through the top behaviors with the greatest odds ratios being "listens," "knowledgeable," "warm," and "confident," with odds ratios of 2.5 (P < .01), 2.97 (P < .01), 3.17 (P < .01), and 6.08 (P < .01), respectively, indicating that these words were associated with a 2x, 3x, and 6x chance of receiving an overall positive score if included in a review.

Regression analysis also showed that inclusion of "long wait" makes a surgeon's review half (.45) as likely to receive a positive sentiment score. Additionally, the inclusion of a "friendly staff" also did not significantly affect the odds of receiving a largely positive review.

Finally, inclusion of the words "pain" and/or "severe pain" were significantly associated with decreased odds of receiving positive reviews (OR = .372 and OR = .263, respectively), whereas the inclusion of "pain free" in surgeon reviews conferred a 4 times greater likelihood that a surgeon received a positive review (OR = 3.96).

Discussion

The field of orthopedics, especially the spinal subspecialty, is very dynamic in terms of both medical practice and surgical approach. Currently, spinal procedures are seeing worldwide increases in volume in conjunction with patients utilizing the internet (i.e., Physician Review Websites) to aid in their surgeon selection process.[1,2,7,8] Thus, in order for surgeons to provide the highest quality of care while successfully recruiting prospective patients, it is crucial that we understand how the spine community is being rated and reviewed online. This current study sought to appreciate the general vernacular being used to describe the highest and lowest rated spine surgeons, while simultaneously identifying characteristics associated with these scores. With 480 surgeons analyzed here, our study represents the second largest spine cohort analyzed using physician review websites. Further, our study processed the most written reviews (8,357) to date and is the only study to do so utilizing written sentiment analysis. Overall, we found that surgeons who are warm, confident, knowledgeable, and listen while providing sufficient pain relief receive the highest scores through online review platforms.

A consistent finding throughout the literature is that orthopedic surgeon's age is significantly associated to their rating positivity online. In general, younger surgeons saw more positive reviews across various orthopedic subspecialties.[8] This point remained constant in sub-analyses of spine surgeons, with younger surgeons amassing significantly higher ratings than older surgeons across multiple physician

Table 5. Multiple logistic regression analysis on clinically relevant keywords.

	2.5% CI	97.5% CI	OR	P val
Approachable	.37	5.57	1.44	.60
Accessible	.50	6.67	1.63	.50
Warm	1.22	8.19	3.17	**.02**
Confident	3.50	10.86	6.08	**<.01**
Listens	1.72	3.63	2.50	**<.01**
Bedside manner	1.16	2.44	1.68	**<.01**
Knowledgeable	2.11	4.18	2.97	**<.01**
Long wait	.13	1.57	.45	.21
Friendly staff	.17	15.03	1.61	.67
Severe pain	.11	.65	.26	**<.01**
Pain	.32	.43	.37	**<.01**
Pain free	2.55	6.13	3.96	**<.01**
Relief	1.58	4.42	2.64	**<.01**
Staff	1.51	1.97	1.73	.12
Open	.71	2.63	1.36	.35
Minimally invasive	.61	3.41	1.44	.40

review websites (e.g., vitals.com, Google.com, and healthgrades.com.[11,12,14] Our findings agree with the current literature as we saw that spine surgeons younger than 50 years of age have significantly more positive written reviews (<50: +.531, >50: +.423; P = .03) and star ratings (<50: 4.59/5.00, >50: 3.99/5.00; P < .01) compared to surgeons older than 50. Given this consistent finding for online reviews of spine surgeons, more experienced surgeons should consider adopting characteristics highlighted later in this study. Further, this remains an interesting point as past literature (articles not focusing on online reviews) has shown that patients indicate no preference for a physician based on their age;[24-26] yet patient's online comments suggest a different narrative. This same point can be stressed for patients' preference on their surgeon's gender. Previous literature on patient review websites has shown that patients report no preference for their spine surgeons based on surgeon gender.[14-16] However, offline studies have indicated that patients may prefer a male or female operating surgeon based on the subspecialty or the patient's gender identity.[27-29] Our results differ from the current physician review website literature. We saw that female surgeons received both lower written review scores (F: +.353, M: +.437; P = .04) and star ratings (F: 3.82/5.00, M: 3.97/5.00; P = .03) compared to male surgeons. However, given our cohort had a considerably larger male population (n = 440) compared to the female population (n = 40), we believe further research is warranted to draw strong conclusions regarding the impact of physician sex on review scores. Additionally, a gender-specific cohort split causes a selection bias, which is impacting the results we saw.

Previous literature has indicated that patients most desire orthopedic physicians who display behavioral qualities such as good bedside manner, listening, and trustworthiness.[8,30-32]

Additionally, Donnally et al., Melone et al., and Kalagara et al. showed that spine surgeons who are confident, punctual, helpful, and answer questions received the most positive ratings online.[13,15,17] In our sentiment analysis, we show that "warm," "confident," "listen," "knowledgeable," and "bedside manner" are all positively associated with surgeons receiving higher written review scores. These findings make intuitive sense as patient's want a surgeon who truly cares for their well-being. Elsewhere in the literature, however, non-personal and non–physician-related characteristics have been linked to alterations in post-visit patient satisfaction. For example, multiple subsidiary characteristics such as physician attire, clinic environment, and scheduling ease have all been shown to significantly influence a patient's healthcare experience.[28,33,34] For spine surgeons in particular, online ratings indicate that long wait times can drive down, while friendly staff help improve, surgeon ratings.[11-13,15] While we did not note "staff," "friendly staff," or "long wait" as having an influence on surgeon ratings or reviews, it has been shown that these variables significantly affect patient post-visit satisfaction in spine practices.[35,36] As (dis)satisfaction can impact how patients review spine surgeons online, and ultimately how future patients will perceive them, these ancillary points should be strongly considered by spine surgeons in order to optimize their overall practice.

In a similar study, Agarwal et al. studied the online ratings of emergency departments and urgent care centers utilizing co-occuring words in text analysis, similar to what was performed in the present study in our word and bigram frequency analysis and multivariate regression.[37] Comparatively, they found that the most positive reviews focused on effective pain management and improved communication. Our bigram frequency analysis and multivariate regression emphasizes this focus on "pain" and pain-descriptors as well. After screening over 8000 reviews for the word-couplings that were most common, pain and its locations inundated the negative reviews. It is clear that lasting pain and the patient's experiences with pain are greatly influencing their perception of their providers' abilities. Pain and pain management is an overwhelming focus in these reviews, indicating how critical it is for physicians to clarify any misconceptions that patients have prior to surgery and to perform proper pain expectation management. Soroceanu et al. in their study on preoperative expectations for patients undergoing lumbar and cervical spine surgery, found that patients who had fulfillment of their expectations had higher postoperative satisfaction.[38] Those who retained more preoperative expectations also had decreased postoperative satisfaction. Both of these results further emphasize how important it is that surgeons devote proper attention to pain expectation management. Mahomed et al. also indicated that 66% of individuals expected no pain after recovery from total joint arthroplasty procedures,[39] further emphasizing the misconceptions that patients may hold. By preemptively addressing pain and the inherent complexity in treating pain, physicians may be able to resolve any of the preconceived notions that surgery is a guaranteed cure to a patient's pain. Surgeons who address these head on are more likely to resolve these misconceptions and patients, if after the procedure they still feel pain, will be less likely to attribute this lasting pain to the lack of abilities of their provider but rather to the fact that pain resolution is inherently difficult and never ensured.

Moreover, the power of this present study is such that we can screen for any descriptive words or phrases and assess how they may influence a physicians' rating. This could prove to be a useful tool for surgeons trying to improve their individual online presence, especially as comments that do not pertain to a surgeon's person in particular can still reflect negatively on them. Further, in recent years minimally invasive and robotic-assisted procedures have gained immense traction within the orthopedic subspecialties, especially in spine. The market for robotic assistance in particular is estimated to reach $2.77 billion by 2022 as more and more surgeons adopt these technologies in their everyday practice.[40] Despite this immense increase in popularity of emerging technologies among providers and in literature, the interest has yet to significantly alter patient opinions. When specifically searching for the frequency of phrases, the word "robot" only appeared once in our entire set of 8357 reviews. Additionally, the multivariate regression for words associated with different types of procedures, "open," "minimally invasive," "robot," all returned insignificant. Future research needs to be performed as patients begin to comment on their experiences with these new surgical methods.

We do recognize that this study is not without its limitations. For this study, only one website, healthgrades.com, was utilized as it was the only website we were able to freely extract data from. This, has the potential to bias the results as they are only sourced from a single site. Additionally, given the subjective nature of the ratings and reviews left on public websites, we were unable to assess a patient's rationale for leaving specific comments. Additionally, we were not able to stratify comments that pertained to physician or non-physician specific remarks in the overall sentiment analysis scores. Further, using the "The Physician Payments Sunshine Act" to curate a list of surgeons meant that we were only able to run analysis on a subset of all spine surgeons. Future directions for this project could be the development of an app or website that would allow providers to input their own websites and reviews to output individualized data on their performance in order to see how patients are reviewing them online. As comments are uploaded and added daily, through a machine learning based app, surgeons would be able to actively monitor their reviews in real time and amend their practices as needed.

Conclusion

This study represents the largest single analysis of written comments left for spine surgeons on physician review websites. We saw that gender did not play a role in the sentiment of

the reviews left for providers in our analysis while age did. In general, positive behavioral attributes contributed significantly to positive reviews, as one would expect. However, although some negative behavioral characteristics contributed to the worst reviews, pain was the largest factor as seen in both the frequency analysis and the multiple logistic regression. This study thus serves to reinforce the importance of practicing proper pain expectation management prior to surgery in order to better serve patients and thus improve patient satisfaction and reviews.

Author Contribution

Declaration of Conflicting Interests

Funding

ORCID iDs

Calista Dominy, BS https://orcid.org/0000-0002-6572-585X
Christopher A. White, MA https://orcid.org/0000-0002-8582-6105
Samuel K. Cho, MD https://orcid.org/0000-0001-7511-2486
Jun S. Kim, MD https://orcid.org/0000-0002-6114-2673

References

1. Grotle M, Småstuen MC, Fjeld O, et al. Lumbar spine surgery across 15 years: Trends, complications and reoperations in a longitudinal observational study from Norway. *BMJ Open*. 2019;9(8):e028743. Doi:10.1136/bmjopen-2018-028743.

2. Kobayashi K, Ando K, Nishida Y, Ishiguro N, Imagama S. Epidemiological trends in spine surgery over 10 years in a multicenter database. *Eur Spine J*. 2018;27(8):1698-1703. Doi: 10.1007/s00586-018-5513-4.

3. Martin BI, Mirza SK, Spina N, Spiker WR, Lawrence B, Brodke DS. Trends in lumbar fusion procedure rates and associated hospital costs for degenerative spinal diseases in the United States, 2004 to 2015. *Spine (Phila Pa 1976)*. 2019;44(5):369-376.

4. Patel PD, Canseco JA, Houlihan N, Gabay A, Grasso G, Vaccaro AR. Overview of minimally invasive spine surgery. *World Neurosurg*. 2020;142:43-56.

5. Burrus MT, Werner BC, Starman JS, et al. Patient perceptions and current trends in internet use by orthopedic outpatients. *HSS J*. 2017;13(3):271-275.

6. Hautala GS, Comadoll SM, Raffetto ML, et al. Most orthopaedic trauma patients are using the internet, but do you know where they're going? *Injury*. 2021;52(11):3299-3303. doi:10.1016/j.injury.2021.02.029.

7. Yu J, Samuel LT, Yalçin S, Sultan AA, Kamath AF. Patient-recorded physician ratings: What can we learn from 11,527 online reviews of orthopedic surgeons? *J Arthroplasty*. 2020; 35(6S):S364-S367.

8. Bernstein DN, Mesfin A. Physician-review websites in orthopaedic surgery. *JBJS Rev*. 2020;8(3):e0158.

9. 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.

10. Li S, Lee-Won RJ, McKnight J. Effects of online physician reviews and physician gender on perceptions of physician skills and Primary Care Physician (PCP) selection. *Health Commun*. 2019;34(11):1250-1258.

11. Donnally CJ 3rd, Li DJ, Maguire JA Jr, et al. How social media, training, and demographics influence online reviews across three leading review websites for spine surgeons. *Spine J*. 2018; 18(11):2081-2090.

12. Donnally CJ 3rd, McCormick JR, Pastore MA, et al. Social media presence correlated with improved online review scores for spine surgeons. *World Neurosurg*. 2020;141:e18-e25.

13. Kalagara S, Eltorai AEM, DePasse JM, Daniels AH. Predictive factors of positive online patient ratings of spine surgeons. *Spine J*. 2019;19(1):182-185.

14. Zhang J, Omar A, Mesfin A. Online ratings of spine surgeons: Analysis of 208 surgeons. *Spine (Phila Pa 1976)*. 2018;43(12): E722-E726.

15. Melone G, Brodell J Jr, Hernandez C, et al. Online ratings of spinal deformity surgeons: Analysis of 634 surgeons. *Spine Deform*. 2020;8(1):17-24.

16. Haglin JM, Eltorai AEM, Kalagara S, et al. Patient-rated trust of spine surgeons: Influencing factors. *Global Spine J*. 2018;8(7): 728-732.

17. Donnally CJ 3rd, Roth ES, Li DJ, et al. Analysis of internet review site comments for spine surgeons: How office staff, physician likeability, and patient outcome are associated with online evaluations. *Spine*. 2018;43(24):1725-1730.

18. Donnally CJ, McCormick JR, Li DJ, et al. How do physician demographics, training, social media usage, online presence, and wait times influence online physician review scores for spine surgeons? *J Neurosurg Spine*. 2018;30(2):279-288.

19. Health Affairs. The physician payments sunshine act. 2021. https://www.healthaffairs.org/do/10.1377/hpb20141002.272302/full/. Accessed June 24, 2021.

20. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2015; Ann Arbor, MI.

21. Keenum AJ, Wallace LS, Stevens AR. Patients' attitudes regarding physical characteristics of family practice physicians. *South Med J*. 2003;96(12):1190-1194. doi: 10.1097/01.smj.0000077011.58103.c1.

22. Abghari MS, Takemoto R, Sadiq A, Karia R, Phillips D, Egol KA. Patient perceptions and preferences when choosing an orthopaedic surgeon. *Iowa Orthop J*. 2014;34:204-208.

23. Galanis C, Sanchez IS, Roostaeian J, Crisera C. Factors influencing patient interest in plastic surgery and the process of selecting a surgeon. *Aesthet Surg J*. 2013;33(4):585-590.

24. Dineen HA, Patterson JMM, Eskildsen SM, et al. Gender preferences of patients when selecting orthopaedic providers. *Iowa Orthop J*. 2019;39(1):203-210.

25. Chung JS, Merkel D, Carter CW, Kraus E, Rizzone K. Gender preferences of youth athletes for their sports medicine providers: A systematic review. *Orthop J Sports Med*. 2021;9(suppl 3). doi:10.1177/2325967121s00108.

26. Alyahya G, Almohanna H, Alyahya A, et al. Does physicians' gender have any influence on patients' choice of their treating physicians? *J Nat Sci Med*. 2019;2(1):29-34. doi:10.4103/jnsm.jnsm_28_18.

27. Kirkpatrick W, Abboudi J, Kim N, et al. An assessment of online reviews of hand surgeons. *Arch Bone Jt Surg*. 2017;5(3):139-144.

28. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: Analysis of 2185 reviews. *Am J Orthop*. 2014;43(8):359-363.

29. Velasco BT, Chien B, Kwon JY, Miller CP. Online ratings and reviews of American orthopaedic foot and ankle surgeons. *Foot Ankle Spec*. 2020;13(1):43-49.

30. Lands VW, Malige A, Nwachuku CO, Matullo KS. The effect of an orthopedic hand surgeon's attire on patient confidence and trust. *HAND*. 2019;14(5):675-683. DOI: 10.1177/1558944717750918.

31. National Research Council, Division of Behavioral and Social Sciences and Education, Committee on Human-Systems Integration, & Committee on the Role of Human Factors in Home Health Care. *The Role of Human Factors in Home Health Care: Workshop Summary*. Washington, DC: National Academies Press; 2010.

32. Zakare-Fagbamila RT, Park C, Dickson W, Cheng TZ, Gottfried ON. The true penalty of the waiting room: the role of wait time in patient satisfaction in a busy spine practice. *J Neurosurg Spine*. 2020;33(1):51-57. doi: 10.3171/2019.12.spine191257.

33. Lehrich BM, Goshtasbi K, Brown NJ, et al. Predictors of patient satisfaction in spine surgery: A systematic review. *World Neurosurg*. 2021;146:e1160-e1170.

34. Agarwal AK, Mahoney K, Lanza AL, et al. Online ratings of the patient experience: Emergency departments versus urgent care centers. *Ann Emerg Med*. 2019;73(6):631-638. doi:10.1016/j.annemergmed.2018.09.029.

35. Soroceanu A, Ching A, Abdu W, McGuire K. Relationship between preoperative expectations, satisfaction, and functional outcomes in patients undergoing lumbar and cervical spine surgery: A multicenter study. *Spine (Phila Pa 1976)*. 2012;37(2):E103-E108.

36. Mahomed NN, Liang MH, Cook EF, et al. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. *J Rheumatol*. 2002;29:1273-1279.

37. D'Souza M, Gendreau J, Austin F, Kim LH, Ho AL, Anand V. Robotic-assisted spine surgery: History, efficacy, cost, and future trends [Corrigendum]. *Robot Surg*. 2019. DOI: 10.2147/rsrr.s238276.

38. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data. *J Med Internet Res*. 2020;22(12):e21418. doi:10.2196/21418.

39. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. "W'en 'Bad''is ""od'": Identifying Personal Communication and Sentiment in Drug-Related Tweets. *JMIR Public Health Surveill*. 2016;2(2):e162. doi:10.2196/publichealth.6327.

40. Shelar A, Huang C-Y. *Sentiment Analysis of Twitter Data*; 2018. 10.1109/CSCI.2018.00251.