Do Crowdsourced Hospital Ratings Coincide with Hospital Compare Measures of Clinical and Nonclinical Quality?

Victoria Perez and Seth Freedman

Objective. To determine the consistency with which government-issued hospital quality ratings and crowdsourced ratings on social media sites identify hospital quality.

Data Sources. Hospital ratings from Facebook, Google, and Yelp were linked with Hospital Compare (HC) measures.

Study Design. Fixed-effects linear regression model.

Principal Findings. Among crowdsourcing sites' best-ranked hospitals, 50–60% were also the best ranked on HC's overall and patient experience ratings; 20% ranked as the worst. Best-ranked hospitals had significantly better clinical quality scores than worst ranked hospitals, but were not more likely to be the highest rated in terms of HC's clinical quality measures alone.

Conclusions. Crowdsourcing sites and HC provide comparable information on patient experience; scores were less consistent in terms of risk-adjusted measures of patient safety and clinical quality.

Key Words. Hospitals, quality, measurement

The Internet has increasingly become a source of information on hospital quality. Since 2005, the Centers for Medicare & Medicaid Services (CMS) have publicly posted hospital ratings based on both patient satisfaction surveys and clinical quality metrics on Hospital Compare (HC). More recently, hospital reviews have become prevalent on crowdsourcing sites, such as Facebook, Google Reviews, and Yelp. As consumers increase their use of these sites, it is important to understand what dimensions of information overlap between reviews on these sites and Hospital Compare. Whereas HC ratings are based on validated survey instruments and Medicare claims data for fixed, lagged time periods, crowdsourced ratings and comments are

based on unstructured feedback of patients or family members and are updated in real time. Our study examined the extent to which ratings on crowdsourced sites would be expected to lead patients to high-quality hospitals based on HC clinical quality and patient experience ratings.

Experimental evidence (Kanouse et al. 2016) has found that when provided with patient satisfaction survey results, clinical performance measures, and unstructured patient comments about providers, individuals spend more time reviewing the information than when only provided survey results and clinical measures, but they are less likely to choose the provider with the highest quantified ratings. Additional experimental evidence (Emmert and Schlesinger 2017) has found that patient perception of clinical quality is unaffected when clinical quality was presented with or without additional narrative comments. Taken together, the experimental studies suggest that patient comments provide insights on factors that matter for individual decision-making, but are not necessarily related to clinical quality. However, star ratings and free-response comments provided without structure may also be misleading to individuals because they are an overall amalgamation of individual experiences and values. For example, extracting reliable and consistent quality measures from patient narratives requires validated protocols that revolve around an *ex ante* set of criteria (Grob et al. 2016).

Early studies of the correlations between HC and crowdsourcing sites were motivated by the questions of whether HC and other sources of online hospital ratings provided similar or different information. One challenge for these studies was low representation of hospitals on crowdsourcing sites. In 2011, only 25 percent of 3,796 hospitals with Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) scores, which make up the patient experience portion of HC, had a Yelp rating (Bardach et al. 2012). Other studies that limit the focus to a single state have found a gap between patient experience and clinical scores. In a comparison of the HC scores and Facebook ratings of the 136 New York hospitals (Campbell and Li 2017), Facebook star rating was correlated with patient satisfaction in 21 of the 23 patient experience measures, but not clinical outcomes or treatment intensity. In terms of the content of written comments, when comparing the content of Yelp reviews to the domains covered by the HCAHPS survey questions, researchers (Ranard et al. 2016) found that Yelp reviewers discussed the domains

Address correspondence to Victoria Perez, Ph.D., and Seth Freedman, Ph.D., School of Public and Environmental Affairs, Indiana University, 1315 E 10th St, Bloomington, IN 47405; e-mails: viaperez@indiana.edu and freedmas@indiana.edu.

covered by the HCAHPS survey, but also wrote about twelve additional domains not covered by the HCAHPS.

In 2015, the Centers for Medicare & Medicaid Services (CMS; Rau 2015) synthesized 75 Hospital Compare (HC) measures into a five-star scale rating. Critics (American Hospital Association 2016) characterized the rating as “confusing for patients and families trying to choose the best hospital.” Subsequent studies (Trzeciak et al. 2016; Wang et al. 2016) explored correlations between HC ratings and crowdsourced site ratings and found weakly positive correlations among selected clinical and patient experience measures within HC and between HC measures and crowdsourced ratings. In a 2017 study (Hefele et al. 2017), among the 20 percent of nursing homes with a Facebook user rating, the CMS five-star clinical composite score was weakly correlated with Facebook reviews.

Extending this work, we used more recent data from the three largest crowdsourced rating sites, Google, Yelp, and Facebook, to compare rankings between these sites and HC within markets. Our approach differed from previous work in two key ways. First, in our more recent data, across these three sites, at least 90 percent of hospitals with HC ratings also have crowdsourced ratings. Second, we made comparisons within hospital markets, as the overall correlation between ratings could be driven by common geographic factors. We asked if a patient was to choose the “best”-rated hospital in their market based on each crowdsourced site, how often would they also be choosing the “best” and how often would they in fact be choosing the “worst” hospital based on the most heavily weighted dimensions of HC’s overall hospital score: patient experience, mortality, readmissions, and patient safety. This study established how the “best” and “worst” rankings in a hospital market depend on the chosen information source.

STUDY DATA AND METHODS

We extracted five measures from the 2016 HC data: the overall hospital star rating, overall patient experience rating, all-cause unexpected 30-day readmission rates, 30-day pneumonia mortality rates, and intestinal infection rates (*Clostridium difficile* laboratory-identified events). The overall hospital five-star rating calculation is a five-step process within seven different domains. We focused on representative measures of the four domains weighted most heavily in the final score calculation (22 percent each): patient experience from HCAHPS, readmission rates, mortality rates, and patient

safety. Within readmission rates, mortality rates, and patient safety measures, we chose the measure from each dimension with the lowest rate of missing scores. Missing scores occurred when hospitals had an insufficient number of cases to meet CMS's public reporting threshold. Therefore, the conditions with the fewest number of missing scores were the most common conditions within each set at the time of analysis.

The sample included acute care hospitals with a score in HC: 2,995 hospitals. HC's overall hospital and overall patient experience ratings were synthesized by CMS into five-star scales. For the other measures, we categorized hospitals by quintiles of the HC measure's distribution.

During July 2017, we collected star ratings from three major crowdsourcing sites, Google, Yelp, and Facebook, via searches using Google's API. These data include all reviews posted between when the hospital first received a review through the data collection period. Each site uses a five-star scale to represent aggregated individual user reviews, although the proprietary algorithms differ across sites. All three sites indicated higher quality with higher star ratings. In terms of the share of hospitals observed in the sample, 98 percent, 93 percent, and 90 percent of hospitals present on HC had a Google, Facebook, or Yelp rating, respectively. Yelp, established in 2004, applies a filter to reviews to proactively filter out false reviews (Kamerer 2014). Google reviews, established in 2012, allow users to leave reviews without an account and, subsequently, ratings were integrated with the Google search engine, raising their visibility. Facebook reviews, established in 2013, were also integrated with Google, but require a log-in for reviewers. Facebook is the only site that permits hospitals to elect not to receive any star rating. If hospitals do elect a rating on Facebook, then, similarly to Yelp and Google, they cannot filter the content of those ratings (Weinberg 2017).

We compared ratings between HC and crowdsourced sites among hospitals within the same market, as defined by hospital referral regions (HRR). Within each market, we determined which hospitals had either the highest star rating on each crowdsourced site or were tied for the highest rating. Across all three sites, the average HRR contained nine hospitals with ratings and an average of 1–1.6 hospitals identified as “best.” We then calculated the fraction of those hospitals with the highest (or tied for highest) and lowest (or tied for lowest) ratings within the market according to HC metrics. For the small number of HRRs where the crowdsourced or HC scores were the same for all hospitals, we classified them all as the “best.”

The first part of the analysis considered the difference in signal for patients that use a simple decision to go to the “best” hospital or to avoid the

“worst” hospital in their market between crowdsourced hospital ratings and ratings produced by CMS. To extend the analysis, we next considered whether the decision to use the “best” hospital based on crowdsourced hospital ratings resulted in statistically significantly higher quality as measured by each HC Metric. We estimated these differences in a linear regression model of the following form:

$$\text{CMS Rating}_h = \alpha + \text{Crowd Sourced Rating}_h \beta + \theta_{\text{Market}} + \epsilon_{\text{Market}} \quad (1)$$

In this specification, the outcomes of interest were at the hospital level. Outcomes were the star rating of overall hospital quality from HC, patient experience, all-cause readmissions, pneumonia mortality, and intestinal infections. HC’s overall hospital rating and the HCAHPs’ patient experience rating were measured on a five-star scale, where five indicated the highest possible score. Hospitals’ scores of readmissions and mortality were the risk-adjusted rates of these outcomes within 30 days following discharge. The hospital-acquired infection was the hospital’s standardized infection ratio (SIR), which compares the actual incidence of infections with a predicted value. SIRs greater than one indicated a higher rate of infection (poorer outcome); SIRs less than one indicated a positive outcome.

The independent covariates of interest were represented by the vector Crowd Sourced Rating_h. It consisted of two indicator variables to signify whether the hospital was ranked as the “best” in its respective market, which we defined as the hospital referral region, or as neither best nor worst in its HRR. The worst ranking was the omitted category. The coefficients of interest (β) presented the association between scores collected from crowdsourced sites and scores generated from medical claims and survey data compiled by CMS.

The regression also included market fixed effects (θ_{Market}). These effects controlled for fixed attributes of markets and led our estimates of β to be driven by comparing hospitals within markets. With this approach, in markets with only low-performing hospitals, the best of those low-performing hospitals was recognized as the best possible choice of available hospitals. Standard errors were clustered at the market level.

STUDY RESULTS

In studies of crowdsourced restaurant ratings, Luca and Zervas (2016) observed a bimodal distribution because clients with extremely positive or

negative experiences were compelled to submit an online review. Contrary to this finding, the distribution of hospital star ratings across the three crowdsourced sites was unimodal, as seen in Figure 1. Given that hospitals on Facebook can elect to not have any scores shown on their page, the distribution was centered higher on the five-star rating scale than either Google or Yelp. Table 1 shows these star ratings are based on an average of 31 reviews per hospital on Yelp (median of 9), 357 on Facebook (median of 209), and 52 on Google (median of 37). Hospital Compare measures and basic hospital characteristics did not vary between the sets of hospitals with reviews on each site.

HC's overall summary measure includes dimensions related to patient experience, clinical quality, and safety. In Figure 2, if users chose the highest ranking hospital in their HRR from a crowdsourcing site, the recommendation would be consistent with the highest ranking hospital based on HC's overall summary measure approximately 50 percent of the time for Facebook and Google and 40 percent of the time for Yelp. The "best" hospital identified by all three crowdsourcing sites had the worst ranking HC score 19–22 percent of the time.

Figure 3 shows that agreement in hospital rankings as the "best" among crowdsourced sites increased by seven to ten percentage points when compared to HC patient experience ratings based on HCAHPS surveys, rather than to overall scores. Hospitals ranked "best" among crowdsourced site ratings were identified as the "worst" in terms of HC patient experience 18–20 percent of the time.

Figure 3 also shows that results for clinical quality measures were consistent across measures of 30-day all-cause readmission rates and 30-day mortality among pneumonia patients. Hospitals ranked as the "best" on crowdsourced sites were the "best" in terms of either clinical quality measure approximately 30 percent of the time. Hospitals ranked as the "best" among crowdsourced site were the "worst" in terms of clinical quality measures 30–37 percent of the time. Finally, hospitals ranked as the "best" on crowdsourcing sites were similarly likely to be identified as the "best" (26–34 percent) or "worst" (23–31 percent) in terms of patient safety.

Based on the regression results presented in Table 2, hospitals ranked best on Facebook had an HC overall score 13 percent higher than the worst hospitals in their market (increasing 0.37 stars on an average of 2.8 stars). For hospitals that were not the best and not the worst (mid-level), the score was six percent higher than the worst hospital (increasing 0.17 stars). Hospitals ranked the best on Yelp and Google had average overall star ratings 12 percent higher

Figure 1: Distribution of Facebook, Google, and Yelp Star Ratings

Source. Authors' analysis of hospital reviews posted to Facebook, Google, and Yelp collected during July, 2017.

Table 1: Characteristics of Hospitals with Crowdsourc Site Ratings

	<i>All Hospitals</i>	<i>Yelp Rated</i>	<i>Facebook Rated</i>	<i>Google Rated</i>
Hospital compare measures				
Hospital overall rating	3.01 (0.85)	3.00 (0.87)	3.01 (0.85)	3.00 (0.85)
Overall hospital rating	3.46 (0.86)	3.45 (0.86)	3.47 (0.86)	3.46 (0.86)
Pneumonia 30-day mortality	16.3 (2.22)	16.3 (2.20)	16.3 (2.21)	16.3 (2.22)
Hospital-wide 30-day readmission	15.6 (0.94)	15.6 (0.95)	15.6 (0.94)	15.6 (0.94)
Colon infection	0.96 (0.51)	0.97 (0.49)	0.96 (0.51)	0.96 (0.51)
Hospital finances				
Adult and pediatric beds	210.4 (197.8)	223.7 (202.8)	212.5 (196.6)	210.7 (197.9)
HHI	0.11 (0.095)	0.11 (0.097)	0.12 (0.096)	0.11 (0.095)
Crowdsourced reviews/hospital				
Mean number of reviews/hospital	—	30.9 (73.0)	356.6 (713.4)	51.7 (49.5)
Median number of reviews/hospital	—	9	209	37
Range number of reviews/hospital		(0–499)	(1–891)	(2–20,884)
Observations	2,955	2,647	2,755	2,892

Notes. Column 1 is the set of hospitals with a Hospital Compare measure. Each subsequent column contains a subset of hospitals that have a crowd sourced site rating.
Source. Authors’ analysis of data from 2016 Hospital Compare database, the 2014 Medicare Cost Report Data, and reviews posted to Facebook, Google, and Yelp collected during July, 2017.

than the worst hospitals. Mid-level hospitals on these sites scored between four percent (Yelp) and six percent (Google) higher than the worst ranked hospital. Similarly, the best-ranked hospitals had 20 percent (Facebook and Google) and 13 percent (Yelp) higher star rating of overall patient satisfaction. Mid-level hospitals had higher scores, but at a smaller magnitude (8–16 percent). These results were all statistically significant at the five percent level.

Best-ranked hospitals on crowdsourced sites had better HC clinical quality and patient safety scores, and most of these differences were statistically significant. Across the three crowdsourcing sites, the best-ranked hospitals had one to two percent lower readmission rates then the worst ranked hospitals, and these differences were statistically significant at the five percent level. Mortality rates among best-ranked hospitals were similarly one to two percent lower, although these differences were only statistically significant at the five

Figure 2: Percent of Hospitals Ranked Best on Facebook, Google, and Yelp that were Ranked either Best or Worst on Hospital Compare's Overall Hospital Rating

Notes. 359 hospitals were ranked “best” within the sample on Facebook. 333 hospitals were ranked “best” within the sample on Google. 491 hospitals were ranked “best” within the sample on Yelp. *Source.* Authors’ analysis of data from 2016 Hospital Compare database and reviews posted to Facebook, Google, and Yelp collected during July, 2017.

percent level on Facebook and the ten percent level on Google. Best-ranked hospitals on Facebook and Google had hospital-acquired infection rates lower than the worst ranked hospital by 14 (statistically significant at the ten percent level) and 13 percent (statistically significant at the five percent level), respectively.¹

We tested the sensitivity of these findings in several ways and found that results were largely consistent with the main findings. Results from these sensitivity analyses are available in the Appendix SA2. First, to address the potential bias of few reviews, we excluded hospitals with fewer than five reviews. Second, we tested the sensitivity of our main findings to three possible conclusions consumers might infer about hospitals that do not have a crowdsourced score: That hospitals without a score are (1) among the best of hospitals in their market; (2) among the worst of hospitals in their market; or (3) neither best

Figure 3: Percent of Hospitals Ranked Best or Worst on Facebook, Google, and Yelp that were Ranked either Best or Worst on Representative Components of Hospital Compare’s Overall Hospital Rating

Notes. (1) Patient Experience and Readmissions: 359 hospitals were ranked “best” within the sample on Facebook. 333 hospitals were ranked “best” within the sample on Google. 491 hospitals were ranked “best” within the sample on Yelp. (2) Pneumonia Mortality: 345 hospitals were ranked “best” within the sample on Facebook. 322 hospitals were ranked “best” within the sample on Google. 471 hospitals were ranked “best” within the sample on Yelp. (3) Intestinal Infections: 257 hospitals were ranked “best” within the sample on Facebook. 210 hospitals were ranked “best” within the sample on Google. 302 hospitals were ranked “best” within the sample on Yelp.
Source. Authors’ analysis of data from 2016 Hospital Compare database and reviews posted to Facebook, Google, and Yelp collected during July, 2017.

nor worst. Third, the HC overall and patient experience scores were assigned to star ratings by CMS so that the majority of hospitals were given scores in the middle of the rating distribution, whereas we assigned clinical quality measure categories based on even quintiles. Therefore, we ensured that the differences in results between patient experience scores and clinical quality were not driven by this difference in distributions. We first reassigned categorical ratings to clinical measures in order to match the star distribution of the HCAHPS ratings, and we initially found that clinical quality measures were slightly more correlated with crowdsourced ratings. However, these distributions resulted in many more markets in which all hospitals were tied at two or three stars and therefore

Table 2: Regression Results: $\text{CMS Rating}_h = \alpha + \text{Crowd Sourced Rating}_h \beta + \theta_{\text{Market}} + \epsilon_h$

	<i>Five-Star Rating</i>		<i>Risk-Adjusted Rate (%)</i>		<i>Standardized Infection Ratio</i>
	(1) <i>Overall</i>	(2) <i>Patient Experience</i>	(3) <i>All-cause Readmission</i>	(4) <i>Pneumonia Mortality</i>	(5) <i>Colon Infection</i>
Facebook					
Best	0.37*** (0.062)	0.61*** (0.059)	-0.27*** (0.064)	-0.39** (0.12)	-0.22+ (0.11)
Neither best/ worst	0.17** (0.051)	0.49*** (0.051)	-0.12* (0.053)	-0.21* (0.093)	-0.22* (0.10)
Constant	2.80*** (0.049)	2.98*** (0.048)	15.5*** (0.051)	15.1*** (0.089)	1.61*** (0.095)
Observations	2,642	2,642	2,642	2,131	1,544
Yelp					
Best	0.32*** (0.057)	0.43*** (0.065)	-0.16** (0.061)	-0.13 (0.10)	-0.12 (0.086)
Neither best/ worst	0.10* (0.046)	0.27*** (0.052)	-0.067 (0.045)	-0.086 (0.083)	-0.034 (0.076)
Constant	2.78*** (0.035)	3.23*** (0.040)	15.6*** (0.035)	15.1*** (0.063)	1.49*** (0.061)
Observations	2542	2542	2542	2123	1593
Google					
Best	0.34*** (0.058)	0.58*** (0.067)	-0.13* (0.063)	-0.20+ (0.12)	-0.19* (0.094)
Neither best/ worst	0.16** (0.049)	0.33*** (0.052)	-0.021 (0.052)	-0.14 (0.095)	-0.049 (0.084)
Constant	2.75*** (0.046)	3.09*** (0.050)	15.5*** (0.050)	15.0*** (0.095)	1.45*** (0.084)
Observations	2,775	2,775	2,775	2,231	1,607

+ $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Source. Authors' analysis of data from 2016 Hospital Compare database and reviews posted to Facebook, Google, and Yelp collected during July, 2017.

all hospitals were considered "best." We then presented an additional set of figures where we excluded these markets where all hospitals were tied on a given measure, and here we again found that patient experience scores were highly correlated with crowdsourced star ratings, but clinical measures were not. Finally, we considered consistency of rankings when looking across crowdsourced sites. The more crowdsourced sites on which a hospital was ranked best the more likely it was to be best in terms of patient satisfaction scores. However, there was still very little correlation between the number of sites a hospital is ranked best on and clinical quality measures.

As an extension of the main results, we examined characteristics of hospitals labeled as “best” on a crowdsourcing site but ranked “worst” on a Hospital Compare measure, relative to (1) hospitals ranked as best on both crowdsourcing site and HC; and (2) hospitals ranked as worst on both crowdsourcing site and HC (Appendix SA2). Across all measures, urban hospitals were more likely to be inconsistently ranked as best on a crowdsourcing site and worst in terms of a HC measure (mismatched), relative to hospitals ranked consistently as worst or best across both a crowdsourcing site and HC. Large teaching hospitals were more likely to be mismatched in terms of HC’s overall hospital and patient experience rating, relative to hospitals that ranked consistently as best or worst between crowdsourcing sites and HC. In terms of patient safety (colon infections) and clinical quality (readmissions and mortality), small, nonteaching hospitals were more likely to be mismatched, relative to hospitals ranked consistently between crowdsourcing sites and HC.

DISCUSSION/CONCLUSION

Our results implied that the “best”-rated hospitals on crowdsourcing sites were more often than not also the “best” hospital according to HC patient experience; although, in about 20 percent of markets, they were the worst, according to HC patient experience. That being said, the correlation between crowdsourced site ratings and HC patient experience scores was much stronger than the correlation between crowdsourced site ratings and clinical quality measures. If patients chose hospitals ranked as the best in their market based on crowdsourced ratings, these hospitals did have slightly higher metrics of clinical quality and patient safety than the hospitals ranked worst in their market, but were not necessarily the best option available. These findings suggest that crowdsourced ratings focused on aspects of hospital care that do not reflect risk-adjusted clinical quality or patient safety measures. We also found Yelp’s ratings are the least correlated with HC ratings. This may be driven by Yelp’s proprietary filtering algorithm, sample size, and rounded scores, but we are unable to definitively determine the cause of this difference.

The weaker correlation between crowdsourced ratings and clinical measures could be due to a number of factors to be explored in future research. First, selection bias could lead patients (or family members of patients) who experienced an adverse outcome to be less likely to review hospitals online. Such bias would inflate crowdsourced rating averages for hospitals that perform worse on clinical quality. Another factor could be that clinical quality

may be less salient to online reviewers because of other aspects of patient experience. In previous research (Domino et al. 2014), the likelihood of medical malpractice suits, an extreme marker of perceived clinical incompetence, was mitigated by a positive patient experience. Finally, individual measures of clinical quality were likely noisier than patient experience. In fact, we found that the various measures of clinical quality available through HC are not strongly correlated with each other.

One limitation of the study design was that we did not observe actual patient choice or actual information considered by patients. Star ratings across sites were similar; the standard deviation of a hospital's star rating across sites was low (0.75). However, we did find that hospitals ranked as best on multiple sites were significantly more likely to be ranked best in terms of HC overall ratings and patient experience. Relatedly, we were unable to observe or control for other sources of information (such as recommendations from personal acquaintances) or factors that may affect patient choice (physician affiliation or insurance coverage) to the extent that these factors vary within HRRs.

Further, one possible concern was that the voluntary nature of social media resulted in participation among only high-achieving hospitals. The data presented here indicated that social media participation was much higher than in previous studies, improving prospects for generalizable findings. The distribution of scores across sites was generally bell-shaped; however, Facebook's rule permitting hospitals to disable the star rating feature on their site appears to have resulted in a positively skewed distribution as seen in Figure 1. We presented a balance table in Table 1 to show the limited differences between hospitals with crowdsourced ratings and hospitals without these ratings.

This was a descriptive study of cross-sectional data. The objective was not to establish a single source of hospital quality; rather, it was to compare the differences in decisions that might result from consulting crowdsourcing sites and HC. Our findings have implications for the salience of information about hospital quality that was available online. For patients who wish to choose the best hospital in their market in terms of patient experience (or to merely avoid the worst hospital in their market), they would likely choose (or avoid) similar hospitals, regardless of whether they seek information from a crowdsourcing site or HCAHPS. However, crowdsourcing ratings did not appear to signal strongly for other metrics of quality addressed in HC, such as patient safety. Taken together with previous evidence that patient narratives can distract consumers from quantified measures of clinical quality (Kanouse et al. 2016), our results suggest a word of warning to consumers. While

crowdsourced sites may provide similar information to patient experience surveys, patients should be encouraged to seek out other sources of information about clinical quality and not focus on the simpler aggregate five-star ratings. This work demonstrates to providers that reviewers on crowdsourcing sites prioritize and rate providers on nonclinical quality. Hospitals thus have an incentive to invest in nonclinical quality and disinvest in clinical quality if they want to appeal to patients online. Therefore, there may be a need to better communicate clinical quality to patients. For example, if HC conveyed a synthesized measure of clinical quality, then it would fill an information gap that is not provided by crowdsourcing sites.

HC's overall rating is a synthesis of up to 57 metrics of patient experience and clinical quality. For patients interested in patient experience, HCAHPS provides a separate score to summarize the 11 metrics in this field. However, if patients wish to prioritize based on clinical quality or patient safety, our findings indicated that crowdsourced ratings were a poor proxy for these dimensions of quality. The only alternative in this case was for patients to individually determine how to weigh the remaining 46 measures. To the extent that some metrics are specific to conditions for which patients may be seeking care, then the individual metrics are of themselves useful. However, for patients broadly interested in dimensions of clinical quality and safety, the findings suggested that a synthesized summary score of these dimensions would be valuable for consumers because these attributes of hospitals were not reflected in crowdsourced ratings.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors gratefully acknowledge institutional support from Indiana University. We also thank Paramdeep Singh for excellent research assistantship. All errors are our own.

Disclosures: None.

Disclaimer: None.

NOTE

1. We considered whether the ability of hospitals to suppress their total crowdsourced rating on Facebook resulted in significantly different regression results from Google or Yelp via a series of Wald tests. Based on these tests, the regression results comparing Hospital Compare to either Facebook or to

Google were not statistically different from each other. However, the regression results comparing Hospital Compare to Yelp differed significantly. The difference was likely due to the coarseness of the Yelp's star rating system: We observed more tied cases of hospital quality because ratings are only reported in units of half-stars, as well as Yelp's use of a algorithm to filter potentially false reviews. We considered whether the ability of hospitals to suppress their total crowd-sourced rating on Facebook resulted in significantly different regression results from Google or Yelp via a series of Wald tests. Based on these tests, the regression results comparing Hospital Compare to either Facebook or to Google were not statistically different from each other. However, the regression results comparing Hospital Compare to Yelp differed significantly. The difference was likely due to the coarseness of the Yelp's star rating system: We observed more tied cases of hospital quality because ratings are only reported in units of half-stars, as well as Yelp's use of a algorithm to filter potentially false reviews.

REFERENCES

- American Hospital Association. 2016. "CMS Releases Hospital Star Ratings." Online Press Release [accessed on August 23, 2017]. Available at <https://www.aha.org/news/headline/2016-07-27-cms-releases-hospital-star-ratings>
- Bardach, N. S., R. Asteria-Peñaloza, W. J. Boscardin, and R. A. Dudley. 2012. "The Relationship between Commercial Website Ratings and Traditional Hospital Performance Measures in the USA." *BMJ Quality & Safety* 22: 194–202.
- Campbell, L., and Y. Li. 2017. "Are Facebook User Ratings Associated with Hospital Cost, Quality and Patient Satisfaction? A Cross-Sectional Analysis of Hospitals in New York State." *BMJ Quality & Safety* 27: 119–29.
- Domino, J., C. McGovern, K. W. Chang, N. E. Carlozzi, and L. J. Yang. 2014. "Lack of Physician-Patient Communication as a Key Factor Associated with Malpractice Litigation in Neonatal Brachial Plexus Palsy." *Journal of Neurosurgery: Pediatrics* 13 (2): 238–42.
- Emmert, M., and M. Schlesinger. 2017. "Hospital Quality Reporting in the United States: Does Report Card Design and Incorporation of Patient Narrative Comments Affect Hospital Choice?" *Health Services Research* 52 (3): 933–58.
- Grob, R., M. Schlesinger, A. M. Parker, D. Shaller, L. R. Barre, S. C. Martino, M. L. Finucane, L. Rybowski, and J. L. Cerully. 2016. "Breaking Narrative Ground: Innovative Methods for Rigorously Eliciting and Assessing Patient Narratives." *Health Services Research* 51 (S2): 1248–72.
- Hefelee, J. G., Y. Li, L. Campbell, A. Barooah, and J. Wang. 2017. "Nursing Home Facebook Reviews: Who Has Them, and How do They Relate to Other Measures of Quality and Experience?" *BMJ Quality & Safety* 27: 130–9.
- Kamerer, D. 2014. "Understanding the Yelp Review Filter: An Exploratory Study." *First Monday* 19 (9). <https://doi.org/10.5210/fm.v19i9.5436>

- Kanouse, D. E., M. Schlesinger, D. Shaller, S. C. Martino, and L. Rybowski. 2016. "How Patient Comments Affect Consumers' Use of Physician Performance Measures." *Medical Care* 54 (1): 24–31.
- Luca, M., and G. Zervas. 2016. "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud." *Management Science* 62 (12): 3412–27.
- Ranard, B. L., R. M. Werner, T. Antanavicius, H. A. Schwartz, R. J. Smith, Z. F. Meisel, D. A. Asch, L. H. Ungar, and R. M. Merchant. 2016. "Yelp Reviews of Hospital Care can Supplement and Inform Traditional Surveys of the Patient Experience of Care." *Health Affairs* 35 (4): 697–705.
- Rau, J. 2015. "Only 251 Hospitals Score Five Stars in Medicare's New Ratings." *Kaiser Health News*.
- Trzeciak, S., J. P. Gaughan, J. Bosire, and A. J. Mazzairelli. 2016. "Association between Medicare Summary Star Ratings for Patient Experience and Clinical Outcomes in US Hospitals." *Journal of Patient Experience* 3 (1): 6–9.
- Wang, D. E., Y. Tsugawa, J. F. Figueroa, and A. K. Jha. 2016. "Association between the Centers for Medicare and Medicaid Services Hospital Star Rating and Patient Outcomes." *JAMA Internal Medicine* 176 (6): 848–50.
- Weinberg, E. 2017. "Facebook Reviews vs. Google & Yelp: Who's Dominating in 2017." *Fead Media*.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix SA1: Author Matrix.

Appendix SA2: Supplemental Analysis.