Kennesaw State University

Analytics and Data Science Institute

datascience.kennesaw.edu

Data Science Research Series

Understanding Online Physician Ratings

A collaborative study with the Weill School of Medicine at Cornell University, Department of Urology

The rise of social media and near-universal online connectivity has transformed the patient-physician relationship. 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. In a recent survey of over two thousand U.S. adults, 59% agreed that physician rating sites were important in their physician choice.

Up to this point, most analyses have focused on how best to transform ratings into a metric for care quality and provider proficiency. In general, applicable conclusions have been limited, and the utility of these reviews continues to be a mystery. In this study, we propose a new method to understand online ratings, focusing on both the factors that generate a single review and the meaning of all reviews as a whole. Our goals for this study were threefold: to assess the relationship between ratings and certain physician demographics and practice patterns, to establish a mechanism by which physicians can benchmark their ratings compared to their peers, and finally, to identify the aspects of the patient experience that strongly impact ratings using narrative text reviews as a guide. By evaluating the meaning of these ratings at each level, across a physician group, for the individual physician, and on a granular level with each rating and narrative review, providers can formulate a more holistic depiction of their individual performance and understand how it compares to their colleagues.

In February 2018, all numeric ratings and 3300 free text narrative reviews for every board-certified FPMRS (Female Pelvic Medicine and Reconstructive Surgery) specialist were extracted from Healthgrades.com (n=523). Prior

studies have confirmed that the overwhelming majority of ratings tend to rest at the extremes, with most values being either 1 or 5. As expected, over 75% of physician ratings in our dataset were 5s, while slightly over 17% were 1s, with approximately 7% of the ratings as 2, 3 or 4. Using nonparametric testing, we analyzed the distribution of the mean rating, median rating, and ratio of 5:1 ratings to determine which measure was the most statistically robust and clinically interpretable. The ratio approach created a wide range and better separation between "good" and "bad" 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 purpose of a self-assessment rubric was to create an easily interpretable table that physicians could use to benchmark themselves among 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 percentile 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.

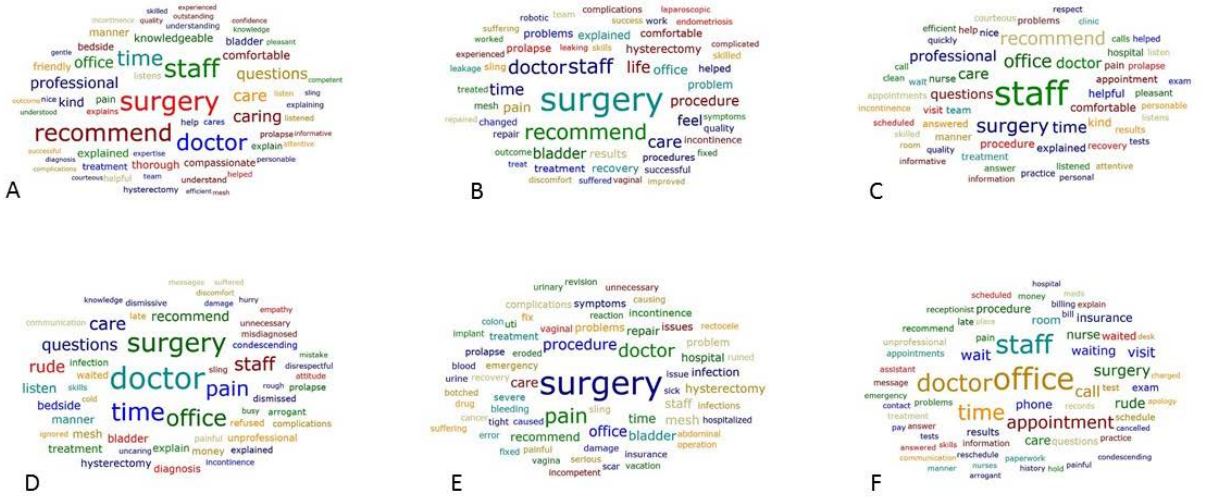
Ratios of 5:1 ratings* among urogynecologists grouped by total number of ratings

Number of ratings	25th percentile	50th percentile	75th percentile	95th percentile
Total (n=523)	1.8	3.5	6.8	16.0
Less than 5 (n=20)	2.0	2.0	3.0	3.0
5 to 10 (n=109)	1.3	3.0	5.0	7.0
10 to 20 (n=201)	1.6	3.3	6.0	13.0
20 to 50 (n=163)	2.0	4.0	8.0	20.0
More than 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.

The 3300 narrative text reviews underwent thematic analysis according to statements pertaining to the physician, staff/practice, clinical outcomes.

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



A: Doctor-oriented reviews with 4 or 5 rating, B: Outcomes-oriented reviews with 4 or 5 rating, C: Staff-oriented reviews with 4 or 5 rating, D: Doctor-oriented reviews with 1, 2, or 3 rating, E: Outcomes-oriented reviews with 1, 2, or 3 rating, F: Staff-oriented reviews with 1, 2, or 3 rating.

Spending time with the patient, listening to them and addressing their questions in physician-related reviews tented to dominate the doctor-oriented reviews. Surgical outcomes, pain, and infections tended to dominate the outcomes-related negative reviews. Staff professionalism, ease with setting appointments and managing billing, and wait time were common categories in the staff/practice-related reviews.

By incorporating a three-step methodological approach using the 5:1 rating ratio and a reliable thematic codebook, physicians can delve into all the factors that formulate each rating that collectively generates their online reputation. As a result, physicians can begin to understand what their ratings mean both as an indicator of their online ratings compared to their peers and as a tool for improving the patient experience within their practice.

Full research publication is currently under review by the Journal of the American Medical Association.