
Research and Applications

Online physician ratings fail to predict actual performance on measures of quality, value, and peer review

Timothy J Daskivich,^{1,2} Justin Houman,¹ Garth Fuller,^{2,3} Jeanne T Black,⁴ Hyung L Kim,¹ and Brennan Spiegel^{2,3,5}

¹Division of Urology, Cedars-Sinai Medical Center, Los Angeles, CA, USA, ²Cedars-Sinai Center for Outcomes Research and Education (CS-CORE), Cedars-Sinai Medical Center, Los Angeles, CA, USA, ³Department of Medicine, Division of Health Services Research, Cedars-Sinai Health System, Los Angeles, CA, USA, ⁴Resource and Outcomes Management Department, Cedars-Sinai Health System, Los Angeles, CA, USA and ⁵Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, CA, USA

Corresponding Author: Timothy J Daskivich, Division of Urology Cedars-Sinai Medical Center, 8635 West 3rd Street, Suite 1070W, Los Angeles, CA 90048, USA. E-mail: Timothy.Daskivich@csmc.edu. Tel: 310-423-4700, Fax: 310-423-1886.

Received 13 March 2017; Revised 18 May 2017; Accepted 21 August 2017

ABSTRACT

Objective: Patients use online consumer ratings to identify high-performing physicians, but it is unclear if ratings are valid measures of clinical performance. We sought to determine whether online ratings of specialist physicians from 5 platforms predict quality of care, value of care, and peer-assessed physician performance.

Materials and Methods: We conducted an observational study of 78 physicians representing 8 medical and surgical specialties. We assessed the association of consumer ratings with specialty-specific performance scores (metrics including adherence to Choosing Wisely measures, 30-day readmissions, length of stay, and adjusted cost of care), primary care physician peer-review scores, and administrator peer-review scores.

Results: Across ratings platforms, multivariable models showed no significant association between mean consumer ratings and specialty-specific performance scores (β -coefficient range, $-0.04, 0.04$), primary care physician scores (β -coefficient range, $-0.01, 0.3$), and administrator scores (β -coefficient range, $-0.2, 0.1$). There was no association between ratings and score subdomains addressing quality or value-based care. Among physicians in the lowest quartile of specialty-specific performance scores, only 5%–32% had consumer ratings in the lowest quartile across platforms. Ratings were consistent across platforms; a physician's score on one platform significantly predicted his/her score on another in 5 of 10 comparisons.

Discussion: Online ratings of specialist physicians do not predict objective measures of quality of care or peer assessment of clinical performance. Scores are consistent across platforms, suggesting that they jointly measure a latent construct that is unrelated to performance.

Conclusion: Online consumer ratings should not be used in isolation to select physicians, given their poor association with clinical performance.

Key words: quality of health care, value of health care, quality assessment

BACKGROUND AND SIGNIFICANCE

In response to a growing awareness of variation in quality among health care providers,^{1,2} patients are increasingly using online consumer ratings to identify high-performing physicians. A 2012 survey

of 2137 US adults found that 65% were aware of physician-rating websites and that 59% reported these sites to be “somewhat important” or “very important” in choosing a physician.³ Furthermore, patients appear willing to trust online reviews to differentiate

the quality of physicians in the absence of other data. A survey of 1000 surgical patients at the Mayo Clinic found that 81% would seek consultation from a physician based on positive reviews alone and 77% would not seek consultation from a physician based solely on negative reviews.⁴ This appears to be a generalizable trend, since similar trends have been reported in Europe.⁵ No doubt responding to the interest of their customers, health care payers are now incorporating online consumer ratings into their physician search tools,⁶ suggesting that insurance companies also believe consumer ratings are trustworthy measures of ground-truth clinical performance. Given this trajectory, it is certain that consumer ratings will play an increasingly prominent role in how patients choose their physicians in the future.

Despite widespread interest, it is unknown whether online consumer ratings of physicians are effective at identifying high-performing physicians or even what they are measuring. The most widely used consumer ratings platforms offer no direction on what criteria should be used in physician assessment, which makes results difficult to interpret. Although many believe that consumer ratings are measuring ancillary components of the patient experience (eg, office environment, staff friendliness) rather than clinical performance,^{7–9} there are no strong data to support this claim, and the limited available data suggest the opposite.^{10–13} For example, large observational studies have shown that online consumer ratings are associated with coarse measures of quality such as board certification, rating of medical education, and physician volume.^{10,12} Health care consumers might be less enthusiastic about consumer ratings if scores were definitively shown to be poor predictors of actual clinical performance.

OBJECTIVE

We sought to determine whether specialist physician ratings from the 5 most popular online platforms predict multidimensional and empirical assessments of physician performance in a mixed cohort of medical and surgical specialists across diverse practice settings at a large urban health delivery system. In order to ensure robust capture of clinical performance, we used multidimensional assessments incorporating objective measures of quality and value¹⁴ of care, as well as subjective peer assessments from primary care colleagues and administrators. We also determined whether consumer ratings are consistent across ratings platforms, which would suggest that they jointly measure a uniform construct. The overarching goal of this study was to measure the validity and reliability of consumer physician ratings to inform patients, providers, and payers about whether and how to use these scores in selecting doctors.

MATERIALS AND METHODS

Data source and participants

We sampled physicians across a variety of medical and surgical specialties at Cedars-Sinai Medical Center, an 886-bed tertiary referral, hybrid academic-community hospital in Los Angeles. This study was conducted as part of a larger initiative within the Cedars-Sinai health system to perform quality ratings of our specialty physicians for the purpose of developing new provider narrow networks. All practicing physicians in the Cedars-Sinai medical delivery network within designated specialties were eligible for inclusion (n=123). Physicians were excluded if they were not using the health system's electronic health record in their ambulatory practice (n=30),

received <3 peer-review survey responses (n=14), or had no online consumer rating data available (n=1).

Variables

Primary predictors

Online consumer physician ratings. We collected consumer ratings for physicians in our sample across the 5 most popular online platforms according to Google Trends: Healthgrades, Vitals, Yelp, RateMDs, and UCompareHealth.¹⁵ Each website invites consumers to rate physicians using a 5-star scale. We collected data on each physician's average rating and the number of ratings that informed the average.

Covariates

We gathered demographic data on physicians in our sample, including race (Caucasian, Asian, or other), gender, practice type (faculty, medical group, or private practice), and years since medical school graduation, from a combination of physician profile data provided by Cedars-Sinai and the California Medical Board website.

Outcomes

Specialty-specific performance score. Specialty-specific performance score is a composite metric that includes outcomes assessing both quality (eg, adherence to Choosing Wisely measures [www.choosingwisely.org]) and value of care (eg, use of generic medications, case mix-adjusted length of stay) (Supplementary Appendix 1). Outcomes included in specialty-specific performance scores varied by specialty and were constructed through an intensive process of stakeholder engagement with specialty physicians. First, we conducted meetings with specialty leadership to review existing performance metrics, solicit new electronic health record-based metrics, and identify measures unique to each specialty. This resulted in a catalog of candidate measures for each specialty. Next, specialty physicians were asked to rate each candidate measure using the modified RAND Appropriateness Method.¹⁶ The top 5 metrics within each specialty comprised the outcomes for the final specialty-specific performance score. Finally, based on feedback from specialty leadership, each metric was assigned a weight between 0 and 100, with 0 representing "not relevant" to performance and 100 being "highly relevant." These relative weights represented the total possible score for each metric, and their sum represented the total possible specialty-specific performance score for each specialty (Supplementary Appendix 1).

Physicians in our sample were scored across metrics for fiscal years 2013–2014, and raw scores were standardized across specialties and transformed to the weighted scale. The total specialty-specific performance score was expressed as a proportion of possible points earned. Metrics were dropped for physicians with <20 qualifying encounters. Score calculations were performed by analysts blinded to consumer rating results.

PCP Survey score. The Primary Care Physician (PCP) Survey is a single-item question measuring the perception of specialist physicians' reputation, quality of care, and value-based care as assessed by primary care colleagues (see Supplementary Appendix 2).

Thirty-two PCPs rated each specialist using a 9-point modified RAND scale ranging from 1 ("completely disagree") to 9 ("completely agree") according to how much they agreed with the statement "Dr. [Name] is a strong partner in value-based health-care." To avoid an order effect, the survey presented specialist profiles in random order unique to each respondent. We used the

average rating for each physician across all respondents in our analysis.

Administrator survey score. The Administrator Survey is a single-item question that assesses perception of specialist physicians by administrators and hospital leaders using the same question structure described above for the PCP Survey. The administrator respondents included 20 members: departmental leaders who typically interact with specialists (eg, nursing, pharmacy); directors of inpatient clinical programs (eg, hospitalist medicine, emergency department, intensive care), and personnel with broad-based knowledge of specialist performance (eg, chief of medical staff, academic dean). We also surveyed groups of relevance for individual specialties (eg, dialysis staff for nephrologists, endoscopy staff for gastroenterologists, and anesthesiologists for surgeons). We used the average rating for each physician across all respondents in our analysis.

Statistical analysis

To visually characterize the relationship of online physician ratings scores with specialty-specific performance scores, we created a scatterplot with a Lowess line smoother overlay for each consumer ratings platform. We also calculated the Spearman's rank correlation coefficient between each platform and performance scores. To formally test whether consumer ratings scores were independently associated with performance scores, we created multivariable linear regression models predicting the mean performance score for each physician across platforms. The primary predictor was the mean consumer ratings score for each physician. Covariates included number of reviews, physician specialty, race, gender, and years since medical school graduation. We characterized the association of consumer ratings with PCP and administrator scores in an identical fashion.

In order to assess whether consumer ratings scores were associated with subdomains of performance scores related to quality of care, we created separate multivariable linear regression models predicting adherence to Choosing Wisely measures and best practice alerts for each consumer ratings platform, while adjusting for the same covariates. In order to assess whether consumer ratings scores were associated with subdomains of the performance score related to value-based practice, we created similar multivariable linear regression models predicting generic medication utilization rate and case mix-adjusted length of stay index by consumer ratings score for each consumer ratings platform, while adjusting for the same covariates.

To investigate the consistency of consumer ratings scores across various platforms, we constructed scatterplots with univariable linear regression overlays for pairwise comparisons of consumer ratings platforms. We then calculated the Spearman's rank correlation coefficient for each comparison. To test whether consumer ratings scores from one platform independently predicted scores from another, we created multivariable linear regression models for pairwise comparisons of platforms. Covariate structure was identical to the aforementioned models.

We used $P < .05$ to denote statistical significance of 2-sided tests. All statistical analyses were performed in Stata 11.0 (Stata Inc., College Station, TX, USA). The Cedars-Sinai Institutional Review Board approved this study.

RESULTS

Characteristics of our sample are listed in [Table 1](#). The median number of consumer ratings per physician ranged from 4 to 9 across

Table 1. Sample characteristics

Variable	Frequency (%) or Median (IQR)
Specialty	
Cardiology	25 (32)
Neurology	11 (14)
Nephrology	10 (13)
Obstetrics/gynecology	9 (12)
Endocrinology	7 (9)
Otolaryngology	6 (8)
Gastroenterology	6 (8)
General surgery	4 (5)
Race	
Caucasian	50 (64)
Asian	19 (24)
Other	9 (12)
Gender	
Male	66 (84)
Female	12 (16)
Type of practice	
Cedars-Sinai Medical Group	36 (46)
Private practice	34 (44)
Faculty group	8 (10)
Years in practice	
Median	33.5 (19–43)
Consumer ratings scores	
Yelp score	4 (3.5–5)
Yelp number of reviews	9 (4–15)
Healthgrades score	4.2 (3.7–4.8)
Healthgrades number of reviews	8 (5–12)
Vitals score	4 (3.5–4.5)
Vitals number of reviews	4 (2–7)
RateMDs score	4 (3.5–4.5)
RateMDs number of reviews	5 (2.5–9)
UCompareHealthcare score	4.5 (4–5)
UCompareHealthcare number of reviews	4 (2–9)
Performance outcomes	
Specialty-specific performance score	0.8 (0.5–0.9)
PCP score	6.8 (5.5–7.8)
Administrator score	6.7 (5.3–7.5)

platforms. There was substantial variability in specialty-specific performance scores (median 0.8, interquartile range [IQR] 0.5–0.9), PCP scores (median 6.8, IQR 5.5–7.8), and administrator scores (median 6.7, IQR 5.3–7.5).

Scatterplots and Lowess line smoother overlays depicting mean consumer ratings scores versus mean specialty-specific performance scores showed no relationship between physician rating and performance score ([Figure 1A](#)). Across all platforms, there was either a negative or weak correlation between consumer ratings score and performance score (range $-0.18, 0.02$). Multivariable models also showed no statistically significant or meaningful association between consumer ratings score and performance score (β -coefficient range, $-0.04, 0.04$) ([Table 2](#)). Sensitivity analyses showed no difference in significance or magnitude of comparisons when models were retested in subgroups of specialties. Very few physicians with low mean consumer ratings scores also had low performance scores; among physicians in the lowest quartile of mean specialty-specific performance scores, only 5%–32% had consumer ratings scores in the lowest quartile across platforms. Multivariable linear regression models found no relationship between consumer scores and subdomains of performance scores reflecting either quality of care (ie, ad-

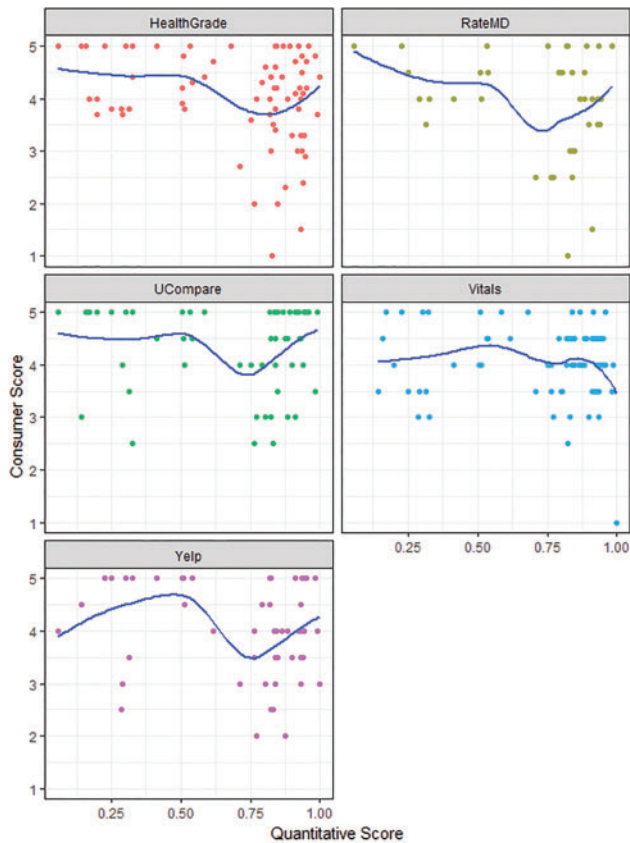


Figure 1. (A) Scatterplots and Lowess line smoother overlays depicting mean consumer ratings scores and mean specialty-specific physician performance scores.

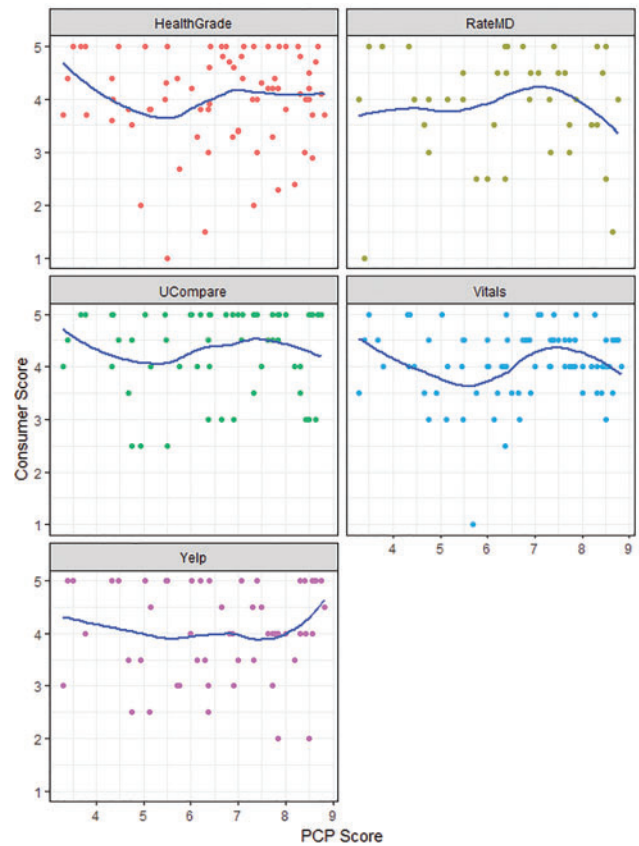


Figure 1. (B) Scatterplots and Lowess line smoother overlays depicting mean consumer ratings scores and mean PCP scores.

herence to Choosing Wisely measures and best practice alerts) or value-based care (ie, use of generic medications and length of stay index).

Scatterplots and Lowess line smoother overlays depicting mean consumer ratings scores versus mean PCP and administrator scores showed no relationship between physician rating and either PCP or administrator peer assessments (Figures 1B and C). Spearman's rank correlations between consumer ratings scores and PCP scores (range -0.04 , 0.07) as well as consumer ratings scores and administrator scores (range -0.02 , 0.18) were weak or negative across all platforms. Multivariable models also showed no significant association between consumer ratings scores and either PCP (β coefficient range, -0.01 , 0.3) or administrator scores (β coefficient range, -0.2 , 0.1) (Table 2).

Consumer ratings of individual physicians were generally consistent across platforms. Scatterplots and linear regression overlays of pairwise comparisons across platforms revealed a positive relationship among all platforms (Figure 2). Furthermore, multivariate models comparing mean consumer ratings scores across platforms showed a positive association that was statistically significant in 5 of 10 comparisons (Table 3).

DISCUSSION

Despite broad interest and uptake of online consumer physician ratings, our study shows that ratings are not associated with core measures of specialist physician performance, as determined by

multidimensional assessments including objective measures of quality of care, value of care, and peer assessments by PCPs and administrative staff. Not only were average ratings not statistically associated with performance outcomes, but the magnitudes of association were also weak, or even negative in some cases. For example, for every point change in consumer ratings score, performance scores only changed between -0.04 and $+0.04$ points across all platforms. Furthermore, subgroups of performance scores assessing quality (ie, adherence to Choosing Wisely measures or best practice alerts) and value of care (ie, use of generic medications when appropriate and case-mix-adjusted length of stay) showed no association with mean physician rating. Consumer ratings scores were poor at identifying low-performing physicians; in our sample, only 5%–32% of the physicians in the lowest quartile of specialty-specific performance had mean ratings scores in the lowest quartile.

Our results support what has been a longstanding belief among health care stakeholders: that consumer rating scores are not reflective of the basic quality and value of care provided by an individual physician,^{7–9} which are unquestionably among the most important criteria to consider when selecting a health care provider. This information has direct implications for the $\sim 80\%$ of health care consumers who are currently selecting physicians based on ratings alone.^{4,5}

While many have postulated that online physician ratings are not a valid metric of physician performance,^{7–9} there are no high-quality data to support this claim. In contrast, consumer ratings do seem to predict quality on the institutional level. Two large retrospective studies showed that web-based patient ratings of hospitals

predicted hard endpoints of institutional quality, including Hospital Consumer Assessment of Healthcare Providers and Systems survey scores¹⁷ as well as hospital readmissions for myocardial infarction, heart failure, and pneumonia.¹⁸ Yet data linking consumer ratings to physician quality are more limited, mainly by the granularity of quality endpoints. A study of National Health Service Choices data in the United Kingdom showed that among the 69% of physicians who were “recommended” in survey reviews, there were moderate associations with patient experience (Spearman’s 0.37–0.48, $P < .001$) and weak associations with measures of clinical process and outcome (Spearman’s less than ± 0.18 , $P < .001$).¹¹ Other studies showed that online physician reviews could be used to predict

physician volume for major operations (as a proxy of quality)¹² as well as board certification status and quality of medical education.¹⁰ However, none of these studies has robustly characterized physician performance across quality, value, and peer-review metrics as in this study; it is unlikely that larger studies will be able to assess physician performance on such a granular and multimodal level.

In judging the validity of consumer physician ratings, we should apply a similar psychometric framework for any other questionnaire. The validity of a questionnaire has many levels, ranging from least to most rigorous, including face validity, content validity, predictive validity, and concurrent validity. Online physician ratings fail to meet basic face validity even before measuring empirical relationships. In the absence of any recommendations on how to assess physicians, it remains unclear whether a 5-star item measures an explicit construct of any kind. Content validity is also poor, since no consistent categories of patient experience are measured. The current study further discredits the construct validity of consumer ratings, including predictive validity (ie, the ability of scores to predict physician performance) and concurrent validity (ie, the ability to distinguish between highly and poorly performing physicians).

Yet, despite the lack of understanding of what consumer ratings of physicians are intended to measure, we found that ratings were relatively consistent across platforms, suggesting that they may jointly measure a latent construct. Nonetheless, it remains unclear what the latent construct being measured is, though in part it may be patient satisfaction. A recent publication found that online reviews were modestly associated with National Committee for Quality Assurance consumer satisfaction scores for insurance plans (Pearson correlation = 0.376).¹⁹ Defining this latent construct is critical for appropriate use of consumer ratings scores, since it will at least offer patients a sense of what these scores mean, which is especially important given their strong influence on provider selection for the majority of patients.

Despite the shortcomings of online physician ratings, it is clear that patient assessments of physicians are important to health care consumers and payers and that this information is here to stay. How can the utility of consumer ratings be improved? First, consumer ratings websites should create a definable construct around which the ratings are intended to measure, and there should be instructions on how to measure physicians across metrics that patients are well positioned to evaluate. These areas of content should be informed by both expert and patient stakeholders to maximize the relevance of the data, thereby optimizing content validity. Second, consumer ratings should be paired with complementary data on quality and value of care, since it appears that these critically important data are not being

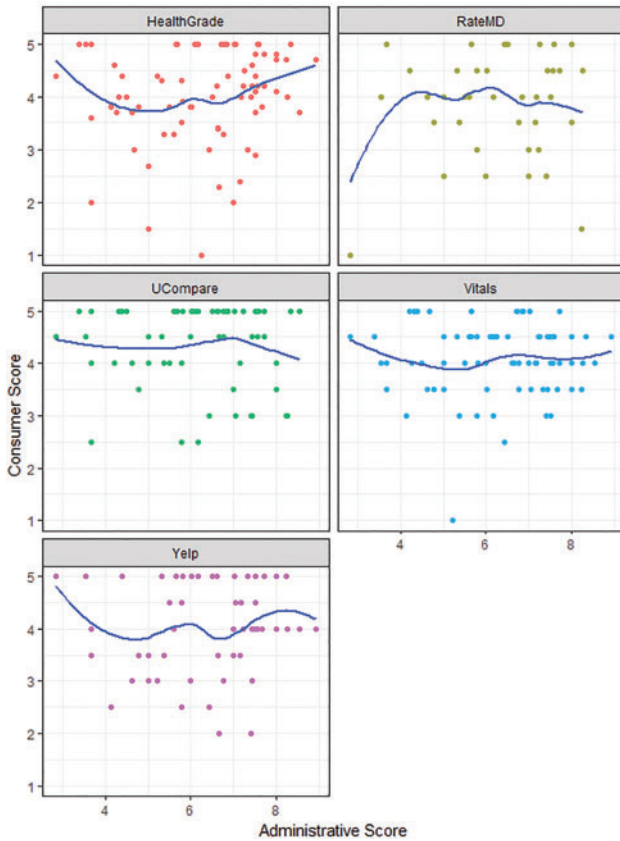


Figure 1. (C) Scatterplots and Lowess line smoother overlays depicting mean consumer ratings scores and mean administrator scores.

Table 2. Multivariable linear regression models predicting mean specialty-specific physician performance scores by mean consumer ratings score by platform

Consumer ratings platform	Specialty-specific performance score (β coefficient, 95% confidence interval)	PCP score (β coefficient, 95% confidence interval)	Administrator score (β coefficient, 95% confidence interval)
Yelp	-0.0001 (-0.06 - 0.06)	-0.01 (-0.4, 0.4)	-0.1 (-0.5, 0.2)
Healthgrades	-0.005 (-0.05 - 0.04)	0.007 (0.3, 0.4)	0.06 (-0.3, 0.4)
Vitals	-0.04 (-0.08 - 0.01)	0.02 (-0.3, 0.4)	-0.2 (-0.5, 0.2)
RateMDs	0.01 (-0.03 - 0.05)	0.2 (-0.3, 0.7)	0.1 (-0.3, 0.5)
UCompareHealthcare	0.04 (-0.004 - 0.08)	0.3 (-0.06, 0.8)	-0.009 (-0.5, 0.4)

Multivariable models adjusted for number of reviews, specialty, race, years since medical school graduation, gender, and practice setting.

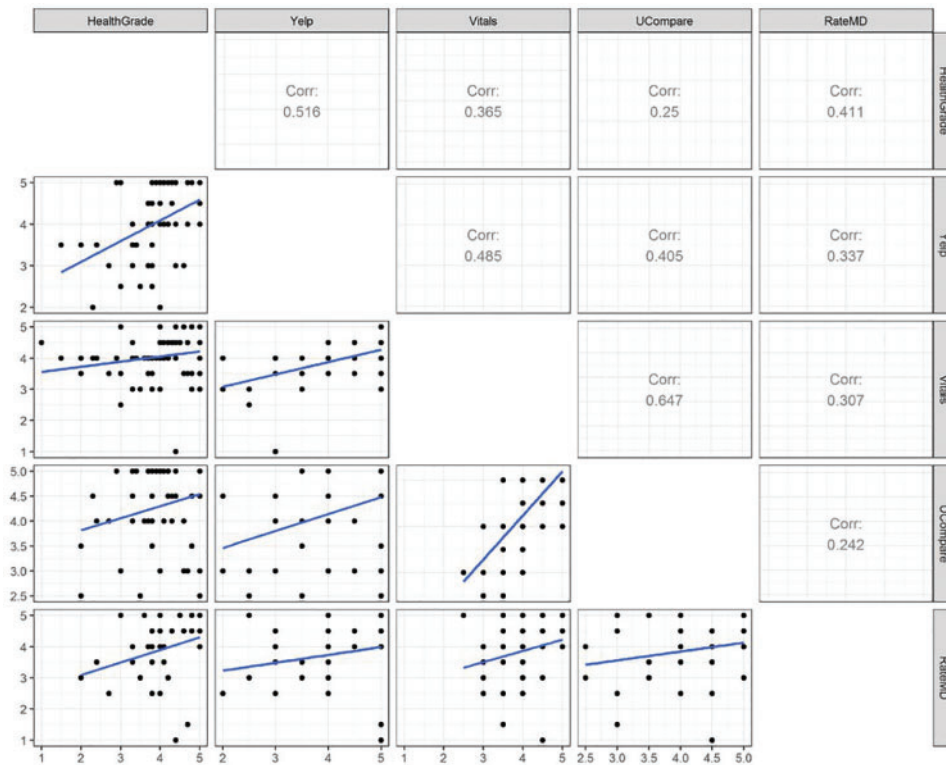


Figure 2. Scatterplots and linear regression overlays comparing mean consumer ratings scores across platforms.

Table 3. Multivariable linear regression models for pairwise predictions of mean consumer ratings score by platform

Primary predictor	Yelp	Healthgrades	Vitals	RateMDs	UCompareHealthcare
Yelp	██████████	0.28 (0.03, 0.52)*	0.52 (0.30, 0.74)*	0.12 (-0.31, 0.55)	0.45 (0.10, 0.81)*
Healthgrades	0.44 (0.05, 0.83)*	██████████	0.12 (-0.13, 0.38)	0.27 (-0.24, 0.79)	0.25 (-0.08, 0.59)
Vitals	0.71 (0.41, 1.02)*	0.13 (-0.14, 0.40)	██████████	0.60 (0.07, 1.13)*	0.96 (0.67, 1.26)*
RateMDs	0.14 (-0.38, 0.67)	0.17 (-0.15, 0.50)	0.29 (0.03, 0.56)*	██████████	0.39 (-0.10, 0.89)
UCompareHealthcare	0.45 (0.10, 0.81)*	0.22 (-0.07, 0.50)	0.53 (0.37, 0.70)*	0.28 (-0.07 - 0.63)	██████████

Multivariable models adjusted for number of reviews, specialty, race, years since medical school graduation, gender, and practice setting.

*P < .05

measured by consumer ratings. Last, companies offering physician rating services should accept the responsibility of helping their customers understand the limitations of consumer ratings when selecting physicians, given that stakes are higher than in other marketplace settings and since health care consumers are poorly positioned to evaluate some aspects of care, such as technical quality and appropriateness.

Our study has important limitations that may affect generalizability. First, although we have collected extensive data on the performance in our sample of physicians, the sample size is relatively small. Nonetheless, despite many comparisons across diverse performance metrics and multiple platforms, we found no signs of any statistically or clinically significant relationships, other than among the consumer scores themselves. Increasing the sample size is therefore unlikely to capture any meaningful association. Second, our sample was drawn from a single center study in an urban tertiary-care setting, which may impact external generalizability to dissimilar environments. However, Cedars-Sinai has a pluralistic provider mix inclusive of academic, managed care, and private practice physicians, which reduces

the sampling bias associated with testing the same question in purely academic settings. Third, we did not have access to the Consumer Assessment of Healthcare Providers and Systems or patient satisfaction data as part of this project, so we could not assess the association of online consumer reviews with these metrics. Last, our objective measures of quality and value of care were heterogeneous across specialties, which obfuscates comparisons among specialties. However, there was no difference in significance or magnitude of comparisons when models were retested in subgroups of specialties.

CONCLUSION

In conclusion, despite the fact that over three-quarters of patients are willing to make provider choices based on consumer physician ratings alone, our study found no significant associations between consumer ratings and a multidimensional assessment of specialist physician performance. Given the strong influence that consumer ratings have on provider selection, it is likely that health care consumers mistakenly

assume that ratings are indicative of some core elements of physician performance. In fact, our data reveal that consumer ratings are especially bad at identifying poorly performing physicians, since only a small minority of these physicians are identified by low scores.

However, consumer ratings are generally consistent across platforms, suggesting that they measure a latent construct that is unrelated to performance. Consumer ratings of physicians should be paired with information on quality and value of care in order to help patients make more informed decisions when selecting providers.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or non-for-profit sector.

COMPETING INTERESTS

The authors have no competing interests to declare.

CONTRIBUTORS

All authors made significant contributions to the research effort and the final manuscript, and all authors approved the final submission.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

REFERENCES

- McGlynn EA, Asch SM, Adams J, *et al*. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348(26):2635–45.
- Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003;349(22):2117–27.
- Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. *JAMA*. 2014;311(7):734–35.
- 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.
- Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. *J Med Internet Res*. 2013;15(8):e187.
- United Healthcare: Healthgrades Patient Satisfaction Ratings*. 2016, <http://health.usnews.com/health-news/patient-advice/articles/2014/12/19/are-online-physician-ratings-any-good>. Published Dec 2014. Accessed August 14, 2017.
- Glover L. Are online physician ratings any good? *US News and World Report*. 2014. <http://www.nytimes.com/2012/03/10/your-money/why-the-web-lacks-authoritative-reviews-of-doctors.html>. Published March 9, 2012. Accessed August 14, 2017.
- Lieber R. The Web Is Awash in Reviews, but Not for Doctors. Here's Why. *New York Times*. March 9, 2012.
- Ellimoottil CLS, Wright CJ. Online physician reviews: The good, the bad, and the ugly. *Bull Am Coll Surg*. 2013;98(9):34–39.
- Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. *J Med Internet Res*. 2012;14(1):e38.
- Greaves F, Pape UJ, Lee H, *et al*. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. *J Med Internet Res*. 2012;14(5):e146.
- Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? *J Med Internet Res*. 2012;14(2):e50.
- Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. *J Med Internet Res*. 2014;16(2):e56.
- Porter, ME. What Is Value in Health Care? *New Engl J Med*. 2010; 363:2477–81.
- Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *J Med Internet Res*. 2011;13(4):e95.
- Fitch K. The RAND/UCLA Appropriateness Method User's Manual. Vol 1269-DG-XII/RE. Santa Monica, CA: Rand Corporation; 2001.
- Ranard BL, Werner RM, Antanavicius T, *et al*. Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. *Health Aff (Millwood)*. 2016;35(4):697–705.
- Bardach NS, Asteria-Penalosa R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. *BMJ Qual Saf*. 2013;22(3):194–202.
- Shetty PR, Rivas R, Hristidis V. Correlating ratings of health insurance plans to their providers' attributes. *J Med Internet Res*. 2016;18(10):e279.