

# Health Affairs

At the Intersection of Health, Health Care and Policy

Cite this article as:

Benjamin L. Ranard, Rachel M. Werner, Tadas Antanavicius, H. Andrew Schwartz, Robert J. Smith, Zachary F. Meisel, David A. Asch, Lyle H. Ungar and Raina M. Merchant

Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care

*Health Affairs* 35, no.4 (2016):697-705

doi: 10.1377/hlthaff.2015.1030

The online version of this article, along with updated information and services, is available at:

<http://content.healthaffairs.org/content/35/4/697>

**For Reprints, Links & Permissions :**

[http://content.healthaffairs.org/1340\\_reprints.php](http://content.healthaffairs.org/1340_reprints.php)

**Email Alertings :** <http://content.healthaffairs.org/subscriptions/etoc.dtl>

**To Subscribe :** <https://fulfillment.healthaffairs.org>

*Health Affairs* is published monthly by Project HOPE at 7500 Old Georgetown Road, Suite 600, Bethesda, MD 20814-6133. Copyright © by Project HOPE - The People-to-People Health Foundation. As provided by United States copyright law (Title 17, U.S. Code), no part of may be reproduced, displayed, or transmitted in any form or by any means, electronic or mechanical, including photocopying or by information storage or retrieval systems, without prior written permission from the Publisher. All rights reserved.

Not for commercial use or unauthorized distribution

By Benjamin L. Ranard, Rachel M. Werner, Tadas Antanavicius, H. Andrew Schwartz, Robert J. Smith, Zachary F. Meisel, David A. Asch, Lyle H. Ungar, and Raina M. Merchant

DOI: 10.1377/hlthaff.2015.1030  
HEALTH AFFAIRS 35,  
NO. 4 (2016): 697–705  
©2016 Project HOPE—  
The People-to-People Health  
Foundation, Inc.

# Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care

**ABSTRACT** Little is known about how real-time online rating platforms such as Yelp may complement the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, which is the US standard for evaluating patients' experiences after hospitalization. We compared the content of Yelp narrative reviews of hospitals to the topics in the HCAHPS survey, called domains in HCAHPS terminology. While the domains included in Yelp reviews covered the majority of HCAHPS domains, Yelp reviews covered an additional twelve domains not found in HCAHPS. The majority of Yelp topics that most strongly correlate with positive or negative reviews are not measured or reported by HCAHPS. The large collection of patient- and caregiver-centered experiences found on Yelp can be analyzed with natural language processing methods, identifying for policy makers the measures of hospital quality that matter most to patients and caregivers. The Yelp measures and analysis can also provide actionable feedback for hospitals.

Since 2006, patient-reported experiences after hospitalization have been collected using the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey.<sup>1,2</sup> HCAHPS survey results are publicly reported on the Centers for Medicare and Medicaid Services Hospital Compare website,<sup>1</sup> which rates all US hospitals that receive Medicare payments on a variety of quality measures.<sup>3</sup> HCAHPS survey scores now drive 25 percent of the financial incentives in the Medicare value-based purchasing program,<sup>4(p50701)</sup> which will eventually penalize hospitals with poor performance by up to 2 percent of their Medicare payments.<sup>4(p51023)</sup> The HCAHPS survey is the current standard for patient-experience-of-care data,<sup>1</sup> but its development dates back to 2002.<sup>2</sup> In the fourteen years since the survey first appeared, the indications for and experiences of hospitalization have changed greatly. Perhaps more importantly,

more than a decade ago patients were not spontaneously publishing their opinions about health care facilities on social media sites where opinions reach the members of the public, who are increasingly comfortable in using them to inform their own decisions.

Evaluations such as the HCAHPS survey are the products of years of measurement research, are fielded and interpreted systematically,<sup>1,2</sup> and have collected a large number of patient responses per hospital.<sup>1</sup> However, they are expensive to deploy,<sup>5</sup> they suffer from low response rates,<sup>6</sup> and there may be significant delays between hospitalization and public reporting of results.<sup>7</sup> Even if the evaluations can give an overall indication of patient satisfaction, they rarely identify the source of perceived problems.<sup>8</sup>

Reviews on social media sites are organic, largely unstructured, and essentially uncurated but are both seemingly haphazard and subject to gaming. Yet the testimonials on social media

**Benjamin L. Ranard** is a doctor of medicine and master of science in health policy research combined-degree student at the Perelman School of Medicine, University of Pennsylvania, in Philadelphia.

**Rachel M. Werner** is a core investigator at the Center for Health Equity Research and Promotion and an associate professor of medicine at the Perelman School of Medicine, University of Pennsylvania.

**Tadas Antanavicius** is an undergraduate student at the University of Pennsylvania.

**H. Andrew Schwartz** is an assistant professor in the Department of Computer Science at Stony Brook University, in Stony Brook, New York.

**Robert J. Smith** is a doctor of medicine student at the Perelman School of Medicine, University of Pennsylvania.

**Zachary F. Meisel** is an assistant professor in emergency medicine at the Hospital of the University of Pennsylvania, in Philadelphia.

**David A. Asch** is executive director of the Center for Health Care Innovation and a professor in the Wharton School and the Perelman School of Medicine, all at the University of Pennsylvania, and a physician in the Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center.

**Lyle H. Ungar** is a professor in the Department of Computer and Information Science at the University of Pennsylvania.

**Raina M. Merchant** (Raina.merchant@uphs.upenn.edu) is director of the Penn Social Media and Health Innovation Lab and an assistant professor in emergency medicine at the University of Pennsylvania.

sites are free, are continuously updated,<sup>9,10</sup> and often reveal exactly what the problem or positive occurrence was that affected the patient's or family member's experience.<sup>8</sup>

Yelp (yelp.com) is a website where users submit star ratings and narrative reviews of local businesses (for example, restaurants, retail stores, hotels), which are then posted for the public to view. It is the 33rd most visited website in the United States and the 155th most viewed site globally,<sup>11</sup> and it has 142 million unique monthly visitors.<sup>12</sup> To date, Yelp is the most widely used freely available commercial website in the United States for hospital ratings.<sup>9</sup> On the site, hospitals are given a rating of one through five stars, which is displayed to the public.<sup>12</sup> The narrative component of reviews often reflects the features of a hospital experience most important to patients, providing information that structured surveys cannot.<sup>8</sup>

Previous work on online patient reviews of hospitals has focused on hospitals in England,<sup>13–15</sup> South Korea,<sup>16</sup> and Germany<sup>17</sup> or has studied US hospitals but has not examined the content of the online reviews.<sup>9,18</sup> In England, the National Health Service (NHS) runs a website, NHS Choices, which allows structured patient reviews to be published;<sup>13,14</sup> however, no such government-run site currently exists in the United States. Little is known about how alternative online sources of patients' and caregivers' experiences of US hospitals may contribute to measuring hospital quality, about what information is contained in these online reviews, and about how this information might be used to identify important experience-of-care measures for hospitalized patients.

We sought to compare the content of all Yelp narrative reviews of all US hospitals that have Yelp reviews to the domains covered by the HCAHPS survey. Our secondary aim was to identify which Yelp topics best correlated with positive or negative Yelp review ratings of hospitals and to correlate Yelp ratings with the HCAHPS survey overall ratings.

## Study Data And Methods

**HOSPITAL CONSUMER SURVEY DATA** We obtained July 2012–June 2013 HCAHPS survey data from the Hospital Compare data set.<sup>19</sup>

**YELP DATA** Using the list of hospitals in the Hospital Compare data set, we identified all hospitals with Yelp reviews posted as of July 15, 2014. We removed hospitals that did not have American Hospital Association (AHA) annual survey data. We also eliminated reviews “not recommended” by Yelp (a measure indicating a review is likely to be fake). Not-recommended re-

views are determined automatically by Yelp's proprietary algorithm that considers a number of factors to try and remove fake reviews (for example, one person posting many reviews from the same computer). According to Yelp, businesses that buy ads cannot influence which reviews are recommended.<sup>20</sup>

**TOPIC GENERATION** We used a type of natural language processing called Latent Dirichlet Allocation to analyze the text of all Yelp narrative reviews of hospitals to produce fifty underlying topics. LDA is widely used to analyze co-occurrences of words in text to produce a predetermined number of topics.<sup>21</sup> Beyond specifying the number of topics to produce, LDA topic generation is fully computer automated. Topics are groups of words (terms) that tend to co-occur. Those groups can then be labeled by human coders based on their content. For example, Latent Dirichlet Allocation generated a topic with co-occurring terms “pain, doctor, nurse, told, medication, meds, gave,” which was then labeled as “pain medications” by coders. We used the implementation of LDA provided by the MALLET package,<sup>22</sup> which has been used previously to analyze the content of other types of social media.<sup>23,24</sup> LDA analysis has also been used previously to characterize online narrative reviews of physicians.<sup>25–27</sup>

**NUMBER OF TOPICS** To choose the number of topics to generate, Latent Dirichlet Allocation was used to generate 25, 50, 100, and 200 topics. Two co-authors served as reviewers and independently rated a blinded sample of each set of topics to determine which predetermined number of topics produced the most coherent topics. If too few topics are generated, the resulting topics are overly broad (for example, from the twenty-five topic list: hospital, medical, care, center, health, people, hospitals). If there are too many topics generated, the topics become extremely specific, with many overlapping topics (for example, from the 200 topic list: ultrasound, pregnant, baby, horrible, bleeding, weeks, miscarriage). After deciding on a predetermined number of topics to generate, we ascertained the significance of the topics by correlating the topics with high and low Yelp review ratings and also by determining the prevalence of the topics in Yelp narrative reviews.

**CODING TOPICS** LDA topics were labeled independently by the coauthor reviewers by viewing the top seven terms in each topic. Adjudication of discrepancies occurred via consensus with a third coauthor reviewer. When concordance was possible, LDA topics were assigned to HCAHPS domains (for example, the topic “pain medications” was placed under the HCAHPS domain Pain Control). Otherwise, topics were as-

# Current hospital ratings based on HCAHPS may be missing the major drivers of patients' overall experiences of care.

signed to new Yelp domains by two of the coauthor reviewers. We assigned each LDA topic to an example quote by selecting a quote from a list of the ten Yelp reviews most associated with each topic. Example quotes may have features of multiple topics, as topics are not mutually exclusive, and several topics may be represented in any single quote from a review.

**HOSPITAL CHARACTERISTICS** We used data from the AHA annual survey to describe hospitals by number of beds, region, teaching status, and ownership. These hospital characteristics have previously been demonstrated to be associated with HCAHPS survey scores.<sup>28</sup>

**STATISTICAL ANALYSIS** We used summary statistics to describe hospital characteristics, number of Yelp reviews, and Yelp ratings for hospitals included in the study cohort. To determine which LDA Yelp topics were associated with high or low Yelp ratings, we correlated those topics with Yelp ratings by calculating the Pearson product-moment correlation coefficient (Pearson's  $r$ ) between each topic and Yelp review rating, and we calculated Bonferroni corrected two-tailed  $p$  values. To determine the prevalence of Yelp topics, each Yelp review was assigned the ten topics most correlated with that review according to the Pearson product-moment correlation coefficient. Prevalence was then calculated based on the percentage of Yelp reviews that contained that topic (as measured by the topic appearing in the top ten topics for a given Yelp review).

We correlated the mean Yelp review rating of each hospital with the HCAHPS survey's overall hospital rating for each hospital and generated Pearson's  $r$ . The number of Yelp reviews per hospital likely affects the correlation; therefore, we calculated the correlation coefficient for different minimum cutoffs of Yelp reviews per hospital. The HCAHPS survey asks patients to give an overall hospital visit rating of 0 (worst) to 10

(best) and reports the results as percentage of patients giving a rating of 9–10, percentage giving 7–8, and percentage giving 6 or lower. We generated a composite HCAHPS overall hospital score for each hospital by using a weighted average of the reported HCAHPS survey scores for an overall hospital visit rating. This study was exempt by the University of Pennsylvania Institutional Review Board.

**LIMITATIONS** There were several limitations to this study. Yelp has inherent selection bias in the reporting of hospital experiences; however, our primary goal was to characterize the content of Yelp reviews. Only 1,352 hospitals had Yelp reviews. It is possible that there were more reviews posted on hospitals not currently reviewed (primarily small, nonteaching, and Midwestern or Southern hospitals), different domains of patient satisfaction might have been identified. With a median of four reviews per hospital, Yelp reviews for many hospitals were too sparse to support robust consumer assessment of an individual hospital's quality. However, analysis of aggregated hospital reviews can still reveal what consumers generally write about (and value) when reviewing hospitals.

## Study Results

There were 4,681 hospitals in the Hospital Compare data set, and 1,451 of these hospitals had Yelp reviews (with a total of 18,058 Yelp reviews). Of the 4,681 hospitals in Hospital Compare, 4,360 hospitals had AHA data, and of these hospitals, 1,352 (31 percent) had Yelp reviews. In aggregate, our final cohort of 1,352 hospitals had 16,862 Yelp reviews, with a median of four Yelp reviews per hospital (interquartile range: 2–13 Yelp reviews). The median date of Yelp review was October 9, 2012 (IQR: April 27, 2011–October 19, 2013). The reviews of hospitals displayed a bimodal distribution of star ratings: 31 percent of reviews gave one star, and 33 percent gave five stars. This bimodal distribution contrasts with the favorably skewed distribution of all Yelp review ratings for all types of businesses and services (see online Appendix Exhibit A1).<sup>29</sup> The median of average Yelp rating per hospital was 3.2 (IQR: 2.47–4.00).

**DESCRIPTION OF HOSPITAL COHORT** The characteristics of the 1,352 hospitals with Hospital Compare data, AHA data, and Yelp reviews, as well as the number of Yelp reviews by hospital characteristic, are reported in Exhibit 1.

**LATENT DIRICHLET ALLOCATION ANALYSIS OF YELP REVIEWS** After reviewing different potential numbers of topics to generate (as described above), we assigned the Latent Dirichlet Allocation model to generate fifty topics from the

## EXHIBIT 1

## Characteristics of hospitals with Yelp reviews, 2005–14

Characteristic	0 Yelp reviews (n = 3,008)		1–4 Yelp reviews (n = 689)		5 or more Yelp reviews (n = 663)		Total (N = 4,360)	
	No.	%	No.	%	No.	%	No.	%
<b>BED SIZE</b>								
0–49	1,173	39%	121	18%	26	4%	1,320	30%
50–199	1,218	40	275	40	217	33	1,710	39
200–399	434	14	206	30	252	38	892	20
400 or more	183	6	87	13	168	25	438	10
<b>REGION</b>								
Northeast	332	11	116	17	121	18	569	13
South	1,379	46	330	48	228	34	1,937	44
Midwest	862	29	119	17	27	4	1,008	23
West	435	14	124	18	287	43	846	19
<b>TEACHING HOSPITAL</b>								
Yes	109	4	46	7	124	19	279	6
No	2,899	96	643	93	539	81	4,081	94
<b>OWNERSHIP</b>								
Government	857	28	112	16	76	11	1,045	24
Nonprofit	1,697	56	430	62	477	72	2,604	60
Physician owned <sup>a</sup>	103	3	25	4	8	1	136	3
Proprietary	351	12	122	18	102	15	575	13

**SOURCE** Authors' analysis of Yelp reviews, 2005–14; and American Hospital Association data, 2008. **NOTE** Percentages may not sum to 100 because of rounding. <sup>a</sup>Includes both nonprofit and proprietary hospitals.

16,862 Yelp reviews. An average of 10.8 (standard deviation: 4.8) topics were covered in a Yelp review. Of the fifty topics, seven (14 percent) were discarded because they were related to names or places (for example, terms such as st [for Saint], hospital, sister, mary's, johns, staff, seton [for Mother Seton]), and two (4 percent) were discarded because the terms could not be used to produce an obvious topic (for example, I'm, don't, time, people, give, you're, guy). The remaining forty-one topics were classified as related to an HCAHPS domain (nine Yelp topics related to seven of the HCAHPS domains, 22 percent) (Exhibit 2) or a new domain not in HCAHPS (thirty-two Yelp topics not related to HCAHPS domains, 78 percent) (see Appendix Exhibit A2).<sup>29</sup> There were twelve new Yelp domains (generated from the thirty-two topics found in Yelp that lacked an analogous HCAHPS domain): cost of hospital visit, insurance and billing, ancillary testing, facilities, amenities, scheduling, compassion of staff, family member care, quality of nursing, quality of staff, quality of technical aspects of care, and specific type of medical care (see Appendix Exhibit A2).<sup>29</sup>

**LATENT DIRICHLET ALLOCATION TOPICS CORRELATED TO YELP RATINGS** Four of the top five (80 percent) Yelp topics most strongly associated with positive Yelp review ratings were not covered by HCAHPS domains. These topics in-

cluded the following: caring doctors, nurses, and staff ( $r = 0.46$ ); comforting ( $r = 0.29$ ); surgery/procedure and peri-op ( $r = 0.23$ ); and labor and delivery ( $r = 0.20$ ). These topics related to the interpersonal relationships of patients with physicians, nurses, and staff with regard to how “caring” or “comforting” they were, or the topics related to specific service lines of the hospital. Two of the top five (40 percent) Yelp topics most strongly associated with negative Yelp review ratings, insurance and billing ( $r = -0.26$ ) and cost of hospital visit ( $r = -0.26$ ), were also not covered by HCAHPS domains (Exhibit 3).

**PREVALENCE OF LATENT DIRICHLET ALLOCATION TOPICS IN YELP REVIEWS** The prevalence of the LDA topics in the Yelp reviews is shown in Appendix Exhibit A3.<sup>29</sup> Nine of the top fifteen (60 percent) most prevalent topics were not in the HCAHPS survey. The top five most prevalent topics were caring doctors, nurses, and staff (46 percent of Yelp reviews); waiting for doctors and nurses (43 percent of Yelp reviews); friendliness of emergency department staff (35 percent of Yelp reviews); nice, friendly staff and nurses (33 percent of Yelp reviews); and patient treatment by physician and information provided (32 percent of Yelp reviews).

**YELP RATING CORRELATED TO OVERALL RATING** The mean Yelp rating per hospital was correlated with the HCAHPS overall rating for each hospi-

**EXHIBIT 2**

**Yelp Latent Dirichlet Allocation topics associated with HCAHPS survey domains, 2005–14**

HCAHPS domain (HCAHPS questions)	Yelp topic (topic terms)	Example quote
Communication with nurses How often did nurses treat you with courtesy and respect? How often did nurses listen carefully to you? How often did nurses explain things in a way you could understand?	Rude doctor/nurse communication <sup>a</sup> (nurse, asked, told, didn't, doctor, questions, rude)	"I asked a question...[the nurse] snapped at me with a raised voice and very condescendingly asked me if there was anything else I did not understand. No one deserves to be treated like that for a question that is asked to their health care PROFESSIONAL."
Communication with doctors How often did doctors treat you with courtesy and respect? How often did doctors listen carefully to you? How often did doctors explain things in a way you could understand?	Patient treatment by physician and information provided (patient, medical, hospital, information, treatment, physician, case)	"For my first visit with [the doctor,] I was given a 3 part brochure listing his education, qualifications, his expectations of you as a patient and what you should expect from him as your surgeon. I have never had any Dr., dentist or other health professional give me any information approaching the thoroughness of that brochure."
Responsiveness of hospital staff How often did you get help as soon as you wanted after you pressed the call button? How often did you get help in getting to the bathroom or in using a bedpan as soon as you wanted?	Nursing responsiveness (nurse, room, bed, hours, minutes, nurses, hour)  Waiting for doctors and nurses (minutes, room, waiting, doctor, wait, hours, nurse)	"[My sister] rang for the nurse b/c she had to go to the bathroom—nurse didn't come for 40 min!!! Even after we went to get her she still took another 20 min!!!" "Checked in @ 7:30 AM no procedure until 12:30 PM. Designated echo doctor changed 3 times. Procedures started @ 1:30 PM took 2 hrs, 4.5 hrs for recovery. The wait was tough..."
Pain control How often was your pain well controlled? How often did the hospital staff do everything they could to help you with your pain?	Pain medications (pain, doctor, nurse, told, medication, meds, gave)	"...[D]o not come here if you have a severe migraine because you will be pumped up full of narcotics, which is not the appropriate treatment."
Cleanliness of hospital environment How often were your room and bathroom kept clean?	Clean, private, nice hospital rooms (room, hospital, rooms, nice, staff, private, clean)	"...[T]he common areas and patient room were clean and decorated smartly (laminated wood floors and tempur pedic mattress for sofa bed!)"
Overall hospital rating What number would you use to rate this hospital during your stay?	Horrible hospital (hospital, place, people, don't, worst, treated, horrible) Great care (wife, hospital, staff, care, good, great, time)	"This place is terrifying. I'd rather bleed to death." "My wife gave birth here and the attention and service was excellent..."
Likelihood to recommend Would you recommend this hospital to your friends and family?	Reviewing hospital experience (hospital, review, experience, ive, im, years, time)	"If you have a choice, go anywhere else. The people, the food, the quality of care is abysmal."

**SOURCE** Authors' analysis of Yelp reviews, 2005–14. **NOTE** Four Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) domains have no associated Yelp topics: communication about medicines, discharge information, care transition, and quietness of hospital environment. <sup>a</sup>Also appears under "communication with doctors."

tal. Pearson's *r* was 0.50 using the 871 hospitals with more than two Yelp reviews per hospital (Exhibit 4). Increasing values of *r* could be seen by increasing the minimum number of Yelp reviews per hospital cutoff—for example, more than five Yelp reviews per hospital (*r* = 0.56), more than fifteen Yelp reviews (*r* = 0.63), and more than twenty-five reviews (*r* = 0.69) (Appendix Exhibit A4).<sup>29</sup>

**Discussion**

Seventy-two percent of US Internet users reported looking online for health information

in 2012,<sup>30</sup> and 42 percent reported looking at social media for health-related consumer reviews.<sup>31</sup> Meanwhile, only 6 percent of Americans had heard of the Hospital Compare website as of 2008.<sup>32</sup> This discrepancy raises the possibility that there are important lessons from the information contained in online consumer reviews that can be used to improve current, more formal rating systems and increase their use in consumer decision making.

This study has three main findings. First, hospitals with at least three Yelp reviews (*n* = 871) had mean Yelp ratings that correlated relatively strongly with an HCAHPS item asking about

EXHIBIT 3

Top five Latent Dirichlet Allocation Yelp topics correlated to high and low Yelp ratings, 2005-14

Topic name	Correlation, Pearson's r	Covered by HCAHPS
<b>YELP TOPICS MOST CORRELATED WITH POSITIVE YELP RATINGS</b>		
Caring doctors, nurses, and staff	0.46	No
Comforting	0.29	No
Clean, private, nice hospital rooms	0.25	Yes
Surgery/procedure and peri-op	0.23	No
Labor and delivery	0.20	No
<b>YELP TOPICS MOST CORRELATED WITH NEGATIVE YELP RATINGS</b>		
Horrible hospital	-0.33	Yes
Rude doctor/nurse communication	-0.29	Yes
Pain control	-0.28	Yes
Insurance and billing	-0.26	No
Cost of hospital visit	-0.26	No

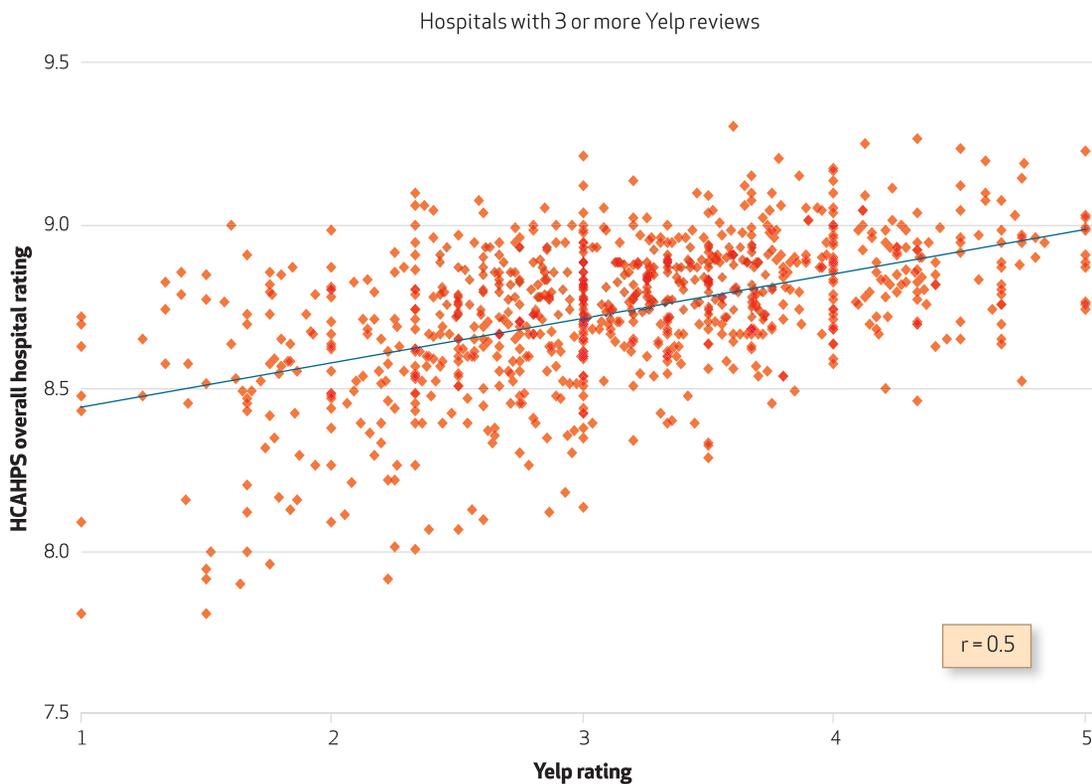
**SOURCE** Authors' analysis of Yelp reviews, 2005-14. **NOTES** If a topic was similar to an already listed topic, it was removed and replaced by the next most associated topic. HCAHPS is Hospital Consumer Assessment of Healthcare Providers and Systems survey. All correlation coefficients are statistically significant ( $p < 0.0001$ ).

overall hospital rating ( $r = 0.50$ ). This builds on older data. Naomi Bardach and colleagues showed that high Yelp review ratings correlated with high HCAHPS overall ratings for 270 hospitals with more than five Yelp reviews.<sup>9</sup> Additionally, Felix Greaves and colleagues demonstrated a correlation in England between patient ratings collected on the NHS Choices website and traditional paper-based survey responses.<sup>13</sup> Our study's correlation is notable considering we only required a minimum of three Yelp reviews per hospital. Second, Yelp reviews cover far more topics than the HCAHPS survey does. While Yelp reviews include information about seven of the eleven HCAHPS domains, twelve additional Yelp domains not covered by the HCAHPS survey were identified from Yelp topics. Third, many of the Yelp topics most associated with strongly positive or negative Yelp ratings are not covered by the HCAHPS survey nor are a majority of the most prevalent Yelp topics.

Our third finding suggests that current hospital ratings based on HCAHPS may be missing the

EXHIBIT 4

Mean Yelp review ratings correlated with HCAHPS survey overall hospital ratings, 2005-14



**SOURCE** Authors' analysis of Yelp reviews, 2005-14; American Hospital Association data, 2008; and Hospital Compare data, 2012-13. **NOTES** Mean Yelp review rating is correlated with Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) overall rating for hospitals with more than two Yelp reviews ( $n = 871$ ;  $p < 0.001$ ). Sensitivity analysis with different cutoffs of Yelp reviews per hospital is shown in Appendix Exhibit A4 (see Note 29 in text).

# The content of Yelp narrative reviews reflects new areas of importance to patients and caregivers.

major drivers of patients' overall experiences of care. When Yelp users give an overall hospital rating, they presumably write a narrative review about aspects that went into their overall rating. When patients respond to the HCAHPS question asking about overall hospital rating, they may (or may not) be assigning an overall rating based on HCAHPS domains. The Yelp topics may provide a more nuanced view of aspects of hospital quality that patients value. For example, in consumer focus groups that were conducted as part of the development of the HCAHPS survey to identify consumers' preferences for hospital quality, "compassion and kindness" of staff was important.<sup>33</sup> Yet "compassion and kindness" is not an item on the HCAHPS survey but instead is broadly reported under "How often did doctors [or nurses] communicate well with patients?" Meanwhile, "compassion of staff" was an important Yelp domain that contained seven of the forty-one Yelp topics. While it is possible for a rating of communication to include an element of compassion, it is also possible to show a lack of compassion and empathy without traditional communication: "[My boyfriend]...was on the floor crying and screaming in pain. Nurses walked by as if we didn't exist!!! Finally...[t]hey stripped him naked and proceeded to do tests on him with very little empathy. The ER was freezing cold, they didn't even offer him a blanket." Patients who are looking for hospitals with empathetic, compassionate staff might rather visit Hospital Compare and view ratings for "How comforting and empathetic were the hospital staff?" instead of, or in addition to, "How often did doctors [or nurses] communicate well with patients?"

This is the first study to use automated computational methods to create *de novo* topics from online hospital reviews and the first study to characterize the content of narrative online hospital reviews in the United States. Two previous studies that characterized *de novo* topics from online reviews or tweets used qualitative coding to describe the content of 200 reviews from twen-

ty English NHS hospitals<sup>14</sup> or 1,000 tweets to English hospitals.<sup>15</sup> This study examined a large number of reviews from a large number of US hospitals, used an automated method that allowed for easier analysis of large data sets, and attempted to identify important domains of patients' hospital experiences by analyzing what patients and their caregivers wrote about when reviewing their care experiences.

In addition to offering a view across broader domains, online consumer rating platforms offer several advantages. Reviews are in real time, without the delay in reporting that the HCAHPS survey experiences. Additionally, narrative reviews can supplement data from the HCAHPS survey by providing actionable insights for hospitals to improve patient satisfaction (for example, patient or caregiver feedback about poor experiences interacting with the billing department or specific rude staff members). Finally, Yelp is easy to use, and consumers are already comfortable with the platform. The more than 140 million users who already look up restaurants and businesses on Yelp may be more likely to turn to services such as Yelp to look up hospitals' ratings instead of going to Hospital Compare.

Despite these advantages, concerns remain about using information from platforms such as Yelp. Unlike the HCAHPS survey, Yelp does not solicit reviews from a random selection of patients discharged from the hospital. Yelp reviews are more likely coming from younger and more educated consumers. Second, hospitals may have difficulty deciding how to incorporate unsolicited online reviews and how to respond to them. In response to sites such as Yelp, several hospitals have begun allowing patients to review doctors directly on the hospital's website. Hospitals contend that this provides transparency and information to patients and allows hospitals to verify that reviews are from actual patients. Critics argue that hospitals are not in a neutral position to determine which reviews to post.<sup>34</sup> Finally, not all hospitals are represented by Yelp reviews, and half of hospitals have four or fewer reviews, which limits usefulness. This concern may be ameliorated in the future if the use of Yelp and similar platforms to rate hospitals continues to increase.

Those who publicly report quality could learn from consumer review websites, and consumer review websites could do more to incorporate systematic assessments of providers. A major barrier to the public reporting movement has been engaging consumers in responding to publicly reported quality information.<sup>35</sup> If features from popular review websites such as Yelp could be replicated in more systematic assessments of

providers, there is potential to increase patients' engagement with them. In fact, the NHS in England has already taken steps in this direction and currently allows consumers to post narrative reviews and star ratings on its public website.<sup>14</sup> Furthermore, Hospital Compare just began displaying star ratings for HCAHPS survey results to make its website more engaging for consumers.<sup>36</sup> Lessons could also be learned from Amazon.com on how to curate large numbers of consumer reviews by highlighting the most helpful positive and negative reviews. Collection of narrative reviews themselves could also be made more systematic.<sup>8</sup> For example, a random selection of patients could be e-mailed or texted after discharge to ask for a narrative review of their hospital visit. Natural language processing approaches could then be used in real time to analyze and incorporate these reviews into online hospital ratings. Consumer websites could also do more to incorporate information from systematic assessments. As a step in this direction, Yelp recently partnered with ProPublica, an

independent provider of investigative journalism, to display certain information from Hospital Compare on the Yelp pages of medical facilities.<sup>37</sup>

## Conclusion

Despite the relatively small number of hospitals that had Yelp reviews, it is apparent to us that online consumer review platforms such as Yelp can supplement information provided by more traditional patient-experience surveys and contribute to patients', providers', and policy makers' understanding and assessment of hospital quality. The content of Yelp narrative reviews mirrors many aspects of the HCAHPS survey but also reflects new areas of importance to patients and caregivers that may have important implications for policy makers seeking to measure the patient experience of hospital quality and hospitals attempting to improve patient satisfaction. ■

Funding and support were provided by the National Institutes of Health, Grant No. K23-10714038 (for Raina M. Merchant) and Grant No. K24-AG047908 (for Rachel M. Werner). The study sponsors had no role in the design and conduct of the study; collection,

management, analysis, and interpretation of the data; preparation, review, or approval of the article; or decision to submit the article for publication. The authors thank Judith A. Long (Perelman School of Medicine, University of Pennsylvania) for her

review of the article and her insightful comments. She received no financial compensation. The opinions expressed are those of the authors and not the Department of Veterans Affairs.

## NOTES

- Giordano LA, Elliott MN, Goldstein E, Lehrman WG, Spencer PA. Development, implementation, and public reporting of the HCAHPS survey. *Med Care Res Rev.* 2010; 67(1):27-37.
- Goldstein E, Farquhar M, Crofton C, Darby C, Garfinkel S. Measuring hospital care from the patients' perspective: an overview of the CAHPS hospital survey development process. *Health Serv Res.* 2005; 40(6 Pt 2):1977-95.
- Centers for Medicare and Medicaid Services. What is Hospital Compare? [Internet]. Baltimore (MD): CMS; 2014 [cited 2016 Feb 11]. Available from: <http://www.medicare.gov/hospitalcompare/About/What-Is-HOS.html>
- Centers for Medicare and Medicaid Services. Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and fiscal year 2014 rates; quality reporting requirements for specific providers; hospital conditions of participation; payment policies related to patient status. Final rules. *Fed Regist.* 2013;78(160):50495-1040.
- Jordan H, White A, Joseph C, Carr D. Costs and benefits of HCAHPS: final report [Internet]. Cambridge (MA): Abt Associates Inc.; 2005 Oct 5 [cited 2016 Feb 11]. Available from: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/downloads/HCAHPSCostsBenefits200512.pdf>
- Centers for Medicare and Medicaid Services. Summary of HCAHPS survey results [Internet]. Baltimore (MD): CMS; 2014 Jul [cited 2016 Feb 11]. Available from: [http://www.hcahpsonline.org/files/Report\\_July\\_2014\\_States.pdf](http://www.hcahpsonline.org/files/Report_July_2014_States.pdf)
- Centers for Medicare and Medicaid Services. HCAHPS public reporting [Internet]. Baltimore (MD): CMS; 2014 Oct 14 [cited 2016 Feb 11]. Available from: [http://www.hcahpsonline.org/files/HCAHPS%20Public%20Reporting\\_April2015\\_Dec2016.pdf](http://www.hcahpsonline.org/files/HCAHPS%20Public%20Reporting_April2015_Dec2016.pdf)
- Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, et al. Taking patients' narratives about clinicians from anecdote to science. *N Engl J Med.* 2015; 373(7):675-9.
- Bardach NS, Asteria-Peñaloza 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.
- Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. *J Med Internet Res.* 2014;16(11):e264.
- Alexa Internet Inc. Site overview: yelp.com [Internet]. San Francisco (CA): Alexa; 2016 [cited 2016 Feb 19]. Available from: <http://www.alexa.com/siteinfo/yelp.com>
- Yelp Inc. About Yelp [Internet]. San Francisco (CA): Yelp; [cited 2016 Feb 11]. Available from: <http://www.yelp-press.com/>
- Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Qual Saf.* 2012;21(7):600-5.
- Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A mixed-methods analysis of patient reviews of hospital care in England: implications for public reporting of health care quality data in the United States. *Jt Comm J Qual Patient Saf.* 2013;

- 39(1):7–15.
- 15 Greaves F, Lavery AA, Cano DR, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. *BMJ Qual Saf*. 2014;23(10):838–46.
  - 16 Jung Y, Hur C, Jung D, Kim M. Identifying key hospital service quality factors in online health communities. *J Med Internet Res*. 2015;17(4):e90.
  - 17 Drevs F, Hinz V. Who chooses, who uses, who rates: the impact of agency on electronic word-of-mouth about hospitals stays. *Health Care Manage Rev*. 2014;39(3):223–33.
  - 18 Glover M, Khalilzadeh O, Choy G, Prabhakar AM, Pandharipande PV, Gazelle GS. Hospital evaluations by social media: a comparative analysis of Facebook ratings among performance outliers. *J Gen Intern Med*. 2015;30(10):1440–6.
  - 19 Data.Medicare.gov. Hospital Compare data: patient survey (HCAHPS)—hospital [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; [last updated 2015 Dec 10; cited 2016 Feb 11]. Available for download (click on Microsoft access database) from: <https://data.medicare.gov/>
  - 20 Yelp Inc. Recommended reviews [Internet]. San Francisco (CA): Yelp; [cited 2016 Feb 11]. Available from: [http://www.yelp-support.com/Recommended\\_Reviews/](http://www.yelp-support.com/Recommended_Reviews/)
  - 21 Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Machine Learn Res*. 2003;3:993–1022.
  - 22 McCallum AK. MALLETT: a machine learning for language toolkit [Internet]. Amherst (MA): University of Massachusetts, Amherst; 2002 [cited 2016 Feb 11]. Available for download from: <http://mallet.cs.umass.edu>
  - 23 Schwartz HA, Ungar LH. Data-driven content analysis of social media: a systematic overview of automated methods. *Ann Am Acad Pol Soc Sci*. 2015;659(1):78–94.
  - 24 Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS One*. 2013;8(9):e73791.
  - 25 Brody S, Elhadad N. Detecting salient aspects in online reviews of health providers. *AMIA Annu Symp Proc*. 2010;2010:202–6.
  - 26 Paul MJ, Wallace BC, Dredze M. What affects patient (dis)satisfaction? Analyzing online doctor ratings with a joint topic-sentiment model [Internet]. Palo Alto (CA): Association for the Advancement of Artificial Intelligence; 2013 [cited 2016 Feb 11]. Available from: <https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/viewFile/7120/6510>
  - 27 Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. *J Am Med Inform Assoc*. 2014;21(6):1098–103.
  - 28 Lehrman WG, Elliott MN, Goldstein E, Beckett MK, Klein DJ, Giordano LA. Characteristics of hospitals demonstrating superior performance in patient experience and clinical process measures of care. *Med Care Res Rev*. 2010;67(1):38–55.
  - 29 To access the Appendix, click on the Appendix link in the box to the right of the article online.
  - 30 Fox S, Duggan M. Health online 2013 [Internet]. Washington (DC): Pew Research Center; 2013 Jan 15 [cited 2016 Feb 11]. Available from: <http://pewinternet.org/Reports/2013/Health-online.aspx>
  - 31 PricewaterhouseCoopers Health Research Institute. Social media “likes” healthcare: from marketing to social business [Internet]. New York (NY): PwC; 2012 [cited 2016 Feb 11]. Available for download from: <http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml>
  - 32 Henry J. Kasier Family Foundation. 2008 update on consumers’ views of patient safety and quality information: summary and chartpack [Internet]. Menlo Park (CA): KFF; 2008 Oct [cited 2016 Feb 11]. Available from: <http://kaiserfamilyfoundation.files.wordpress.com/2013/01/7819.pdf>
  - 33 Sofaer S, Crofton C, Goldstein E, Hoy E, Crabb J. What do consumers want to know about the quality of care in hospitals? *Health Serv Res*. 2005;40(6 Pt 2):2018–36.
  - 34 Ramey C. Long Island hospital posts doctor ratings: North Shore-LIJ is the first in New York area, and among the few in the U.S., to do so. *Wall Street Journal* [serial on the Internet]. 2015 Aug 26 [cited 2016 Feb 11]. Available from: <http://www.wsj.com/articles/long-island-hospital-posts-doctor-ratings-1440635377>
  - 35 Ketelaar NA, Faber MJ, Flottorp S, Rygh LH, Deane KH, Eccles MP. Public release of performance data in changing the behaviour of health-care consumers, professionals, or organisations. *Cochrane Database Syst Rev*. 2011;(11):CD004538.
  - 36 Centers for Medicare and Medicaid Services. Technical notes for HCAHPS star ratings [Internet]. Baltimore (MD): CMS; 2014 Sep [cited 2016 Feb 11]. Available from: [http://www.hcahpsonline.org/files/HCAHPS\\_Stars\\_Tech\\_Notes\\_9\\_17\\_14.pdf](http://www.hcahpsonline.org/files/HCAHPS_Stars_Tech_Notes_9_17_14.pdf)
  - 37 Stoppelman J. Yelp’s consumer protection initiative: ProPublica partnership brings medical info to Yelp. *Yelp Official Blog* [blog on the Internet]. 2015 Aug 5 [cited 2016 Feb 11]. Available from: <http://officialblog.yelp.com/2015/08/yelps-consumer-protection-initiative-propubica-partnership-brings-medical-info-to-yelp.html>