Understanding Your Online Ratings: A Methodological Analysis Using Urogynecologists in the United States

Sarah A. Huber, MD,† Jennifer Priestley, PhD,† Khushabu Kasabwala, MD,* Bogdan Gadidov, MS,† and Patrick Culligan, MD**

Objective: The influence of online physician ratings is growing, yet their interpretation remains difficult. Our aim was to analyze the online content of urogynecologists on 1 website to transform these ratings into practical tools for care improvement.

Methods: This cross-sectional analysis studied the ratings and reviews of every board-certified urogynecologist listed on Healthgrades.com. The ratio of 5:1 ratings was compared between various physician characteristics and practice qualities. Four investigators classified narrative reviews into one or more of the following themes: about the (1) physician, (2) clinical outcomes, (3) and/or staff. The content of the narrative reviews was analyzed, and word clouds were created to understand the primary motivators behind ratings.

Results: In February 2018, the Healthgrades pages for 689 urogynecologists were evaluated, and 523 physicians were included in the study. Higher 5:1 ratios were found among men versus women (4.0 vs 3.0; $P < 0.01$), and OB-GYN–trained versus urology-trained (4.0 vs 2.2; $P < 0.01$) physicians. A benchmarking rubric was developed to illustrate the 5:1 ratio distribution for all physicians stratified by number of ratings. A total of 3300 narrative reviews were assigned themes with strong inter- and intrarater reliability (Table 3). Physician qualities most influenced extreme scores (1 or 5 stars), whereas average reviews were more influenced by staff. Commonly discussed physician qualities included professionalism, time with patient, and counseling.

Conclusions: Using the 5:1 rating ratio and simplified review themes as tools, physicians can understand what their ratings signify both as an indicator of their online reputation compared with their peers and as a means for improving the patient experience.

Key Words: social media, ratings, reviews, quality improvement

(*Female Pelvic Med Reconstr Surg* 2019;25: 193–197)

Online physician review platforms are visited by countless patients every day, with at least 50% of patients crediting a website in their selection of a provider.[1–3] In a recent large survey of US adults, 59% agreed that physician rating sites were important in their physician choice.[4] One of the most popular websites, Healthgrades.com, attracts millions of viewers a month and contains over 7 million public reviews.[5]

Despite their popularity, the true impact of these ratings, specifically their relevance in clinical care and practice management, remains unknown. Until now, analyses have focused on the relationship between ratings and either clinical quality of care metrics or physician demographics.[6–8] In general, results have been mixed, applicable conclusions are limited, and the use of these reviews remains a mystery.

In reality, online ratings and narrative reviews reflect the complete patient experience, highlighting the patient's *perception* of the quality of care along with the facilitative environment and personnel interactions. In other words, to reduce these ratings to surrogates for clinical outcomes would be to ignore the complexity of the patient experience. There is a clear knowledge gap in how best to independently interpret physician ratings as an authentic measure of the patient experience and how to use them as the main metric to assess one's online reputation.

We propose a new method to understand online ratings, focusing on both the factors that generate a single review and the significance of all reviews as a whole. Our goals for this study were 3-fold: to assess the relationship between ratings and physician qualities, to establish a mechanism by which physicians can benchmark their ratings compared with their peers, and finally, to identify the aspects of the patient experience that strongly impact ratings using narrative text reviews as a guide.

MATERIALS AND METHODS

This was a descriptive cross-sectional study analyzing online provider ratings and accompanying text reviews on the Healthgrades.com website. Healthgrades is the most commonly visited physician rating website, welcoming nearly 40% of the total web traffic of online physician ratings platforms.[9] In February 2018, a physician search was performed on the Healthgrades website, filtered by "Female Pelvic Medicine and Reconstructive Surgery" (FPMRS) board certification status. Each physician identified was cross-referenced with the American Urogynecologic Society's (AUGS) membership directory (www.augs.org). The publicly available ratings and reviews for each identified physician were extracted from the website for analysis. We received authorization from Healthgrades Inc to report our findings through publication.

Our methodological plan took the following steps. We first examined the relationship between physician ratings and certain physician demographic factors such as age, sex, years in practice, general specialty (urology or obstetrics/gynecology), and fellowship training, all of which were obtained from Healthgrades and confirmed through AUGS. Geographic region and community type were determined by the address of the practice as listed on Healthgrades. Physicians were placed in one of 5 US geographic regions. Those who were located within one of the 20 most populous US metropolitan areas were classified as working in an urban community. Practice type was based on information provided by the physicians in the AUGS membership directory. Finally, physicians listed as a Castle Connolly Top Doctor were compared with those not listed. Castle Connolly is a private consumer research firm that compiles an annual list of highly ranked US physicians determined by a variety of factors including board certification, malpractice history, appointments, and peer nomination. Of note, it is not influenced by any financial incentives from motivated providers and is not a paid marketing tool. We were interested in how the patient perception of top quality physician care compares to how physicians view their peers represented by Top Doctor status. Comparative analysis via Kruskal-Wallis and Mann-Whitney testing was performed with SAS v9.4 (Cary, North Carolina).

We began by assessing whether differences between groups are best depicted in terms of mean ratings, median ratings, or the

From the *Weill Cornell Medicine, New York, NY; and †Kennesaw State University, Kennesaw, GA.

Correspondence: Patrick Culligan, Center for Female Pelvic Health, Department of Urology, Weill Cornell Medicine, 525 E 68th Street, F-9 West, New York, NY 10065. E-mail: pjc9007@med.cornell.edu.

The authors have declared they have no conflicts of interest.

DOI: 10.1097/SPV.0000000000000676

ratio of 5:1 ratings. Prior studies have confirmed that the majority of online ratings tend to rest at the extremes, with most values being either 1 or 5.[8–10] As expected, over 75% of physician ratings in our data set were 5 stars and 17% were 1 star. With nonparametric testing, the ratio approach created a wide distribution and better separation between "good" and "bad" performance ratings compared with median or mean ratings. Based on these findings, we opted to use the 5:1 ratio as our metric for comparative testing and for the creation of a self-assessment rating rubric.

The self-assessment rubric serves as an easily interpretable tool that physicians could use to benchmark themselves in reference to their peers. The rubric was stratified by the total number of ratings for a given physician, as we anticipated the 5:1 ratio values would be affected by the number of ratings per physician. The ratio values for physicians in the 25th, 50th, 75th, and 95th percentiles for each stratified group were calculated and entered in the rubric, so that a given physician would be able to quickly assess their standing within the group.

The narrative text reviews of each physician were extracted from Healthgrades along with their corresponding star rating of 1 to 5. To easily understand the motivations generating the content of these text reviews, we established a simplified codebook with reproducible themes. A qualitative assessment of the reviews highlighted the following 3 themes for the reviews: provider related, staff/practice related, and outcomes related. Reviews that used vague terms to describe one's overall care experience without going into any detail were not included in any subsequent analysis. Four independent investigators, composed of 2 urogynecology providers and 2 lay persons, assigned themes to each review. If a review contained remarks about more than 1 theme, it was assigned to each relevant theme group. The frequency of the theme groups was calculated within each numerical rating cohort, 1 to 5, to better understand what factors into a high or low rating. Interrater reliability of the thematic categorizations was assessed using percent agreement and Fleiss κ.

Intrarater testing of the thematic analysis was performed 3 months later. An online randomizer selected 100 of the original reviews, which were reassigned to theme groups by one of the original investigators who was blinded to her prior assignments. The new theme assignments were compared with the original assignments, using Cohen κ.

TABLE 1. Ratio of 5:1 Ratings by Physician Demographics and Practice Features

Physician Demographic, n (%)*	Total N = 523	Ratio of 5:1 Ratings Median (IQR)	*P*
Sex			
Female	220 (42.1)	3.0 (1.5–5.4)	<0.001
Male	303 (57.9)	4.0 (2.0–7.0)	
Residency training			
OB-GYN	361 (69.0)	4.0 (2.0–7.0)	<0.001
Urology	162 (31.0)	2.2 (1.1–4.0)	
Fellowship training†			
No	363 (69.4)	3.2 (1.8–6.7)	0.43
Yes	160 (30.6)	4.0 (2.0–6.9)	
Years in practice‡			
<10	45 (9.1)	4.5 (2.1–8.0)	0.22
10–19	212 (43.0)	3.0 (1.7–6.6)	
≥20	236 (47.9)	3.5 (1.9–6.7)	
Region of practice location			
Northeast	131 (25.0)	3.7 (1.9–6.1)	0.58
Southeast	147 (28.1)	6.7 (1.9–6.6)	
Midwest	108 (20.7)	3.3 (1.5–7.0)	
Southwest	55 (10.5)	3.3 (2.0–6.8)	
West	82 (15.7)	3.0 (1.7–5.0)	
Top 20 metropolitan area			
No	308 (58.9)	3.1 (1.7–6.0)	0.21
Yes	215 (41.1)	4.0 (2.0–8.2)	
Practice affiliation§			
Academic	71 (30.6)	3.3 (1.4–6.0)	0.005
Teaching community	97 (41.8)	5.0 (2.7–7.7)	
Nonteaching community	64 (27.6)	4.2 (2.0–8.7)	
Castle Connolly Top Doctor‖			
No	440 (84.1)	3.1 (1.7–6.7)	0.06
Yes	83 (15.9)	4.5 (2.4–7.2)	

Kruskal-Wallis independent samples testing.

*Physicians with no 1-star ratings were not included as no ratio calculation existed. No physicians had 1-star ratings without any 5-star ratings.

†Based on reporting through Healthgrades.com. Accreditation Council for Graduate Medical Education accredited FPMRS/Female Urology fellowships only.

‡Practice history was missing in 37 physician profiles on Healthgrades.com

§Based on review of AUGS Membership Directory (accessed February 2017). Information was unavailable for 291 physicians.

‖Status obtained through review of Castle Connolly public websites (accessed February 2017).

TABLE 2. Ratios of 5:1 Ratings Among Urogynecologists Grouped by Total Number of Ratings

No. Ratings	25th Percentile	50th Percentile	75th Percentile	95th Percentile
Total (N = 523)	1.8	3.5	6.8	16.0
<5 (n = 20)	2.0	2.0	3.0	3.0
5–10 (n = 109)	1.3	3.0	5.0	7.0
10–20 (n = 201)	1.6	3.3	6.0	13.0
20–50 (n = 163)	2.0	4.0	8.0	20.0
>50 (n = 30)	6.1	10.6	23.8	76.0

Values correspond to the median value of the 5:1 ratios within each group. Physicians with all 5-ratings or all 1-ratings were excluded.

To provide some clarification of the various motivators under each theme, we created word clouds for each theme group in positive-rating groups (4 or 5 stars) and for negative-rating groups (1 or 2 stars) based on word frequency. The frequency counts were filtered using our codebook, which was created from identified codes and categories assessed during the earlier thematic analysis. Thematic analysis and word clouds were performed with ATLAS.ti v8 (Berlin, Germany).

RESULTS

In February 2018, all numeric ratings and 3300 free text narrative reviews for every board-certified urogynecologist were extracted from Healthgrades.com. Of the total 689 urogynecology physicians listed on Healthgrades, 36 had no ratings and were therefore excluded. In addition, another 130 physicians with only 5-star ratings and therefore no 5:1 ratio were excluded, leaving 523 physicians for analysis. The total number of urogynecologists listed in Healthgrades was compared with the number listed in the AUGS directory, which confirmed adequate representation. Table 1 illustrates the demographic distribution of the included providers. The average physician age was 50.5 years (range, 37–83 years), with the majority completing a residency in OB-GYN (69%) versus urology (31%). There were 94% of physicians who completed a FPMRS fellowship that were board-certified OB-GYNs. Over 90% of included physicians were in practice for at least 10 years. The geographical and community-type distribution of providers showed a balanced coverage of the United States.

Interestingly, the excluded 130 physicians with only 5-star ratings tended to be younger (48.8 vs 51.0 years; $P < 0.001$) and had less experience (17.9 vs 19.9 years; $P = 0.02$). As would be expected, they also had fewer total ratings (10.4 vs 21.0; $P < 0.001$). There were no significant differences between this group and the included group in any of the other physician or practice characteristics.

The ratio of 5:1 ratings, relative to each demographic and practice feature variable, is depicted in Table 1. There was a significant difference in the median 5:1 ratio between male and female providers (4.0 vs 3.0 ratio; $P < 0.001$) and between OB-GYN–trained and urology-trained providers (4.0 vs 2.2; $P < 0.001$). Practice affiliation also had a notable difference, with teaching community practice providers having the highest median ratio (5.0), compared with nonteaching community practice and academic practice providers at 4.2 and 3.3, respectively ($P = 0.005$).

Table 2 depicts the online rating self-assessment rubric by which physicians can determine their rating status compared with their peers. Of note, the differences in the 5:1 ratios by percentile increase exponentially as the number of ratings increases. Whereas the 5:1 ratio for physicians with 5 to 10 ratings in the 95th percentile is 5 times greater than those in the 25th percentile, it is 12 times greater between similar percentiles for physicians with more than 50 ratings.

The 3300 narrative text reviews underwent thematic analysis as mentioned previously. Theme assignments were made according to statements pertaining to the physician, staff/practice, and clinical outcomes. The interrater reliability using these themes to categorize the reviews was high as shown in Table 3. The intrarater reliability was also high, showing consistency in thematic assignment over time.

Figure 1 depicts the distribution of the thematic categories for each corresponding rating. As expected, the patient's perception of the physician had the greatest influence on very positive (5) or very negative (1) ratings. Although less so for the "clinical outcomes" theme, outcomes-related commentary did appear consistently in the text of very positive or very negative reviews. Ratings in the middle range (2, 3, and 4) were more greatly influenced by staff and practice qualities.

The words clouds of positive (rating 4 or 5 stars) and negative (rating 1, 2, or 3 stars) reviews are illustrated in Figure 2. Of note, the qualities of spending time with the patient, listening to them, and addressing their questions play a prominent role in physician-related reviews. Surgical results, pain, and infections tended to dominate the outcomes-related negative reviews. Staff professionalism, ease with making appointments and billing, and wait time were common topics in the staff/practice-related reviews.

DISCUSSION

Our 3-step approach illustrated in this study generated a comprehensive appraisal of online ratings that considered their statistical features along with the underlying organic motivators, thus transforming online ratings into an interpretable tool for physician assessment. Calculation of one's 5:1 ratio along with in-depth understanding of the factors behind ratings allows physicians to easily assess their performance and look for ways to improve their practice.

Among the physicians included in this study, interesting differences in ratings were noted between male versus female and OB-GYN versus urology providers. In general, male providers tended to have a higher ratio of very positive to very negative

TABLE 3. Reliability Testing of Thematic Categorization Scheme for Online Reviews

	Doctor	Results	Staff
Interrater reliability (κ)*	0.656	0.789	0.809
Intrarater reliability (κ)†	0.694	0.859	0.917

*Fleiss κ coefficient calculated from analysis by 4 independent investigators using 3300 online reviews.

†Cohen κ coefficient calculated from a 3-month test-retest analysis by 1 investigator using a subset of 100 online reviews.

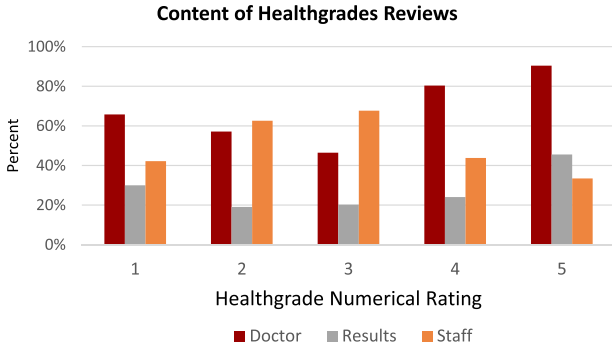

FIGURE 1. Frequency of each review theme by numerical rating.

reviews compared with their female colleagues. The same was true for OB-GYNs compared with urologists. Although it is not in the scope of this study to explain why these differences exist, it may be of importance in the consideration for future studies within the same cohort or for a different specialty.

Another such future study consideration should include creating similar self-assessment rubrics for other specialties and subspecialties using our approach as a template. In an earlier published study, we identified differences in rating distributions based on medical specialty and, in particular, between surgical and nonsurgical specialties.[8] Comparing rubrics across multiple specialties can further highlight those interspecialty differences and can demonstrate the degree of variability within a specialty group.

The basic themes identified in the narrative review analysis correspond to a recent purely qualitative study by Asanad et al,[11] which identified similar motivational themes. Their conclusion and earlier studies also support our findings that the patient-physician interaction has a stronger influence on a positive or negative review than other aspects of care.[12–14] For providers who hope to "boost" their online ratings, addressing any shortcomings in this area should be the priority. This would include increasing the perceived amount of direct patient interaction, thorough counseling, and personalized follow-up.

The strengths of our study include the large cohort of 689 urogynecologists, which is to-date one of the largest and most comprehensive online ratings studies for a single specialty, providing a diverse physician sample. This makes our results more resistant to the statistical weaknesses inherent in nonnormally distributed data and to the biases inherent in online reviews. Confirmation of physician characteristics using the AUGS member directory also added validity to our findings.

For the qualitative analyses, every narrative review of a urogynecologist was read and included in the analysis, creating a large and robust qualitative data set. The involvement of non–medically trained researchers during the coding and thematic analysis ensured that the thematic findings were reflective of both a provider's and a patient's interpretation of the reviews.

One potential limitation is the partial reliance on physician-driven profile updates on both the Healthgrades and AUGS platforms. Therefore, certain demographic information and board certification status may not be current. Attempts were made to ensure physician information was up to date.

Although the self-assessment rubric is informative for physicians with a large number of ratings, it was less useful for those with only a few ratings. As would be expected, the variability in the 5:1 rating ratio for physicians with more ratings tended to be greater than in those with 5 or fewer ratings. Therefore, we would advise physicians with less than 5 ratings to wait until they have more ratings before assessing their status.

Regarding the narrative reviews, we cannot exclude the possibility that the current reviews on Healthgrades had been altered or were subjected to external motives, termed "fake" reviews. In a large-scale study of reviews on Yelp, roughly 20% of reviews were proven to be fake in 2013.[10] However, Healthgrades internally blocks fake reviews by preventing multiple reviews to be submitted from one e-mail address and requiring a 2-step verification system for new users to prevent fake accounts. Furthermore, our use of the 5:1 ratio variable eliminates physicians with exclusively 5-star ratings, which may be suggestive of manipulation. Influences outside of the sphere of care, such as nonmedical interactions with physicians or staff, prior medical experiences, or unrelated events surrounding a visit, may also inappropriately impact reviews.

Last, as with any optional feedback format, particularly for an online public website, the degree of nonresponse bias will alter the data. Only patients with internet access who are social media savvy and aware of these websites will contribute a rating, and of those, typically only patients with a strong motivation will leave both a rating and a review. In our study, of the 11,000 ratings, only 30% wrote a corresponding narrative review.

The findings in this study highlight the value of online ratings as a tool for patient recruitment. In our office, we have put

FIGURE 2. Word clouds of coded terms in positive and negative reviews of physicians, outcomes, and staff.

into practice a new policy that involves directly asking our patients to leave feedback with a rating plus review after their visit with us. We also provide them with a handout that includes a list of physician rating websites with a QR code that links directly to our profile pages. We explain to them that their contribution not only helps other consumers select new providers but also helps us to improve as a practice and provides necessary feedback to optimize our care.

Embracing online physician ratings and the motivators behind them enhances our understanding of how we perform as physicians. Future quality-based studies should treat these as a new measurable variable of the patient experience. In our study, we performed a comprehensive analysis of the ratings and reviews of urogynecologists in one of the most popular physician rating platforms, Healthgrades.com. We hope that our methodology is used in future studies from other specialties to further validate its use. Furthermore, we hope our approach can be used in physician quality improvement projects as a new measure of the patient experience that complements traditional clinical outcome measures.

ACKNOWLEDGMENT

This study was supported by The Frederick J. and Theresa Dow Wallace Fund of the New York Community Trust.

REFERENCES

1. Holliday AM, Kachalia A, Meyer GS, et al. Physician and patient views on public physician rating websites: a cross-sectional study. *J Gen Intern Med* 2017;32(6):626–631. doi:10.1007/s11606-017-3982-5.
2. Emmert M, Meier F, Pisch F, et al. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. *J Med Internet Res* 2013;15(8):e187. doi:10.2196/jmir.2702.
3. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. *BMC Health Serv Res* 2015;15:416. doi:10.1186/s12913-015-1099-2.
4. Hanauer DA, Zheng K, Singer DC, et al. Public awareness, perception, and use of online physician rating sites. *JAMA* 2014;311(7):734–735. doi:10.1001/jama.2013.283194.
5. Harris Institute. *Healthgrades: Patient Sentiment Report.* Healthgrades Operating Company Inc. 2018. Available at: https://www.healthgrades.com/content/patient-sentiment-report. Accessed July 22, 2018.
6. Okike K, Peter-Bibb TK, Xie KC, et al. Association between physician online rating and quality of care. *J Med Internet Res* 2016;18(12):e324. doi:10.2196/jmir.6612.
7. Trehan SK, DeFrancesco CJ, Nguyen JT, et al. Online patient ratings of hand surgeons. *J Hand Surg Am* 2016;41(1):98–103. doi:10.1016/j.jhsa.2015.10.006.
8. McGrath RJ, Priestley JL, Zhou Y, et al. The validity of online patient ratings of physicians: analysis of physician peer reviews and patient ratings. *Interact J Med Res* 2018;7(1):e8. doi:10.2196/ijmr.9350.
9. Kadry B, Chu LF, Kadry B, et al. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *J Med Internet Res* 2011;13(4):e95. doi:10.2196/jmir.1960.
10. Luca M, Zervas G. Fake it till you make it: reputation, competition, and yelp review fraud. *Manage Sci* 2016;62(12):3412–3427. doi:10.1287/mnsc.2015.2304.
11. Asanad K, Parameshwar PS, Houman J, et al. Online physician reviews in female pelvic medicine and reconstructive surgery: what do patients really want? *Female Pelvic Med Reconstr Surg* 2018;24(2):109–114. doi:10.1097/SPV.0000000000000503.
12. Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. *Soc Sci Med* 2001;52(4):609–620. doi:10.1016/S0277-9536(00)00164-7.
13. Attree M. Patients' and relatives' experiences and perspectives of "good" and "not so good" quality care. *J Adv Nurs* 2001;33(4):456–466.
14. Smith AL, Nissim HA, Le TX, et al. Misconceptions and miscommunication among aging women with overactive bladder symptoms. *Urology* 2011;77(1):55–59. doi:10.1016/j.urology.2010.07.460.