

# Bifurcation of Patient Reviews: An Analysis of Trends in Online Ratings

Lara L. Devgan, MD, MPH, FACS\*

Elizabeth J. Klein, BA†

Stephen Fox, PhD‡

Tugce Ozturk, PhD‡

**Background:** Online reviews have become increasingly important drivers of healthcare decisions. Data published by the Pew Research Center from 2016 suggest that 84% of adult Americans use online rating sites to search for information about health issues. The authors sought to analyze physician reviews collected from a large online consumer rating site to better understand characteristics that are associated with positive and negative review behavior.

**Methods:** Published patient reviews from RealSelf were sampled over a 12-year period (June 2006 to August 2018). SQL, Python, and Python SciPy were used for statistical analysis on 156,965 reviews of 10,376 unique physicians. Python VADER was used to quantify consumer sentiment with review text as input.

**Results:** Surgical procedures tended to be higher rated than nonsurgical treatments. The highest-rated procedures were breast augmentation, rejuvenation of the female genitalia, and facelift. The lowest-rated surgical procedures were buttock augmentation, rhinoplasty, and eyelid surgery. The mean physician rating was 4.6, with 87% of reviews being 5-star and 5% being 1-star. Sentiment analysis revealed positive consumer sentiment in 5-star reviews and negative sentiment in 1-star reviews.

**Conclusions:** These findings suggest that online reviews of doctors are polarized by extreme ratings. Within the surgical category, significant differences in ratings exist between treatments. Perceived problems with postprocedural care are most associated with negative reviews, whereas satisfaction with a physician's answers to patient questions is most associated with positive reviews. Polarization of physician reviews may suggest selection bias in reviewer participation. (*Plast Reconstr Surg Glob Open* 2020;8:e2781; doi: [10.1097/GOX.0000000000002781](https://doi.org/10.1097/GOX.0000000000002781); Published online 23 April 2020.)

## INTRODUCTION

With global internet usage at a historic peak, online reviews are increasingly driving healthcare decisions.<sup>1</sup> Web forums and social media platforms have become an extension of the “word-of-mouth” endorsement, significantly impacting patient choice of procedure and provider. Data published by the Pew Research Center in 2014 suggest that 84% of adult Americans currently use online rating sites to search for information about health issues and up

to 80% of patients trust online reviews as much as personal recommendation.<sup>2,3</sup> These statistics are particularly meaningful in the context of cosmetic plastic surgery, a field where procedures are elective and patients behave as critical consumers.

Although online reviews have democratized access to health information, yielding transparency and a more informed patient population, they are also a potential source of hazardous, misrepresented data. Because surgical websites are not subject to the strict regulations of peer-reviewed scientific publications, they can be wrought with substantive errors: information gathered on esthetic surgery websites is inaccurate or misleading 34%–89% of the time.<sup>4,5</sup> Moreover, up to 85% of plastic surgeons believe that information curated on online forums and blogs is harmful to patients.<sup>5</sup>

Authorship selection bias further complicates discussion of online physician reviews. Patients who have

From the \*Department of Plastic Surgery, Lenox Hill Hospital and Manhattan Eye, Ear and Throat Hospital, New York, N.Y.; †Department of Plastic Surgery, New York University School of Medicine, New York, N.Y.; and ‡Department of Plastic Surgery, RealSelf, Seattle, Wash.

Received for publication December 29, 2019; accepted February 25, 2020.

Copyright © 2020 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: [10.1097/GOX.0000000000002781](https://doi.org/10.1097/GOX.0000000000002781)

**Disclosure:** Dr. Devgan was the Chief Medical Officer/Chief Medical Editor for RealSelf from 2018-2020. Stephen Fox is the Director of Data Science at RealSelf. Tugce Ozturk is a Communications Data Scientist at RealSelf. Elizabeth Klein has no financial information to disclose.

extreme opinions, both positive and negative, are more likely to post unsolicited commentary on procedures and providers than consumers with more moderate attitudes.<sup>6,7</sup> Ratings may be based on nonmedical, interpersonal factors that have little to do with procedural outcome or patient safety.<sup>8</sup> Nevertheless, online physician reviews are a valuable resource for patients. With 37%–81.8% of patients visiting cosmetic surgery web platforms before their first consultation with a physician, review sites have the power to disseminate information to a broad consumer base and significantly impact healthcare choices.<sup>5</sup>

In our technologic era, one in which anyone can author and publish an opinion, it is important to determine how to best curate and verify resources for patients. Although some research has been conducted in this field, there is insufficient analysis of the factors which motivate effusive reviews of cosmetic procedures. Evaluating patterns in esthetic surgery reviews may be helpful for assessing the reliability of physician rating sites.<sup>7,9</sup> Such analysis will also offer insight into the attitudes consumers are exposed to before surgical consult so that physicians can optimize patient counseling.

The purpose of this study was to analyze physician reviews collected from a large online consumer rating site to better understand characteristics that are associated with positive and negative review behavior.

## METHODS

To study online physician ratings, we analyzed all patient reviews published on RealSelf ([www.reaself.com](http://www.reaself.com)) over a 12-year period (June 2006 to August 2018). RealSelf is an online, crowd-sourced cosmetic surgery forum, which features patient reviews and rating scales. These include a provider star rating, set on a scale of 1–5 stars, and a consumer “worth it” rating, which allows visitors to evaluate procedures as “worth it,” “not worth it,” or “not sure.”

Reviews with <70 words, without a star rating or without an associated treatment procedure, were excluded from analysis. SQL and Python (Python Software Foundation, Wilmington, Del.) were the primary means for extracting and analyzing data. Python and SciPy (Python Software Foundation, Wilmington, Del.) were used for statistical analysis. Lexicon, or a “dictionary of sentiment,” analysis of written reviews was accomplished using Python VADER (Valence Aware Dictionary and sEntiment Reasoner; Python Software Foundation, Wilmington, Del.), a computational tool designed to assess attitudes and emotions expressed in social media text. VADER is sensitive to both polarity (positive/negative) and intensity (strength) of emotions. The model labels each word in the lexicon according to its semantic orientation and assigns a polarity rating. The compound score of a sentence can then be computed by summing all lexicon ratings and normalizing them on a sentiment scale of –1 (very negative) to +1 (very positive). Compound scoring takes into account punctuation, capitalization, slang, emojis, and emoticons when assessing strength of attitude and belief. The Bonferroni statistical correction applied when several

tests were performed simultaneously. Statistical significance was set for a  $P$  value of <0.05.

## RESULTS

In our RealSelf data analysis, 285,031 patient reviews of 12,253 unique physicians were abstracted, of which 156,965 reviews of 10,376 unique physicians met inclusion criteria. We identified 122,810 reviews of surgical treatments, but only 34,155 reviews of noninvasive, injectable, or laser treatments (Table 1).

Surgical procedures tended to be higher rated than nonsurgical treatments (mean rating, 4.70 versus 4.52). Highest-rated surgical procedures were breast augmentation (mean rating, 4.81;  $n = 29,100$  reviews), rejuvenation of the female genitalia (mean rating, 4.79;  $n = 1,979$  reviews), and facelift (mean rating, 4.78;  $n = 5,591$  reviews). There was no statistically significant difference between these 3 treatment categories as tested by a pairwise  $t$  test with Bonferroni correction ( $P = 0.04 > P_{\text{adjusted}} = 0.025$ ). Lowest-rated surgical procedures were buttock augmentation (mean rating, 4.49;  $n = 8,199$  reviews), rhinoplasty (mean rating, 4.58;  $n = 10,220$  reviews), and eyelid surgery (mean rating, 4.70;  $n = 4,001$  reviews). Total number of reviews for surgical and nonsurgical procedure segments and individual procedures is represented in Table 1, and summarized percentages of provider ratings for surgical and nonsurgical procedure segments and individual procedures are presented in Table 2.

Analysis of provider star ratings yielded a mean physician rating of 4.5, with 136,377 (87%) 5-star reviews and 8,161 (5%) 1-star reviews (Fig. 1). The factors most often associated with 1-star ratings were aftercare and follow-up, whereas for 5-star ratings, it was answering patients’ questions. Sub-rating analysis showing which categories most identify with overall 1- and 5-star ratings is summarized in Figure 2. Sentiment analysis revealed positive consumer sentiment in 5-star reviews (median, 0.98; mean, 0.83) and negative sentiment in 1-star reviews (median, –0.40; mean, –0.10). The distribution of sentiment scores in these ratings is illustrated graphically in Figure 3.

Sentiment word clouds were built by taking the top 25% most positive sentiment reviews and the bottom 25% most negative sentiment reviews (Fig. 4). These word clouds illustrate which words are most predictive of positive or negative sentiment reviews, where the size of the word is indicative of how frequently it shows up in the reviews. It is important to note that there are common words on both the negative and positive word clouds, indicating that the same words and phrases can mean different things in different contexts. The word cloud of RealSelf review sentiment analysis shows that “happy, result, go back, pre op, post op, much better” are some of the phrases that are highly predictive of a positive sentiment review and the phrases “scar, pain, bad, surgery, time, doctor, skin, procedure” are predictive of a negative sentiment review.

**Table 1. Total Number of Reviews for Surgical and Nonsurgical Procedure Segments and Individual Procedures**

Procedure Segment		Procedure				
Surgical	Nonsurgical	Breast Augmentation	Cosmetic Toxins	Eyelid Surgery	Lip Augmentation	Female Genitalia Rejuvenation
122,810	34,155	29,100	5,067	4,001	1,177	1,979

**Table 2. Percentages of 1-, 2-, 3-, 4-, and 5-star Provider Reviews by Procedure**

	1-star (%)	2-star (%)	3-star (%)	4-star (%)	5-star (%)
Procedure segment					
Surgical	4.6	1.4	1.8	4.0	88.4
Nonsurgical	7.5	2.2	3.0	5.7	81.6
Procedure					
Breast augmentation	2.9	0.8	1.1	2.9	92.2
Buttock augmentation	6.2	2.5	4.5	9.4	77.3
Cosmetic toxins	5.0	1.2	1.2	2.2	90.3
Eyelid surgery	5.2	1.3	1.4	2.2	89.9
Facelift	4.0	1.0	0.9	1.1	93.1
Hair transplant	4.8	1.1	1.1	3.7	89.3
Lip augmentation	5.0	1.2	1.8	2.2	89.8
Rhinoplasty	7.4	1.8	1.8	3.1	85.9
Female genitalia rejuvenation	3.1	1.1	1.1	2.9	91.8

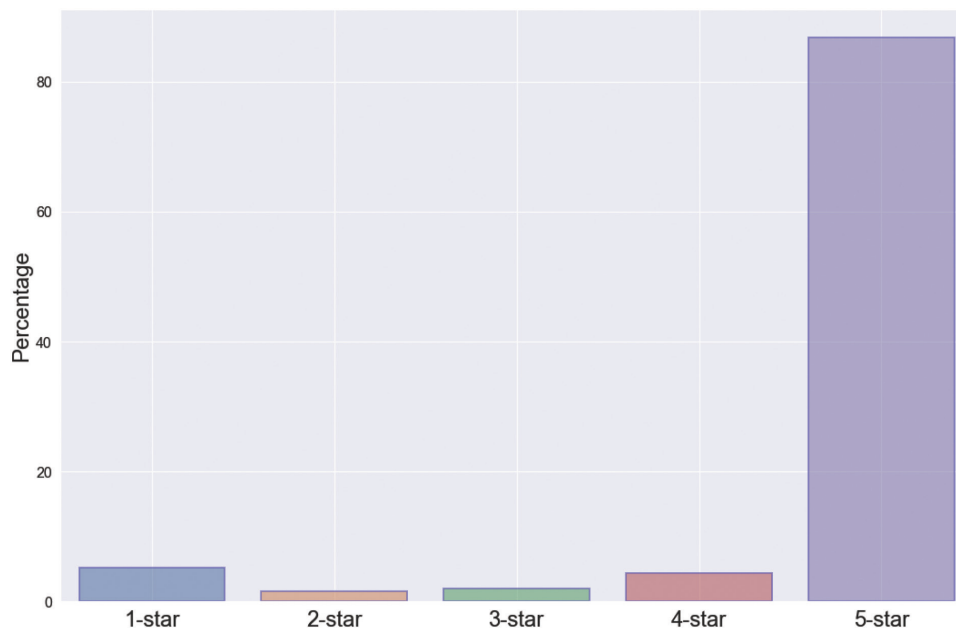
**DISCUSSION**

Online reviews are increasingly driving consumer behavior, including that of patients seeking cosmetic surgery. Although only 19% of Americans sought out online health information in 2001, that value increased to nearly 60% by 2010. Moreover, up to 59% of patients describe online reviews as at least “somewhat important” in choosing a healthcare provider.<sup>10</sup> Given these data, our analysis was designed to better understand the web information patients are interacting with and to identify the factors associated with patient satisfaction.

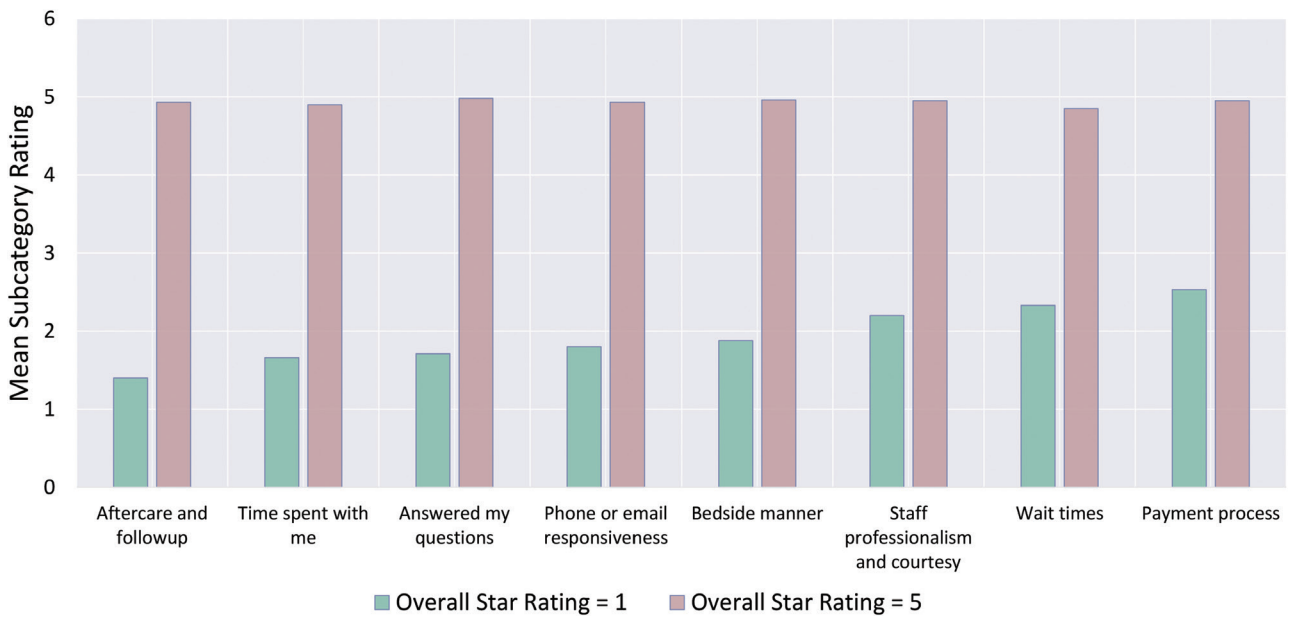
In our study, we found online reviews of physicians to be polarized. Extreme opinions on both ends of the

spectrum were represented with far greater frequency than moderate ones. This was evident in our lexicon analysis and in the distribution of star ratings we observed. Although mean physician rating of RealSelf doctors was 4.5, 87% were 5-star reviews, and 5% were 1-star reviews. Five-star reviews were associated with positive consumer sentiment (median, 0.98; mean, 0.83), and 1-star reviews were associated with negative sentiment (median, -0.40; mean, -0.10).

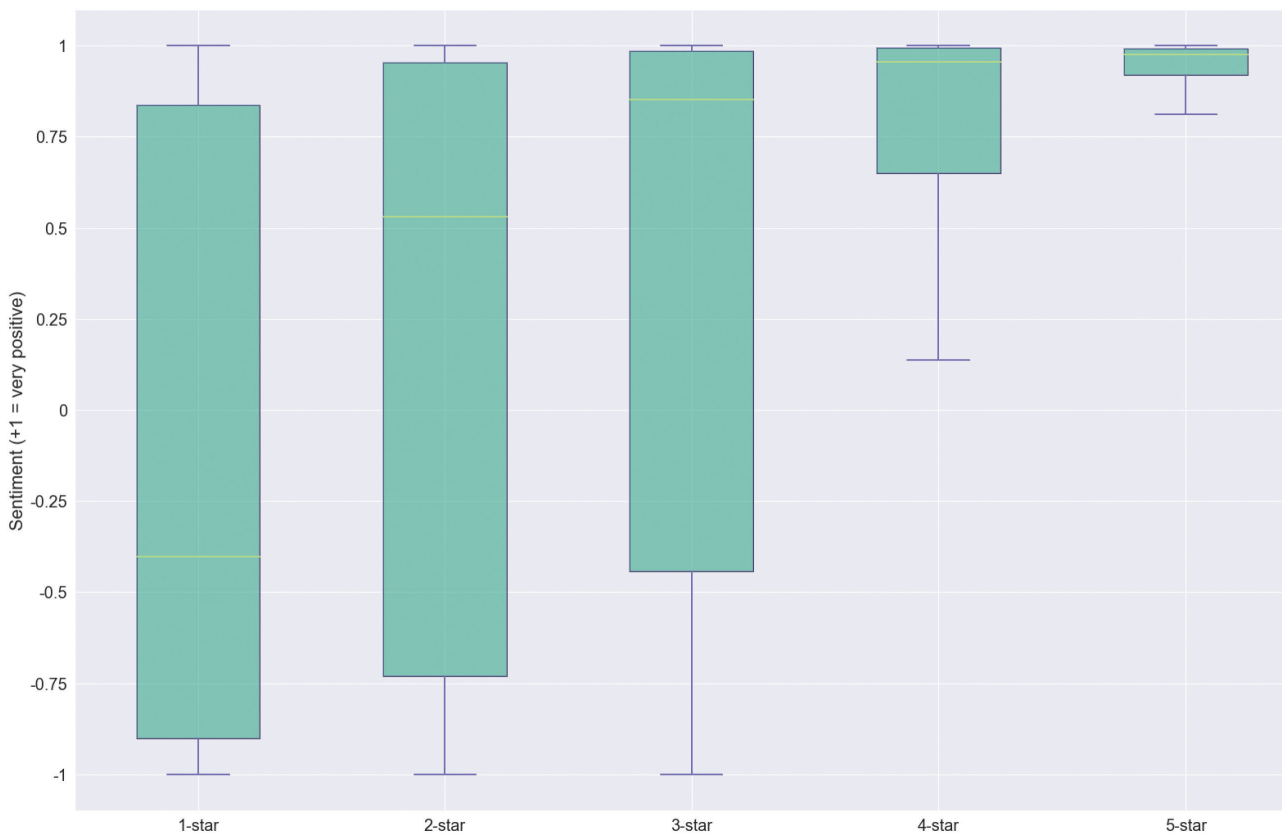
These findings suggest selection bias in the patients rating procedures and providers. They also corroborate results published in other studies. The analysis of online reviews for breast augmentation by Dorfman et al<sup>7</sup> noted a similar bimodal distribution in procedure ratings. Its authors



**Fig. 1.** Percentage of 1-, 2-, 3-, 4-, and 5-star provider reviews for all procedures. Extreme bifurcation in reviews was observed, with 87% of reviews being 5-star reviews and 5% of reviews being 1-star reviews.



**Fig. 2.** Sub-rating analysis showing which categories identify with overall 1- and 5-star ratings. The factors most often associated with 1-star ratings were aftercare and follow-up, whereas for 5-star ratings, it was answering patient questions.



**Fig. 3.** Variations in sentiment score distribution for 1-, 2-, 3-, 4-, and 5-star provider reviews. Sentiment analysis revealed an association between positive consumer sentiment and 5-star reviews, and an association between negative consumer sentiment and 1-star reviews.

determined that “strong feelings elicit strong reactions, [which] convert to voluntary acts of review online.”<sup>7</sup> Other researchers have proposed similar explanations of reviewer behavior.<sup>6,9</sup> Khansa et al<sup>6</sup> described higher word counts in

negative online reviews than positive ones and purported that patients are more motivated to share their experience when dissatisfied with results. More broadly, the number of reviews published per physician on a given site is a small



## Words Associated with Positive Reviews



## Words Associated with Negative Reviews



**Fig. 4.** Sentiment word cloud showing the phrases associated with positive and negative reviews. The size of the word is indicative of how frequently it shows up in reviews.

fraction of the total number of patients treated by those same physicians. Opinions expressed in online reviews cannot be generalized to the patient population at large.

Bifurcation in cosmetic surgery reviews has other important implications. Extremes in review behavior may generate increased web traffic and more profoundly impact purchase behavior than uniformly positive ratings. Northwestern University's Spiegel Research Center found that 82% of consumers seek out negative reviews when gathering information about a product. Consumers proceed to spend 4× as long on review sites where they can interact with negative ratings. In addition, purchase likelihood peaks for products with ratings in the range of 4.0–4.7 and then decreases as ratings approach 5.0.<sup>9</sup> As previously noted, the average rating of RealSelf physicians in this study was 4.5, precisely within this effective purchase conversion range. In moderation, the presence of negative reviews does not deter consumers, but instead attracts them, establishing review site credibility. This is important to consider when discussing the impact of bifurcated reviews on patient decision-making.

Various factors contribute to the relative proportion of positive and negative reviews published for a given procedure or provider. We observed that surgical procedures tend to be higher rated than nonsurgical treatments (mean rating, 4.70 versus 4.52). The highest-rated surgical procedures were breast augmentation, rejuvenation of the female genitalia, and facelift. The

lowest-rated surgical procedures were buttock augmentation, rhinoplasty, and eyelid surgery. Surprisingly, the procedures which received the highest and lowest ratings were different than those which RealSelf consumers deemed most and least “worth it” in data published by Domanski and Cavale.<sup>11</sup> The procedures we identified also differed from the cosmetic interventions in the medical literature that have the highest and lowest satisfaction scores.<sup>11</sup>

These findings suggest that high and low procedure ratings may have little to do with the procedure itself—at least as far as complications and outcomes are concerned. We identified that the factors most associated with 1-star ratings were aftercare and follow-up, whereas the factors most associated with 5-star ratings were answering patients' questions. Similar studies of Google and Yelp reviews have identified entirely different associated factors.<sup>7</sup> The wide range of items identified in the literature as motivating positive and negative review behavior indicates the need for further research on this subject.

Despite these inconsistencies, online reviews could have utility in guiding patients toward satisfying health-care decisions.<sup>12</sup> The factor most associated with 5-star ratings in our study, answering patients' questions, was also the most highly correlated factor with postprocedural satisfaction in a nationwide survey of plastic surgery patients. From the surgeon's perspective, this finding reiterates the importance of empathy and communication in treating patients. It also highlights a strategy for maintaining a positive online presence.

There are several limitations to this study. RealSelf is a single site and thus reflects only the attitudes and behaviors of individuals who choose to visit it. It is difficult to generalize our findings beyond RealSelf visitors because we have not included data from the websites of the other large corporate players.

In addition, we must take into account that online reviewers represent a small percentage of total plastic surgery patients and that RealSelf does not collect demographic data on its users. We cannot be certain that online reviews reflect the real-life preferences of cosmetic surgery patients more globally or know the degree to which these reviews influence in-person consumer behavior.

Finally, we are unable to conclude from our analysis why surgical procedures are higher rated than nonsurgical treatments or why procedures that receive the highest and lowest star ratings are different from those which RealSelf consumers deem most and least “worth it.” Further research is needed to determine the factors that wield the greatest influence on positive and negative review behavior.

## CONCLUSIONS

Online reviews of cosmetic surgeons are polarized by dichotomous ratings, with extremes in reviewer behavior suggesting authorship selection bias. Polarization of ratings may impact the degree to which reviews influence consumer follow through.

Surgical procedures tend to be higher rated than nonsurgical procedures. Within the surgical category, significant differences in ratings exist between treatments. Perceived issues with postprocedural care are most associated with negative reviews, whereas satisfaction with a physician's answer to patient's questions is most associated with positive reviews.

Additional studies must be conducted to understand the motivators of positive and negative review behavior and to illuminate the ways plastic surgeons can harness this information to improve their practice. We believe that effectively utilizing digital data is critical to understand the patients' perspective and experience.

**Lara L. Devgan, MD, MPH, FACS**

Department of Plastic Surgery  
Lenox Hill Hospital and Manhattan Eye, Ear and Throat  
Hospital  
969 Park Avenue  
New York, NY 10028  
E-mail: [lara@laradevganmd.com](mailto:lara@laradevganmd.com)

#### REFERENCES

1. Janik PE, Charytonowicz M, Szczyt M, et al. Internet and social media as a source of information about plastic surgery: comparison between public and private sector, a 2-center study. *Plast Reconstr Surg Glob Open*. 2019;7:e2127.
2. Fox S; Pew Research Center. The social life of health information. Available at <https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/>. Accessed April 7, 2020.
3. Qiu CS, Hockney SM, Turin SY, et al. A quantitative analysis of online plastic surgeon reviews for abdominoplasty. *Plast Reconstr Surg*. 2019;143:734–742.
4. Szycha P, Zieliński T, Rykała J, et al. Role of the internet in communication between patient and surgeon before rhinoplasty. *J Plast Surg Hand Surg*. 2012;46:248–251.
5. Montemurro P, Porcnik A, Hedén P, et al. The influence of social media and easily accessible online information on the aesthetic plastic surgery practice: literature review and our own experience. *Aesthetic Plast Surg*. 2015;39:270–277.
6. Khansa I, Khansa L, Pearson GD. Patient satisfaction after rhinoplasty: a social media analysis. *Aesthet Surg J*. 2016;36:NP1–NP5.
7. Dorfman RG, Purnell C, Qiu C, et al. Happy and unhappy patients: a quantitative analysis of online plastic surgeon reviews for breast augmentation. *Plast Reconstr Surg*. 2018;141:663e–673e.
8. Strech D. Ethical principles for physician rating sites. *J Med Internet Res*. 2011;13:e113.
9. Maslowska E, Malthouse EC, Bernritter SF. Too good to be true: The role of online reviews' features in probability to buy. *Int J Advert*. 2017;36:142–163.
10. Hanauer DA, Zheng K, Singer DC, et al. Public awareness, perception, and use of online physician rating sites. *JAMA*. 2014;311:734–735.
11. Domanski MC, Cavale N. Self-reported “worth it” rating of aesthetic surgery in social media. *Aesthetic Plast Surg*. 2012;36:1292–1295.
12. Chen K, Congiusta S, Nash IS, et al. Factors influencing patient satisfaction in plastic surgery: a nationwide analysis. *Plast Reconstr Surg*. 2018;142:820–825.