The relationship between commercial website ratings and traditional hospital performance measures in the USA

Naomi S Bardach,[1,2] Renée Asteria-Peñaloza,[2] W John Boscardin,[3] R Adams Dudley[2]

[1]Department of Pediatrics, University of California San Francisco, San Francisco, California, USA
[2]Philip R. Lee Institute for Health Policy Studies, University of California San Francisco, San Francisco, California, USA
[3]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

Correspondence to
Naomi S Bardach, Department of Pediatrics, University of California San Francisco, 3333 California St. Suite 265, San Francisco, CA 94118, USA; bardachn@peds.ucsf.edu

Editor's choice
Scan to access more free content

ABSTRACT

Background Our goal was to compare hospital scores from the most widely used commercial website in the USA to hospital scores from more systematic measures of patient experience and outcomes, and to assess what drives variation in the commercial website scores.

Methods For a national sample of US hospitals, we compared scores on Yelp.com, which aggregates website visitor ratings (1–5 stars), with traditional measures of hospital quality. We calculated correlations between hospital Yelp scores and the following: hospital percent high ratings (9 or 10, scale 0–10) on the 'Overall' item on the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey; hospital individual HCAHPS domain scores (eg, nurse communication, pain control); hospital 30-day mortality; and hospital 30-day readmission rates.

Results Of hospitals reporting HCAHPS (n=3796), 962 (25%) had scores on Yelp. Among hospitals with >5 Yelp ratings, the correlation of percent high ratings between Yelp and HCAHPS was 0.49 (p<0.001). The percent high ratings within each HCAHPS domain increased monotonically with increasing Yelp scores (p≤0.001 for all domains). Percent high ratings in Yelp and HCAHPS were statistically significantly correlated with lower mortality for myocardial infarction (MI; −0.19 for Yelp and −0.13 for HCAHPS) and pneumonia (−0.14 and −0.18), and fewer readmissions for MI (−0.17 and −0.39), heart failure (−0.31 and −0.39), and pneumonia (−0.18 and −0.27).

Conclusions These data suggest that rater experiences for Yelp and HCAHPS may be similar, and that consumers posting ratings on Yelp may observe aspects of care related to important patient outcomes.

INTRODUCTION

Commercial websites that post consumer ratings of businesses are increasingly popular. They have begun including ratings of physicians and hospitals, and the number of healthcare ratings grew dramatically from year 2005 to 2010.[1] These websites generally have online visitors give a numerical rating and then report the average of the individual visitors' ratings as the hospital score.[2] Concerns about these ratings include selection bias of raters and the raters' lack of medical expertise to judge the quality of care being delivered.[3 4]

The commercial sites' scores are salient to clinicians and hospital leaders because these sites have high consumer traffic. For instance, one popular commercial site, Yelp.com, that has a mostly US presence (though it has expanded to cities internationally),[5] reports more than 50 million unique visitors each month, and has 20 million user ratings available on its site.[6–8] The vast majority of these visitors are seeking information about other types of businesses or services, but consumer awareness of these sites is clearly high, with Yelp ranked in the top 50 most popular websites in the USA.[9] By contrast, only 6% of Americans have heard of the US Centers for Medicaid and Medicare Services' (CMS) national hospital public reporting website, hospitalcompare.hhs.gov.[10] The commercial sites' scores are also salient because consumers are actively using this type of online media to inform their healthcare choices. A recent survey found that 42% of US consumers have used social media

to access health-related consumer reviews of treatments or physicians, with 41% saying that information found via social media would affect their hospital choice.[11]

Given the use of these kinds of data by consumers, it will likely be important to hospitals and clinicians to understand how many hospitals are receiving consumer ratings on commercial websites, how much variation there is in the scores that websites assign to hospitals based on consumer ratings, and whether those scores in any way reflect on clinical care. Since the scores on the commercial websites are generated by consumers, the most relevant comparator in the USA is scoring from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). In the USA, HCAHPS is widely considered the industry standard for assessing patient experience, and is a metric in the new value-based purchasing programme of the US Center for Medicare and Medicaid Services.[12] Prior research has shown that overall HCAHPS scores are associated with performance in certain domains of patient experience and some processes of care.[13][14] HCAHPS scores are derived from survey data from randomly selected patients, which is usually gathered through modes that do not require internet access, such as a telephone or mailed surveys. The HCAHPS results are then risk-adjusted using a published model.[15][16] This process is highly standardised and well validated, while the approaches used by commercial websites are neither. Given the differences in commercial website and HCAHPS methodologies, it is possible that the scores are driven by different domains of care or that the commercial website scores are not associated with HCAHPS or with other previously validated quality indicators.

Another approach to evaluating the commercial site scores is to assess their relationship with clinical outcomes of care. For instance, prior studies of online ratings from the UK suggest that the opinions of consumers who choose to visit a website and leave a rating may be correlated with readmissions, mortality and infections.[17][18] However, these studies examined the website of the National Health Services (NHS), the national government healthcare agency in the UK, which may attract very different types of raters. The NHS website also solicits ratings using a structured feedback instrument, so these prior findings may not generalise to unprompted raters on commercial sites.

In order to better understand the quality of the information available to consumers using commercial site ratings, we systematically searched for commercial websites scoring hospitals nationally. For the site with the most hospitals scored, we assessed the number of hospitals with ratings and the dates of ratings, and assessed the correlation between these hospitals' scores and important traditional measures of quality: the HCAHPS survey; 30-day mortality rates; and 30-day readmission rates. To better understand what drives consumer visitor scores, we also analysed the relationship between the commercial site's scores and individual HCAHPS domains (eg, nurse communication or pain control).

METHODS
We use 'rating' to refer to an individual patient's response to HCAHPS or an individual consumer's post on a commercial website, and 'score' to refer to measures that summarise the individual ratings at the hospital level.

Data
We used publicly available hospital quality scores, accessible to consumers trying to learn about hospitals.

Hospital sample
All hospitals which had HCAHPS, mortality and readmissions data on the CMS website were included.

Commercial websites with publicly available ratings
Since there are multiple commercial websites offering provider scores,[2] we first determined which sites were likely to be found by consumers searching for hospital information (see online supplementary eMethods 1). We only looked at sites rating US hospitals, since the performance on the standardised patient experience measure and the outcomes measures was uniformly available only for a national set of US hospitals. We excluded 'members-only' sites requiring a subscription fee, such as Angie's List, because this membership requirement may exacerbate the selection bias and greatly limits who can post and use ratings. We therefore used Yelp.com data for the study as the only free, publicly available commercial website with scores on multiple hospitals in multiple US states (see online supplementary eMethods 1).

Yelp visitor demographics and scored businesses
Yelp's user demographics from 2011 are reported in an online network analysis of social media sites: 55% female, proportions of users by age are: ~75%, 25–54 years ~17%, 55–65+ years and ~8%, 0–25 years.[19] There are no data available regarding Yelp visitors' use patterns. Yelp reports numbers of businesses scored in individual cities by 22 categories, including restaurants, shopping and automotive, as well as hospitals and doctors (see online supplementary eTable 1).[20] For all categories, Yelp reports that as of the second quarter of 2012, 56% of ratings on the site were 4 or 5 stars, 14% 3 stars, and 20% 1 or 2 stars.[21]

Yelp data
Yelp allows visitors to rate hospitals from 1 to 5 stars, then aggregates the ratings and reports the aggregate score. Yelp uses an automated filter to exclude reviews that are deemed malicious or self-promoting, and

ratings are included only from visitors who are frequent reviewers.[22] The exclusion and inclusion algorithm for ratings, including the definition of 'frequent reviewer,' is proprietary.

We limited our data gathering to 1 week in time, starting 21 March 2011, and included in the analysis all ratings posted prior to that week. This cross-sectional approach mimics the approach a consumer is likely to take: looking for information about hospitals once at a particular point in time, rather than looking at some hospitals at one time point and other hospitals at another time point and synthesising the information over weeks to months.

We gathered the following data elements for each hospital: whether it had any Yelp ratings, the reported summary star scores for the hospital (integer and half integer values, rounded up or down from the exact average), the frequency each star rating was given, and rating dates. Since we could find no prior reports of the inter-rater reliability of collection of Yelp scores, two individuals independently collected the online Yelp data for each hospital, using the procedures described in online supplementary eMethods 2. The kappa statistic for agreement on the Yelp score for individual hospitals was 0.98. Discrepancies on any data point were reconciled by one of the authors (RAP) (see online supplementary eMethods 2).

Hospital characteristics
We obtained hospital characteristics from the American Hospital Association survey, using variables previously shown to be associated with HCAHPS scores (table 1).[13] In our comparison of hospital characteristics for groups of hospitals according to number of Yelp ratings (none, 1–5, or >5) we only analysed hospitals with American Hospital Association data (table 1).

HCAHPS data and the domains of patient experience
CMS reports HCAHPS[23] results that include both overall assessment of the hospital and eight composite quality domains, which summarise a total of 27 HCAHPS survey items. The overall rating of the hospital is on a scale of 0 (worst) to 10 (best). This overall rating has been found to be highly correlated ($r=0.87$) with the other global item: whether the patient would recommend the hospital to family and friends (possible responses: definitely yes, probably yes, probably no, and definitely no).[13] Therefore, we focused on the overall numerical rating, which is more similar to the star rating method. Prior to

Table 1 Characteristics of hospitals with no, few or more than five Yelp visitor ratings

Characteristics	No Yelp visitor ratings*†	One to five Yelp visitor ratings*‡	More than five Yelp visitor ratings*
Size, n (%)			
1–99 beds	1302 (47.5)	112 (16.8)	17 (6.5)
100–299 beds	1039 (37.9)	326 (48.9)	102 (38.8)
≥300 beds	399 (14.6)	229 (34.3)	144 (54.8)
Region, n (%)			
West	316 (11.5)	202 (30.2)	159 (60.5)
New England	116 (4.2)	44 (6.6)	14 (5.3)
Mid Atlantic	271 (9.9)	77 (11.5)	25 (9.5)
South Atlantic	461 (16.8)	110 (16.5)	21 (8.0)
Central	1576 (57.5)	234 (35.1)	44 (16.7)
Profit Status, n (%)			
Not-for-profit, public	559 (20.4)	83 (12.4)	37 (14.1)
Not-for-profit, private	1662 (60.7)	454 (68.1)	194 (73.8)
Commercial	519 (18.9)	130 (19.5)	32 (12.2)
Teaching, n (%)§	108 (3.9)	93 (13.9)	70 (26.6)
Urban, n (%)	2464 (89.9)	640 (96.0)	263 (100.0)
Hospital expenses in 10s of thousands of dollars per inpatient day, mean (95% CI)	43.9 (37.2 to 50.7)	44.0 (31.0 to 57.8)	41.4 (38.2 to 44.6)
Nurses per 10,000 patient days, mean (95% CI)	72.4 (60.7 to 84.1)	59.5 (57.2 to 66.3)	57.8 (54.6 to 60.9)
Medicaid patients, mean % (95% CI)	16.6 (16.2 to 16.9)	18.1 (17.2 to 18.9)	18.6 (17.1 to 20.2)

*Only hospitals with American Hospital Association data on hospital characteristics were included. Subsequent analyses included all hospitals with Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) data. Total number of observations (n) in each group for this analysis: no Yelp visitors: 2740 of 2834 with HCAHPS data; 1–5 Yelp visitor ratings: 667 of 692; more than 5 Yelp visitor ratings: 263 of 270.
†p Values for comparisons between hospitals with more than 5 ratings and hospitals with no ratings were <0.001 except for hospital expenses (p=0.50), nurse staffing levels (p=0.018), and percent Medicaid patients (p=0.013).
‡p Values for comparisons between hospitals with more than 5 ratings and hospitals with 1–5 ratings were ≤0.001 except for profit status (p=0.029), hospital expenses (p=0.67), nurse staffing (p=0.65), and percent Medicaid patients (0.55).
§'Teaching hospitals' are members of the Council of Teaching Hospitals and Health Systems.

reporting, CMS adjusts scores for the mode of administration, as well as for eight factors related to the patient (eg, age, education, health status) in order to address non-response and other biases.[23] We used data reported on the CMS website in March 2011, reflecting survey responses and response rates for each hospital from October 2009 to September 2010.[24]

Outcome measures

We assessed the correlation between percent of respondents giving a hospital a high (or low) overall rating in HCAHPS versus on Yelp. For HCAHPS, CMS provides only the percent of respondents giving a low (0–6 out of 10), middle (7–8) or high (9–10) rating. Because of this, we could not calculate mean HCAHPS scores, so we did not use mean Yelp scores as an outcome. We used a development subsample of the hospitals to determine how to group the Yelp 1–5 star ratings into high and low categories that would be most similar to the HCAHPS groupings (see online supplementary eMethods 3). Based on these findings, we defined a low rating on Yelp as a 1-star or 2-star rating, and a high rating as a 4-star or 5-star rating. We then calculated the correlations between percent high ratings on HCAHPS and Yelp and the percent low ratings on HCAHPS and Yelp in the entire dataset.

For our primary analyses, we only included hospitals with a minimum number of Yelp ratings, just as most public reporting sites require a minimum number of eligible patients to report each measure.[25] Since there is no precedent for what that minimum number of Yelp ratings should be, we chose a primary requirement of more than 5 Yelp ratings. To determine whether correlations might strengthen as the number of Yelp ratings increased, we compared the correlations between Yelp and HCAHPS hospital scores at different levels of the number of Yelp ratings required (5 or fewer, more than 5, more than 10, more than 15).

Mortality and readmissions rates

We looked for evidence of similarity between Yelp and HCAHPS scores by assessing whether the magnitude and direction of the correlations between percent high ratings on Yelp and hospitals' risk-standardised 30-day mortality and readmission rates for myocardial infarction (MI), heart failure and pneumonia as reported on the CMS website in March 2011[26] were similar to the magnitude and direction of correlations between those outcomes and percent high HCAHPS scores.

Statistical analyses

We used t tests assuming unequal variances and χ^2 tests to compare hospital characteristics between hospitals with and without Yelp ratings.

All correlations were assessed using Pearson's correlation coefficients. To address the impact of varying numbers of Yelp raters per hospital, we calculated weighted correlation coefficients with weights inversely proportional to the number of Yelp raters.

We evaluated the relationship between summary star score on Yelp and HCAHPS domain items using a test of trend of mean percent of positive answers ('always' had a positive experience) for each domain over 4-star score categories: 1–1.5 stars, 2–2.5, 3–3.5 or 4–5 (since there were only two hospitals with star scores of 5, these were grouped with 4–4.5, instead of being evaluated separately).[27] We chose to analyse these groupings instead of grouping by quartile of exact average star scores for hospitals, since the summary star scores are the data available to providers and consumers when reviewing the sites, and so the star groupings are more clinically relevant. Analyses were performed using STATA V.12 (Stata Corp). The study was considered exempt by the committee on human research at the University of California San Francisco.

RESULTS

There were 3796 hospitals with HCAHPS, mortality and readmissions data available on the CMS website (HospitalCompare.hhs.gov), of which 25% (n=962) had at least one Yelp rating. The mean response rate to the HCAHPS survey was 33%. For hospitals with any ratings available on Yelp, the mean number of consumer ratings per hospital was 6.5 (95% CI 5.7 to 7.3; median 2; range: 1–159; total ratings=6260). For hospitals with more than five ratings on Yelp (n=270), the mean percent of high ratings on HCAHPS and Yelp were 65% (95% CI 64% to 66%) and 56% (95% CI 53% to 58%), respectively. The mean star score was 3.3 stars, and 74% had scores of 3 stars or better.

The number of ratings was increasing over time: 22% of all study ratings were from 2009 and 34% from 2010. Posted ratings were recent, with 60% of hospitals having their most recent rating dated in 2011 or in the last quarter of 2010.

Characteristics of the hospital sample

Table 1 shows that hospitals that had more than five ratings on Yelp were more frequently large, Western, non-profit, academic and urban compared with those with no Yelp ratings (and hence no Yelp score). Hospitals with more than five Yelp ratings also had lower nurse staffing and a higher percent Medicaid patients compared with unrated hospitals.

Relationships between Yelp scores and HCAHPS overall scores

Among hospitals with more than five Yelp ratings, the correlation between the percent high ratings from HCAHPS and from Yelp was 0.49 (figure 1), while the correlation between the percent low ratings for each system was 0.47. The correlations varied with the number of individual ratings on Yelp, increasing
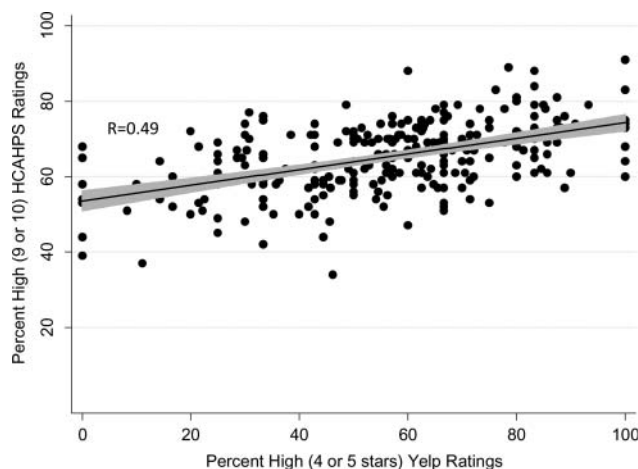
Figure 1 Relationship between percent of high ratings on Yelp and Hospital Consumer Assessment of Healthcare Providers and Systems. Hospitals with more than 5 Yelp ratings (n=270) were included. The line-of-fit (solid black line) and 95% confidence interval (grey shading) is portrayed. R, Pearson's product moment correlation coefficient, used weights inversely proportional to the number of Yelp raters. p Value for correlation is <0.001.

when hospitals with five or fewer ratings were excluded, and then changing only slightly when excluding hospitals with fewer than 11 or 16 ratings (table 2).

Relationships with outcomes

The percent high Yelp ratings was correlated with lower readmission rates for all conditions and with lower mortality for MI and pneumonia. The direction and magnitude of these correlations were similar to correlations with the percent high HCAHPS ratings (table 3).

Potential explanatory factors for Yelp scores

We stratified hospitals by star score (table 4) to assess for differences in performance within individual HCAHPS domains. In all cases, increasing Yelp scores were associated with consistently increasing HCAHPS domain scores, with $p \leq 0.001$ in every domain (table 4). In terms of overall HCAHPS ratings, hospitals in the top category of star scores had 18% more of 9 or 10 ratings than hospitals in the bottom category. Domains in which hospitals in the top star score

category had more than 10% better performance than hospitals in the bottom category were communication with nurses, availability of help and pain control.

DISCUSSION

This study provides a portrait of consumers' ratings of hospitals on a widely used commercial website. We found that many hospitals have already been rated, and that the number of hospital ratings increased over time, though the average number of ratings per hospital was still low. While most hospitals scored 3 stars or better, some did not. This is likely to be important to the hospitals and to the people who work there, since the potential audience for these ratings is large, potentially larger than the audience at CMS' website and other non-profit public reports.[10] [28]

The correlation we found between Yelp star scores and HCAHPS overall scores (0.49) is unexpectedly high; 0.50 is often cited in interpretation of correlations in the social sciences as the cut-off between 'moderate' and 'strong' correlation, although there is no universal consensus about interpreting correlations.[29] We had hypothesised that there would be

Table 2 Correlations between percent of high and low ratings on Yelp and Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)

Correlations between HCAHPS and Yelp on:	Hospitals with:			
	At least one visitor rating on Yelp (n=962)	More than 5 visitor ratings on Yelp (n=270)	More than 10 visitor ratings on Yelp (n=148)	More than 15 visitor ratings on Yelp (n=102)
Percent high ratings*	0.40	0.49	0.49	0.48
Percent low ratings†	0.40	0.47	0.48	0.47

Weighted Pearson's product moment correlation coefficients, using weights inversely proportional to the number of Yelp raters. p Values for all correlations <0.001.
*The percent high rating for HCAHPS refers to the percent of respondents giving the hospital a overall rating of 9 or 10 on a scale of 0–10. The percent high rating for Yelp refers to the percent of respondents giving the hospital an overall rating of 4 or 5 stars.
†The percent low rating for HCAHPS refers to the percent of respondents giving the hospital an overall rating of below 7. The percent low rating for Yelp refers to the percent of respondents giving the hospital an overall rating of 1 or 2 stars.

Table 3 Correlations between clinical outcomes within 30-days and average Yelp score

Outcome	Correlation with percent high Yelp score	p Value	Correlation with percent high ratings on HCAHPS	p Value
Mortality rate, MI	−0.19	0.005	−0.13	0.052
Mortality rate, Heart Failure	−0.01	0.89	0.01	0.85
Mortality rate, Pneumonia	−0.14	0.028	−0.18	0.004
Readmissions, MI	−0.17	0.011	−0.39	<0.001
Readmissions, Heart Failure	−0.31	<0.001	−0.39	<0.001
Readmissions, Pneumonia	−0.18	0.005	−0.27	<0.001

Hospitals with >5 Yelp ratings (n=270) were analysed. Weighted Pearson's product moment correlation coefficients, using weights inversely proportional to the number of Yelp raters.
Data for 30-day mortality and readmission rates come from the Centre for Medicaid and Medicare Services public reporting website, HospitalCompare.hhs.gov.
HCAHPS, Hospital Consumer Assessment of Healthcare Providers and Systems; MI, myocardial infarction.

only a weak relationship between star scores and HCAHPS scores for several reasons. First, the approaches differ in sampling, with the commercial website using a self-selected sample of visitors who might be expected to have more extreme views than randomly solicited HCAHPS respondents. Second, the consumers providing the ratings may differ in case mix (ie, proportion of medical, surgical or obstetrical hospitalisations), a factor that is adjusted for in the HCAHPS scores but not the star scores.[15 16] In addition, internet users tend to be younger than the general population,[30 31] while respondents to phone and mailed surveys, such as HCAHPS, tend to be older,[32] and so the Yelp raters may be younger than the HCAHPS respondents. These differences between the two populations may have confounded the relationship between HCAHPS and Yelp scores. For instance, HCAHPS adjusts for age, since it has been shown that younger patients give lower ratings.[15] By contrast, commercial websites do not, which may explain the lower percent of high ratings on the commercial website than HCAHPS (56% vs 65%).

Though these differences in methods between the two approaches made the correlation unexpected, our other analyses supported the finding, including the consistent pattern that better performance within each HCAHPS domain was associated with higher Yelp scores. Yelp raters are not prompted to reflect on any of those domains specifically, so this robust pattern suggests that these domains are likely important to Yelp raters in a way that is similar to HCAHPS respondents. The HCAHPS domains that vary most with Yelp scores (nurse communication, availability of help and pain control) have previously been shown to be domains that are strongly associated with high overall HCAHPS ratings.[14] Increasing the potential significance of these findings for hospitals is the 18% difference in percent high HCAHPS ratings between the top and bottom categories of hospitals by star scores. This difference in HCAHPS scores is a much larger difference than differences associated with other hospital factors, such as nurse staffing, profit status, census region and percent Medicaid patients (range of associated performance differences: 0–7.3%).[13]

Table 4 Relationships between Yelp score and Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) ratings and individual HCAHPS domains

Overall star score category	1 or 1.5 stars (n=11), %	2 or 2.5 stars (n=47), %	3 or 3.5 stars (n=131), %	4, 4.5, or 5 stars (n=81), %	p Value for trend
Percent 9 or 10 overall HCAHPS rating	52	61	65	70	p<0.001
Communication with nurses*	62	68	70	74	p<0.001
Pain control*	57	64	66	69	p<0.001
Help available*	46	54	56	59	p<0.001
Communication with doctors*	71	76	77	78	p<0.001
Medications explained*	52	54	57	58	p<0.001
Discharge instructions*	76	79	80	82	p=0.001
Clean room*	60	64	65	68	p<0.001
Quiet room*	45	49	49	53	p<0.001

Hospitals with >5 Yelp ratings (n=270) were analysed.
*Percents are the mean percentages of patients who 'always' had a positive experience with respect to the individual domain item of the Hospital Consumer Assessment of Healthcare Providers and Systems survey listed. A test of trend was performed over the 4 Overall Star Score categories for each individual domain.[27] 'Help available' refers to whether help was available when needed; 'Discharge Instructions' refers to the quality of discharge instructions.[23]

In addition, better scores in Yelp are correlated with lower mortality rates for MI and pneumonia, and lower readmission rates for multiple conditions, with similar direction and magnitudes as HCAHPS correlations with these outcomes. These findings reinforce the early observations from the UK,[17] and imply that consumers posting ratings on commercial websites may be observing meaningful aspects of hospital quality of care. The −0.19 correlation between Yelp scores and mortality for MI approaches the −0.25 correlation previously observed between a composite measure of technical processes of MI care and mortality.[33] Though many technical processes of care have a strong evidence base that they can improve outcomes, and patient ratings have not correlated in some settings with technical aspects of care,[34] it may be that patients observe additional elements of care delivery that are less technical but still important.

Although correlations do not prove causation, taken together, the data suggest that the ratings posted on the commercial website may be capturing experiences similar to those driving the more systematically collected HCAHPS ratings, and that improvements in either patient experience measure may be associated with improvements in patient outcomes. Assuming that there are differences in rating populations for each measure, the associations with improvements in patient outcomes may constitute a reason for hospitals to attend to star scores, since they may reflect the experience of a population whose input they receive less of through other means.

A recent study of Yelp data from the restaurant sector by Anderson and Magruder suggests that this source of data has the capacity to change consumer behaviour. Using regression discontinuity estimates around star reporting cut-points, they found that an extra half-star rating causes restaurants to sell out 19 percentage points more frequently (30% vs 49%), with larger impacts when alternate information is more scarce.[35] Of course, consumers have more choice among restaurants than hospitals and can switch restaurants more easily than hospitals, so it is unclear how much this behavioural finding about restaurants will occur with hospital ratings. However, Anderson and Magruder's findings do suggest that voluntary online ratings may warrant the attention of hospital leadership and individual clinicians on the possibility that they may change consumer choice behaviour.[36 37]

Several limitations deserve mention. Our findings may not be generalisable to hospitals that were not rated, which tend to be smaller, non-teaching hospitals in the middle of the USA. Further research will be needed, in the event that these hospitals become rated, to determine whether the relationships we found persist. Second, there are other commercial websites that provide ratings which we did not analyse, but the website studied has the most site visitors currently, ranking within the top 50 most visited sites in the USA.[8 9 38] Because of the small number of reviews available on other websites at the time of our data collection, it was unlikely that we would have been able to conduct stratified analyses to better understand differential selection biases across websites, and so we chose to focus on one website. Third, though the Yelp audience is large,[6 7 9] we do not know how many people look at hospital ratings. However, we found increasing numbers of hospital reviews in the recent time period, which suggests that there is likely an increasing audience as well. In addition, the ratio of doctors to auto mechanics reviewed is approximately one or higher for five of the six major metropolitan areas we looked at (see online supplementary eTable 1), suggesting that consumer interest in healthcare provider ratings may be similar to interest in ratings of other common services. Therefore, for hospitals that are not yet rated on Yelp, this study suggests that the scores may be worthy of some form of attention as hospitals start to garner reviews either on Yelp or other similar websites. Finally, commercial websites can filter the ratings posted, usually with the primary intent of excluding malicious or self-promoting reviews and ratings, but such filters are not transparent. There have been claims in the popular press that businesses can manipulate their ratings,[6 39] although Yelp has refuted those claims.[6 22 40] If the filtering system was not functioning well, or was being manipulated, one would expect there to be more extreme and more positive views represented, which we did not find for hospitals with at least five reviews posted. Although the filtering system may be functioning as Yelp describes, this controversy highlights the importance of understanding whether the ratings posted seem to reflect the underlying performance as measured by validated, standardised methods such as HCAHPS.

In summary, this analysis of consumers' ratings of hospitals on a widely used, publicly available commercial website provides insight into how voluntary raters on this site view hospital performance. As the consumers providing these ratings may be a very different population from HCAHPS respondents, they may represent a distinct source of feedback for hospitals. Our analysis of their ratings suggests that these consumers may have an important perspective on variations in quality of inpatient care. Lastly, there is evidence from this data that performance on star scores varies, and our analysis identified domains of care that may drive these variations. For these reasons, hospital leaders and clinicians may want to learn how their own institutions are scored, and consider these scores as one useful input in their overall plans for quality improvement.

necessarily represent the views of the California HealthCare Foundation.

REFERENCES

1. Gao GG, McCullough JS, Agarwal R, *et al*. 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:e38.
2. Kadry B, Chu LF, Gammas D, *et al*. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *J Med Internet Res* 2011;13:e95.
3. McCartney M. Will doctor rating sites improve the quality of care? No. *BMJ* 2009;338:b1033.
4. Lagu T, Lindenauer PK. Putting the public back in public reporting of health care quality. *JAMA* 2010;304:1711–12.
5. Browse Business Reviews by City. 2012. http://www.yelp.com/locations (accessed 5 Jul 2012).
6. Smith R, Fowler GA. Rate this: Yelp IPO targets $2 billion. *The Wall Street Journal* November 9, 2011.
7. Tsotsis A. Yelp files for IPO to raise $100 million. *TechCrunch* November 17, 2011.
8. Site Analytics for Angieslist.com. http://siteanalytics.compete.com/angieslist.com/ (accessed 5 Apr 2012).
9. Site Analytics for Yelp.com. http://siteanalytics.compete.com/yelp.com/ (accessed 5 Apr 2012).
10. 2008 Update on Consumers' Views of Patient Safety and Quality Information. 2008. http://www.kff.org/kaiserpolls/upload/7819.pdf (accessed 28 Jan 2012).
11. Social media 'likes' healthcare: from marketing to social business. April 2012. http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml (accessed 1 June 2012).
12. HCAHPS: Patient's Perspectives of Care Survey | US Center for Medicaid and Medicare Services. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS.html (accessed 5 Sep 2012).
13. Jha AK, Orav EJ, Zheng J, *et al*. Patients' perception of hospital care in the United States. *N Engl J Med* 2008;359:1921–31.
14. Rothman AA, Park H, Hays RD, *et al*. Can additional patient experience items improve the reliability of and add new domains to the CAHPS hospital survey. *Health Serv Res* 2008;43:2201–22.
15. O'Malley AJ, Zaslavsky AM, Elliott MN, *et al*. Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv Res* 2005;40(6 Pt 2):2162–81.
16. HCAHPS: Mode and Patient-Mix Adjustment Page. http://www.hcahpsonline.org/modeadjustment.aspx (accessed 28 Jan 2012).
17. Greaves F, Pape UJ, King D, *et al*. Associations between Web-based patient ratings and objective measures of hospital quality. *Arch Intern Med* 2012;172:435–6.
18. Greaves F, Pape UJ, King D, *et al*. Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Quality & Safety* 2012;7:600–5.
19. 2012 Social Network Analysis Report—Demographic—Geographic and Search Data Revealed. http://www.ignitesocialmedia.com/social-media-stats/2012-social-network-analysis-report/#yelp (accessed 12 Sep 2012).
20. Best Restaurants, Bars, Salons, Spas, and more! | Yelp. 2012. http://www.yelp.com/find (accessed 11 Sep 2012).
21. FAQ | Yelp Rating Distribution. 2012. http://www.yelp.com/faq#rating_distribution (accessed 10 Sep 2012).
22. Yelp.com Frequently Asked Questions. 2011. http://www.yelp.com/faq (accessed 28 Sep 2011).
23. HCAHPS: Hospital Care Quality Information from the Consumer perspective. 2011. http://www.hcahpsonline.org (accessed 2 Sep 2011).
24. Medicare Hospital Compare Quality of Care. 2011. http://hospitalcompare.hhs.gov/ (accessed 1 Mar 2011).
25. The Joint Commission. Guidelines for Publicizing Hospital National Quality Improvement Goals. 2012. http://www.jointcommission.org/accreditation/guidelines_for_publicizing_nqigs.aspx. (accessed 28 Jan 2012).
26. Hospital Compare: A quality tool provided by Medicare. http://www.hospitalcompare.hhs.gov/ (accessed Mar 2011).
27. Vittinghoff E. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*. New York: Springer, 2005.
28. Bardach N, Hibbard JH, Dudley RA. *Users of Public Reports of Hospital Quality: Who, What, Why, and How? An aggregate analysis of 16 online public reporting Web sites and users' and experts' suggestions for improvement*. Rockville, MD 2011.
29. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd edn. Hillsdale, NJ: L. Erlbaum Associates, 1988.
30. Who's Online: Internet User Demographics: Pew Research Center's Internet and American Life Project. May 2011. http://pewinternet.org/Trend-Data/Whos-Online.aspx (accessed 29 Jan 2012).
31. Jones S. Generations Online in 2009. 2009. http://pewresearch.org/pubs/1093/generations-online (accessed 29 Jan 2012).

32 Elliott MN, Zaslavsky AM, Goldstein E, *et al*. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res* 2009;44(2p1):501–18.

33 Bradley EH, Herrin J, Elbel B, *et al*. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA* 2006;296:72–8.

34 Chang JT, Hays RD, Shekelle PG, *et al*. Patients' global ratings of their health care are not associated with the technical quality of their care. *Ann Intern Med* 2006;144:665–72.

35 Anderson M, Magruder J. Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database*. *Econ J* 2012;122:957–89.

36 Hibbard JH, Stockard J, Tusler M. It isn't just about choice: the potential of a public performance report to affect the public image of hospitals. *Med Care Res Rev* 2005;62:358–71.

37 Hibbard JH, Stockard J, Tusler M. Hospital performance reports: impact on quality, market share, and reputation. *Health Aff (Millwood)* 2005;24:1150–60.

38 Site Analytics for RateMDs.com. http://siteanalytics.compete.com/angieslist.com/ (accessed Apr 2012).

39 Richards K. Yelp and the Business of Extortion 2.0. *East Bay Express* February 18, 2009.

40 Weise K. A Lie Detector Test for Online Reviewers. *Bloomberg Business Week* September 29, 2011.

eMethods1

Since there are several commercial websites offering scores, we first determined which sites were likely to be found by consumers searching for hospital information. In the fall of 2010, using three popular search engines (Google, Yahoo, and Bing), one research associate searched for websites that provide hospital ratings of patient experience (search terms: "hospital comparison" "hospital quality" "best hospitals" "hospital comments"). Since our interest was in sites in which consumers self-select to post ratings, we excluded sites that only presented the CMS HCAHPS data (e.g., HealthGrades) or offered no patient experience ratings at all. While several other sites have consumer visitor physician ratings,[1,2] we found that there were few sites with consumer visitor hospital ratings. We therefore used Yelp.com data for the study as Yelp was the only commercial website with consumer visitor ratings that had publicly available scores on multiple hospitals in multiple states. After our online search and after data collection was complete, a study from November 2011 used a different method to find provider rating websites.[1] We reviewed their list of 10 websites and confirmed that Yelp was the only website that was a consumer visitor rating website with publicly available scores on multiple hospitals in multiple states.

eMethods2

To find hospitals listed in the HCAHPS database in Yelp, we used the set of procedures listed below, which were initially piloted in small batches of 10-20 hospitals. The data was entered using an online form. For the initial search for the hospital in Yelp, we used Google.com since we found in our pilot searches that Google's search engine more effectively found the hospitals in Yelp than the Yelp search engine. The Search Field for each hospital was created by concatenating three fields: the name of the hospital in the HCAHPS database, the address for the hospital in the HCAHPS database, and a field containing: YELP.COM.

TASK INSTRUCTIONS:

1. Go to Google.com

2. Enter the Hospital information provided in the GOOGLE SEARCH FIELD WITHOUT QUOTATION MARKS. Here is an example of what you would enter in Google: ALAMEDA HOSPITAL 2070 CLINTON AVE ALAMEDA CA 94501 YELP.COM

3. Click Search and look through the Google results for the Hospital in Yelp. The URL for the Google result link should begin with www.yelp.com <http://www.yelp.com/> .

4. If you find a Yelp listing in Google, click on the URL. Note: You may have multiple Yelp results among the Google results for the Hospital. Click on the first result.

5. In the field below called STAR RATINGS GIVEN:

ENTER 0 If you find the hospital listed in Yelp with NO star ratings

ENTER 1 if there are ANY STAR RATINGS given for the hospital

NOTE: Only enter 1 or 0. We do not want to know what the ratings are, just whether you can find the hospital in Yelp and whether there are any ratings

6. In the field below called YELP HOSPITAL URL paste the URL (web address) of the Yelp Hospital result you based your STAR RATINGS GIVEN decision on . Here is an example of what you would enter: http://www.yelp.com/biz/alameda-hospital-alameda

7. If there are NO YELP.COM URLS in the first page of the Google results, go to Yelp.com and do the following:

8. In the Yelp website, enter the information provided in the YELP SEARCH FOR FIELD and the YELP NEAR FIELD WITHOUT QUOTATION MARKS. Here is an example of what you would enter in Yelp: "Search for": KAISER "Near": 3288 MOANALUA RD HONOLULU HI 96819

9. If your search in Yelp yields no results try again by using just the term: Hospital in the "Search for" field together with information provided in the YELP NEAR FIELD. Here is an example: "Search for": HOSPITAL "Near": 3288 MOANALUA RD HONOLULU HI 96819

10. After performing these searches in Yelp, in the field below called STAR RATINGS GIVEN

ENTER 0 If you find the hospital listed in Yelp with NO star ratings

ENTER 0 If you do not find the hospital in Yelp at all

ENTER 1 if there are ANY STAR RATINGS given for the hospital

NOTE: Only enter 1 or 0. We do not want to know what the ratings are, just whether you can find the hospital in Yelp and whether there are any ratings

11. In the field below called YELP HOSPITAL URL paste the URL (web address) of the Yelp Hospital result you based your STAR RATINGS GIVEN decision on. Here is an example of what you would enter: http://www.yelp.com/biz/kaiser-permanente-medical-care-program-honolulu-2

12. For Hospitals with STAR RATINGS GIVEN in the field below called TOTAL NUMBER OF REVIEWS enter the number of reviews next to the star rating as illustrated by the example below.

For those hospitals with ANY STAR RATINGS, the following further steps were followed.

1. In **"Overall Yelp Star Rating"**, write the overall rating given by Yelp for the hospital. This "Overall Yelp Star Rating" is illustrated by a number of stars and should be directly below the Hospital Name. Note: "Overall Yelp Star Rating" can ONLY take on the following values: 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, or 5.
2. In **"Total Number of Yelp Reviews"**, write the number of reviews reported by Yelp for the hospital. The number of reviews is reported to the right of the Overall Star Rating. Do not include filtered reviews.

Click on Rating Details (a small graphic right next to the Overall Star Rating and the Number of Reviews) and a window will pop-up with Rating Details,

3. In **"Number of 5 Star Reviews"**, write down the number of 5 Star Reviews provided in the Rating Details pop-up window.
4. In **"Number of 4 Star Reviews"**, write down the number of 4 Star Reviews provided in the Rating Details pop-up window.
5. In **"Number of 3 Star Reviews"**, write down the number of 3 Star Reviews provided in the Rating Details pop-up window.
6. In **"Number of 2 Star Reviews"**, write down the number of 2 Star Reviews provided in the Rating Details pop-up window.
7. In **"Number of 1 Star Reviews"**, write down the number of 1 Star Reviews provided in the Rating Details pop-up window.

Sort Yelp Reviews by DATE

8. In **"Most Recent Review Date"**, write down the month, day and year of the most recent review (the newest review) in this format mm/dd/yyyy. For example, February 17, 2010 would be written 02/17/2010. Note we are NOT interested in filtered reviews.
9. In **"Oldest Review Date"**, write down the month, day and year of the oldest review (the first ever review) in this format mm/dd/yyyy. For example, March 12, 2005 would be written 03/12/2005.

Scroll through all the Yelp Reviews (sorted by DATE)

10. In **"Number of Reviews 2010"**, count the number of reviews given between January 2010 and December 2010 and <u>write down the total number</u>. <u>We are NOT interested in filtered reviews.</u>
11. In **"Number of Reviews 2009"**, count the number of reviews given between January 2009 and December 2009 and <u>write down the total number</u>. <u>We are NOT interested in filtered reviews.</u>

A REMINDER OF THE IMPORTANT NOTES STATED ABOVE: When you visit the Yelp Hospital URL provided below Yelp reviews are not sorted by DATE but by "Yelp Sort". You need to sort the Yelp Reviews **by Date** in order to find the Most Recent Review Date and the Oldest Review Date. Also, Hospitals with <u>many reviews (>40) have multiple pages of reviews</u> in Yelp so you will have to <u>scroll to the bottom of the Yelp review page and click on the "next page"</u> in order to get the total Number of 2010 Reviews (Jan. 2010 - Dec. 2010) and the total Number of 2009 Reviews (Jan. 2009 to Dec. 2009). Also Yelp filters certain reviews. <u>Ignore all filtered reviews.</u>

eMethods3

For HCAHPS, CMS provides only the percent of respondents giving a low (0-6), middle (7-8), or high (9-10) rating. CMS uses only the percent high ratings as the score for the hospital. Because of this, we were not able to calculate mean HCAHPS ratings. The publicly available data on Yelp includes the star score (the approximate mean of all individual ratings) and the number of visitors who gave 1-, 2-, 3-, 4-, or 5-star ratings for each hospital. To compare the percent high ratings from HCAHPS and Yelp, we had to develop a system to group the Yelp star ratings into low, middle, and high groups. However, there is no prior research describing a grouping process of 5-star ratings that is analogous to CMS' groupings of the HCAHPS ratings on the 0-10 scale, nor has CMS published the rationale for its groupings. Therefore, to develop a system of grouping Yelp ratings, we randomly selected a subsample of 100 hospitals. We used this subsample to compare the percentage of high and low ratings on HCAHPS and Yelp using different approaches to grouping the Yelp ratings. We calculated the correlation coefficient and mean differences between the percent high HCAHPS overall ratings and percent high Yelp ratings, under three different definitions of "high" for the Yelp ratings: only 5 star ratings, 4-5 star ratings, and 3-5 star ratings. Similarly, we calculated the correlation coefficient and mean differences between the percent low HCAHPS overall ratings and percent low Yelp ratings, under three different definitions of "low" for the Yelp ratings: only 1 star ratings, 1-2 star ratings, and 1-3 star ratings. We then chose the grouping of Yelp scores that correlated most closely with the HCAHPS groups and had the smallest mean differences in the development sample and used that grouping method in subsequent analyses. The subsequent analyses were performed on the entire population of hospitals that had Yelp scores.

eTable1. Number of Businesses reviewed on Yelp.com by Category in Geographically Diverse Cities

	Doctors	Hospitals	Auto Repair	Doctors/ Auto Repair Shop*
Boston	291	53	295	0.99
Chicago	909	86	591	1.54
Houston	255	41	256	1.00
Saint Louis	84	14	127	0.66
Atlanta	457	27	506	0.90
Oakland	448	14	366	1.22

* Because there are no denominators available for these numbers, we showed the ratio of numbers of doctors reviewed to number of auto repair businesses reviewed to illustrate the variation by region in interest in healthcare providers. We showed the ratio with auto repair businesses because the type of information asymmetry is similar for auto mechanics and doctors, and the number of doctors per capita is more similar to auto repair businesses per capita than is the number of hospitals per capita.

References

1. Kadry B, Chu LF, 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.
2. Bardach N, Hibbard JH, Dudley RA. *Users of Public Reports of Hospital Quality: Who, What, Why, and How? An aggregate analysis of 16 online public reporting Web sites and users' and experts' suggestions for improvement. .* Rockville, MD2011.