ORIGINAL RESEARCH

OPEN ACCESS

Measuring patient-perceived quality of care in US hospitals using Twitter

Jared B Hawkins,[1,2] John S Brownstein,[1,2,3] Gaurav Tuli,[4,5] Tessa Runels,[2] Katherine Broecker,[2] Elaine O Nsoesie,[2,3,6] David J McIver,[2] Ronen Rozenblum,[7,8] Adam Wright,[7,8] Florence T Bourgeois,[2,3] Felix Greaves[9]

For numbered affiliations see end of article.

Correspondence to
Dr Jared B Hawkins, HealthMap BCH3393, 300 Longwood Ave, Boston, MA 02115, USA; jared.hawkins@childrens.harvard.edu

Open Access
Scan to access more
free content

Linked

▶ http://dx.doi.org/10.1136/bmjqs-2015-004489
▶ http://dx.doi.org/10.1136/bmjqs-2015-004827

CrossMark

ABSTRACT

Background Patients routinely use Twitter to share feedback about their experience receiving healthcare. Identifying and analysing the content of posts sent to hospitals may provide a novel real-time measure of quality, supplementing traditional, survey-based approaches.

Objective To assess the use of Twitter as a supplemental data stream for measuring patient-perceived quality of care in US hospitals and compare patient sentiments about hospitals with established quality measures.

Design 404 065 tweets directed to 2349 US hospitals over a 1-year period were classified as having to do with patient experience using a machine learning approach. Sentiment was calculated for these tweets using natural language processing. 11 602 tweets were manually categorised into patient experience topics. Finally, hospitals with ≥50 patient experience tweets were surveyed to understand how they use Twitter to interact with patients.

Key results Roughly half of the hospitals in the US have a presence on Twitter. Of the tweets directed toward these hospitals, 34 725 (9.4%) were related to patient experience and covered diverse topics. Analyses limited to hospitals with ≥50 patient experience tweets revealed that they were more active on Twitter, more likely to be below the national median of Medicare patients (p<0.001) and above the national median for nurse/patient ratio (p=0.006), and to be a non-profit hospital (p<0.001). After adjusting for hospital characteristics, we found that Twitter sentiment was not associated with Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) ratings (but having a Twitter account was), although there was a weak association with 30-day hospital readmission rates (p=0.003).

Conclusions Tweets describing patient experiences in hospitals cover a wide range of patient care aspects and can be identified using automated approaches. These tweets represent a potentially untapped indicator of quality and may be valuable to patients, researchers, policy makers and hospital administrators.

INTRODUCTION

Over the past decade, patient experiences have drawn increasing interest, highlighting the importance of incorporating patients' needs and perspectives into care delivery.[1 2] With healthcare becoming more patient centred and outcome and value driven, healthcare stakeholders need to be able to measure, report and improve outcomes that are meaningful to patients.[2–5] These outcomes can only be provided by patients, and thus systems are needed to capture patient-reported outcomes and facilitate the use of these data at both an individual patient level and the population level.[2 3] Structured patient experience surveys such as Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) are common methods designed to assess patients' perception of the quality of their own healthcare.[4 6 7] A major drawback with these surveys is the significant time lag—often several months—before official data are released, making it difficult for patients and other concerned stakeholders to be informed about current opinions on the quality of a given institution. Moreover, these surveys traditionally have low response rates,[4 8] raising concerns about potential response and selection bias in the results.

Social media usage is pervasive in the USA, with most networks seeing growth in their user base each year. As of 2014, approximately one out of five adults actively use Twitter; while most popular with adults under 50 years old, the network has seen significant growth in the 65 and older population in the past

year.[9] Although there are legitimate privacy, social, ethical and legal concerns about interacting with patients on social media,[10–12] it is clear that patients are using these venues to provide feedback.[13–21] In addition, the use of social media data for health research has been gaining popularity in recent years.[22 23]

Sentiment analysis of social media is useful for determining how people feel about products, events, people and services.[24–28] It is widely used in other industries, including political polling[29–32] and brand/reputation management.[33 34] Researchers have also been experimenting with sentiment analysis of social media for healthcare research.[13 14 16 17] Sentiment can be determined in several ways, with the goal being to classify the underlying emotional information as either positive or negative. This can be done either purely by human input or by an algorithm trained to complete this process based on a human-classified set of objects, and reliability is largely a function of the method used.

We seek to describe the use of Twitter as a novel, real-time supplementary data stream to identify and measure patient-perceived quality of care in US hospitals. This approach has previously been used to examine patient care in the UK.[17] While there was no correlation between Twitter sentiment and other standardised measures of quality, the analysis provided useful insight for quality improvement. Our aims are to provide a current characterisation of US hospitals on Twitter, explore the unsolicited patient experience topics discussed by patients, and determine if Twitter data are associated with quality of care, as compared with other established metrics.

METHODS

Hospital Twitter data

We compiled a list of Twitter accounts for each hospital in the USA. The October 2012–September 2013 HCAHPS report served as our hospital master list and included 4679 hospitals. We used Amazon's Mechanical Turk (AMT)—an online tool that allows large, tedious jobs to be completed very quickly by harnessing the efforts of crowd-sourced employees[35]—to identify a Twitter account for each hospital. Two AMT workers attempted to identify an account for each hospital, with any disagreements resolved by manual inspection (FG). We used the services of DataSift, a data broker for historical Twitter data, to obtain all tweets that mentioned any of these hospitals during the 1-year period from 1 October 2012 to 30 September 2013. Mentions were defined as tweets if they were specifically directed toward a hospitals' Twitter account (ie, they included the full hospital Twitter handle, such as '@BostonChildrens'). During this time frame, we found 404 065 tweets that were directed at these hospitals. Tweets and associated metadata were cleaned and processed by custom Python scripts and stored in a database (MongoDB) for further analysis. This study only analysed tweets that were completely public (ie, no privacy settings were selected by the user) and that were original tweets—we ignored all retweets (tweets from another individual that have been reposted) to ensure we were capturing unique patient experience feedback. Furthermore, there were no personal identifiers used in our analysis, and thus there was no knowledge of the users' identities. The study was approved by the Boston Children's Hospital Institutional Review Board, which granted waiver of informed consent.

Machine learning classifier

We manually curated a random subset of hospital tweets to identify those pertaining to patient experiences—defined as patient's, friend's, or family member's discussion of healthcare experience. Some examples of patient experience included: interactions with staff, treatment effectiveness, hospital environment (food, cleanliness, parking, etc), mistakes or errors in treatment or medication administration, and timing or access to treatment. Curation was achieved via two methods. The first method used a custom web-app that facilitated the curation of randomly selected tweets from the database, by allowing multiple curators (including TR and KB) to label them if related to patient experiences. Each tweet was labelled by two curators, and only those tweets labelled identically were used for the analysis. The second method of data curation used AMT for crowd-sourced labelling. Again, multiple curators, who were classified by Amazon as being highly experienced in the field of sentiment analysis (Master Workers),[36] labelled each tweet, and only those tweets that agreed in their labelling were used. To test if curators for both methods were classifying tweets reliably, we calculated inter-rater agreement and Cohen's κ values[37 38] between raters. Because we used multiple pairs of curators for the first method and AMT can use hundreds of individual curators for a project, we focused our efforts on the pairs of curators that were most prolific between specific sets of curators. The most prolific raters in the web-based method, representing four curators out of eight total and 52% of all classified tweets, showed an average agreement of 94.7% and an average κ value of 0.425 ($p < 0.001$) across 12 620 tweets. The analogous raters for the AMT method, representing 20 curators out of 210 and 5% of classified tweets, had an average agreement of 97.7% and an average κ value of 0.788 ($p < 0.001$) across 529 tweets. After multiple rounds of curation, curator pairs had rated 24 408 tweets using the web-app (overall agreement of 90.64%) and 15 000 tweets using AMT (overall agreement of 80.64%). These two sets were combined to create a training set of 2216 tweets relating to patient experiences and 22 757 tweets covering other aspects of the hospital.

Original research

This training set was used to build a classifier that could automatically label the full database of tweets. The machine learning approach looks at features of the tweets (eg, number of friends/followers/tweets from the user, user location and the specific words used in the tweet, but never username) and uses this information to develop a classifier. For the text of a tweet, we used a bag-of-words approach and included unigrams, bigrams and trigrams in the analysis. Specifically, we compared multiple different classifiers (naive Bayes and support vector machine) and subjectively selected the best classifier based on a variety of metrics such as F1 score, precision, recall and accuracy. Building the classifier was an iterative process and we retrained and improved the classifier over many rounds of curation. We used 10-fold cross-validation for evaluating the different classifiers, and selected a support vector machine classifier with an average accuracy of 0.895. This classifier on average had an F1 score of 0.806, precision of 0.818, and recall of 0.795.

Sentiment calculation

We used natural language processing (NLP) to measure the sentiment of all patient experience tweets. Sentiment was determined using the open-source Python library TextBlob.[39] The sentiment analyser implementation used by TextBlob is based on the Pattern library,[40] which is trained from human annotated words commonly found in product reviews. Sentiment scores range from −1 to 1, and scores of exactly 0.0 were discarded, as they typically indicate that there was not sufficient context. The average number of patient experience tweets for all hospitals was 43. To ensure that there were enough tweets to provide an accurate assessment of sentiment, we calculated a mean sentiment score for each hospital with ≥50 patient experience tweets (n=297).

Hospital characterisation

We compared the proportion of hospitals in each of the following American Hospital Association (AHA) categorical variables between the highest and lowest sentiment quartiles: region, urban status, bed count, nurse-to-patient ratio, profit status, teaching status and percentage of patients on Medicare/Medicaid. We compared nurse-to-patient ratio and percentage of patients on Medicare/Medicaid with the median national value. We used the following Twitter characteristics (measured in August 2014) for sentiment correlation and quartile comparison: days that the account has been active; number of status updates; number of followers; number of patient experience tweets received; and number of total tweets received.

Topic classification

We again used AMT to identify which topics were being discussed in the patient experience tweets. A total of 11 602 machine-identified patient experience tweets were classified by AMT workers as belonging to one or more predefined categories. Only tweets with agreed-upon labels were further analysed; this totalled 7511 tweets (overall agreement of 64.7%); of these, 3878 were identified as belonging to a patient experience category, and 3633 were found to be not truly about patient experience. Owing to the sheer number of topics, we calculated average agreement and Cohen's κ values for both workers for each topic. We found that the topics Food, Money, Pain, General, Room condition, and Time had an average agreement of 91.7% and a moderate κ of 0.52 (p<0.001), while the topics Communication, Discharge, Medication instructions, and Side effects had an average agreement of 97.4% and a low κ of 0.18 (p<0.001).

Hospital surveys

We emailed contacts with formal positions in the office of patient or public relations (or equivalent) of the 297 hospitals with ≥50 patient experience tweets (111 unique Twitter accounts) and asked them to provide feedback regarding their use of Twitter for patient relations. If employees could not be identified, either the department email (n=44) or general contact email (n=40) was used. Contact was attempted twice, with a second email sent 9 days after the first if necessary. The questions asked were: (1) "Does your hospital monitor Twitter activity?"; (2) "Do you follow-up with patients regarding comments they make on Twitter?"; and (3) "Are you aware that patients post about their hospital/care experience on Twitter?". Informants were told their participation in the study was voluntary, confidential and anonymous.

Comparison with validated measures of quality of care

We chose two validated measures of quality of care. The first was HCAHPS, the formal US nationwide patient experience survey. The intent of the HCAHPS is to provide a standardised survey instrument and data-collection methodology for measuring patients' perspectives on hospital care, which enables valid comparisons to be made across all hospitals. Like other traditional patient surveys, the HCAHPS is highly standardised and well validated.[4 6 7] We focused on the percentage of patients who rated a hospital a 9 or 10 (out of 10), which has been shown to correlate with direct measures of quality,[4] although we also looked at the percentage of patients who gave a 0 to 6 rating (not shown). We analysed data from the HCAHPS period 1 October 2012–30 September 2013. The second validated measure of quality of care was the Hospital Compare 30-day hospital readmission rate calculated from the period 1 July 2012–30 June 2013. This is a standardised metric covering 30-day overall rate of unplanned readmission after discharge from the hospital and includes patients admitted for internal medicine, surgery/gynaecology, cardiorespiratory, cardiovascular and neurology services.[41]

The score represents the ratio of predicted readmissions (within 30 days) to the number of expected readmissions, multiplied by the national observed rate.[42]

Statistical analysis

We used Pearson's correlation to assess the linear relationship between numeric variables, Fisher's exact test to compare proportions between categorical variables, and a two-tailed independent t test to compare the means of quantiles. Bonferroni correction was used to adjust for multiple comparisons. Multivariable linear regression was used to adjust for potential confounders such as: region, size, bed count, profit status, rural/urban status, teaching status, nurse-to-patient ratio, percentage of patients on Medicare and percentage of patients on Medicaid. Twitter account confounders (total statuses, total followers, and total days since account creation) were measured in August 2014. Additional Twitter covariates were the total number of patient experience tweets received during the study period and whether or not the hospital had a unique Twitter handle (as opposed to sharing with a

larger healthcare network). A Wald test was used to test for trend significance.

RESULTS

Characteristics of US hospitals on Twitter

Of the 4679 US hospitals identified, 2349 (50.2%) had an account on Twitter; this included data from 1609 Twitter handles (as many hospitals within a provider network share the same Twitter handle). During the 1-year study period, we found 404 065 total tweets directed towards these hospitals (data from 1418 Twitter handles, representing 2137 hospitals); of these, 369 197 (91.4%) were original tweets (data from 1417 Twitter handles, representing 2136 hospitals). The classifier tagged 34 725 (9.4%) original tweets relating to patient experiences and 334 472 (90.6%) relating to other aspects of the hospital. Patient experience tweets were found for 1065 Twitter handles, representing 1726 hospitals (36.9%).

Table 1 describes the common characteristics for all of the hospitals with Twitter accounts. Overall, the mean number of patient experience tweets received

Table 1 Characteristics of US hospitals using Twitter

| Metric	Followers (n=2349)		Sentiment (n=297)		Proportion of hospitals in sentiment quartiles		
	Median	IQR	Median	IQR	Highest quartile	Lowest quartile	p Value
Region							0.392
Northeast	666	188–2686	0.278	0.124–0.377	0.27	0.31	
Midwest	981	176–2881	0.296	0.263–0.332	0.33	0.20	
West	437	118–1426	0.213	0.213–0.293	0.02	0.17	
South	832	183–2522	0.300	0.280–0.334	0.39	0.31	
Urban							0.379
Yes	1087	303–3069	0.293	0.244–0.334	0.77	0.93	
No	364	70–1871	0.301	0.263–0.334	0.23	0.07	
Bed count							0.037*
Small (<100)	439	72–2198	0.312	0.270–0.338	0.41	0.13	
Medium (100–299)	622	166–2182	0.294	0.270–0.334	0.24	0.30	
Large (300+)	1610	527–3592	0.280	0.222–0.331	0.35	0.57	
Nurse-patient ratio							0.395
Above national	853	151–3078	0.301	0.270–0.338	0.59	0.39	
Below national	741	182–2199	0.283	0.223–0.334	0.41	0.61	
Profit status							<0.001*
Public	237	48–1549	0.263	0.112–0.299	0.06	0.26	
Private non-profit	1115	281–3008	0.301	0.263–0.334	0.88	0.74	
Private for-profit	327	103–934	0.280	0.281–0.326	0.06	0.00	
Teaching hospital							0.242
Yes	1359	382–3498	0.285	0.223–0.332	0.45	0.67	
No	527	119–2005	0.301	0.270–0.334	0.55	0.33	
Medicare							0.617
Above national	605	138–2185	0.298	0.265–0.334	0.55	0.37	
Below national	1187	247–3228	0.294	0.228–0.334	0.45	0.63	
Medicaid							1
Above national	756	160–2459	0.295	0.227–0.338	0.61	0.65	
Below national	819	187–2828	0.298	0.270–0.334	0.39	0.35	

*p<0.05.
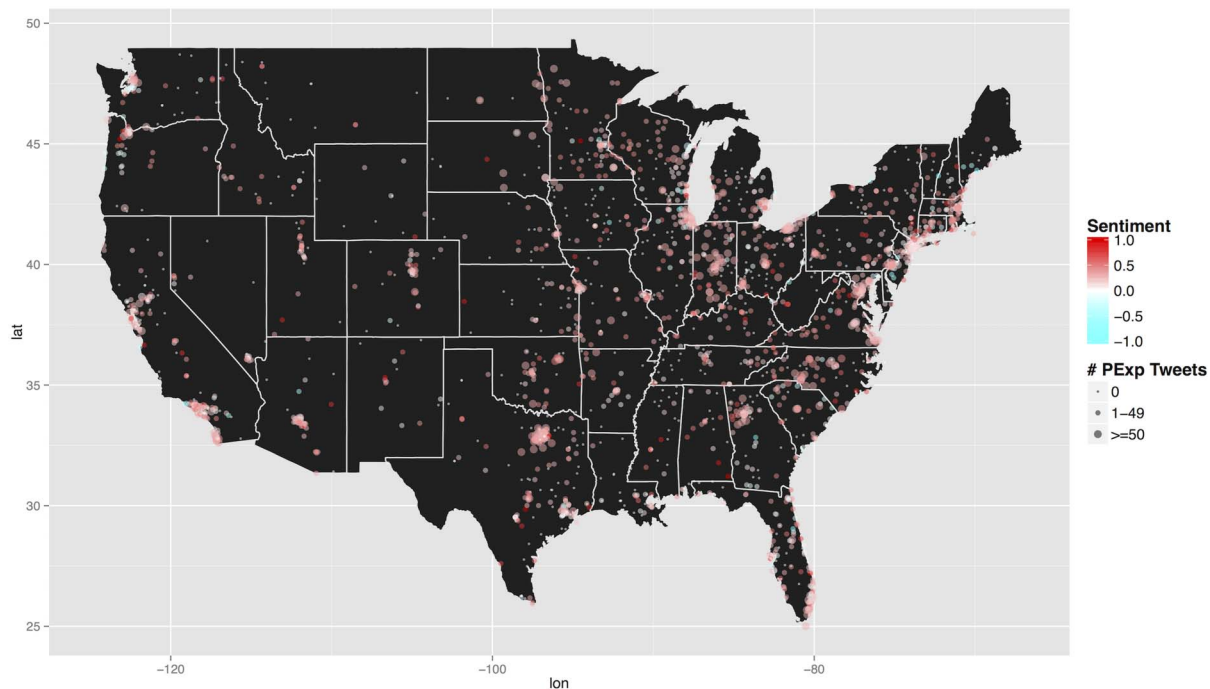
Figure 1 Geographical distribution of all US hospitals on Twitter (n=2349). Hospitals are coloured by mean sentiment, and sized by the number of patient experience tweets received in the 1-year study period. Sentiment ranges from −1 (negative) to 1 (positive).

for all hospitals during the 1-year study period was 43. The median sentiment values for the highest and lowest quartiles were 0.362 and 0.211, respectively. The proportion of hospitals in the profit status (p<0.001) and bed count (p=0.037) categories was significantly different between the highest and lowest sentiment quartiles, with public and larger hospitals over-represented in the lowest sentiment quartile.

We found no correlation between sentiment and Twitter characteristics, except a weak negative correlation (r=−0.18, p=0.002) with total days the account was active. When the highest and lowest quartiles were compared after hospitals had been ranked based on these characteristics, only the total number of

tweets was shown to have an effect on sentiment (p=0.002).

Hospitals with 50+ tweets were more active on Twitter, as they had more posts and followers (p<0.001), but their accounts were not older. In addition, hospitals with more patient experience posts were more likely to be below the national median of Medicare patients (p<0.001), above the national median for nurse/patient ratio (p=0.006), and be a non-profit hospital (p<0.001). Figure 1 shows the geographical distribution of all US hospitals on Twitter, highlighting sentiment and number of patient experience tweets received.

Topic classification
We identified the topics of patient experience that were discussed in a random subset of tweets (table 2). Box 1 includes some specific examples of each topic.

Use of Twitter data by US hospitals
Of the 297 hospitals surveyed about Twitter use, 49.5% responded. All hospitals indicated that they monitored Twitter closely, actively interacted with patients via Twitter, and were aware that patients post about their care experiences. Box 2 includes some additional representative feedback received.

Linking Twitter data to quality of care
In the univariate analysis, we found a significant difference between percentage of people giving an HCAHPS rating of 9 or 10 for hospitals that have a Twitter account compared with those that do not

Table 2 Topic classification

Topic	Count	Ratio of +ve/−ve Tweets	Sentiment median
Discharge	6	0.500	−0.096
Time	313	0.514	−0.150
Side effect	10	0.667	−0.150
Communication	205	0.884	−0.039
Money	222	0.917	−0.028
Pain	37	1.000	−0.007
Room condition	41	1.769	0.140
Medication instructions	10	2.000	0.138
Food	35	2.625	0.250
General	2999	6.734	0.467
Totals	3878	3.762	0.400

Topics are ordered on the basis of the ratio of positive to negative tweets.

Box 1   Patient experience tweets

- Discharge instructions (including care at home)
  - "…epic fail on my TKA discharge"
- Time management
  - "12 hrs in the ER…..not good [hospital]. My father is just now getting a room. #notgoodenough #er #getitright"
- Treatment side effects
  - "Im on 325 mg aspirin coming off blood thinners after blood clott found in lung will this affect my heart??"
- Communication with staff
  - "I know it's Monday/Flu season but waiting 3.5 h for pediatrics to call me back is a little much"
- Money concerns
  - "Hey [hospital], can you hold off on the collections calls until my bill is actually due? Please try to keep it classy."
- Pain management
  - "[hospital] pediatric ER sucks! No doctors to assist, nurses in the back having coffee while there is a sick child in pain in empty ER"
- Conditions of rooms/bathrooms
  - "what do you have against coat hooks? 3 exam rooms, 2 locations in 1 day and not 1 place to hang my coat and bag!"
- Medication instructions
  - "[hospital] gave my mom a prescription for a discontinued medicine. #Silly hospital. @Walgreens saved the day!"
- Cafeteria food
  - "skipped dinner last night because of your terrible cafeteria food. Are you trying to get more patients? #eathealthy #healthcare"
- General satisfaction/dissatisfaction with procedure and/or staff
  - "I'm thankful for the [hospital], their staff, my Doctors and to be treated as a person, not just a patient."

Box 2   Hospital responses regarding their use of patient Twitter posts

- "[O]ur goal is to respond to patient comments within an hour of their post. If the comment can be addressed via Twitter, we direct them to appropriate resources online. We are extremely careful to abide by HIPAA guidelines and the protection of patient privacy."
- "We also do social listening to find out what topics are important to our patient families and supporters so we can take a proactive approach and participate in the conversation by providing information and help."
- "We've had patients tweet us from waiting rooms, we've even had patients tweet us from their hospital beds!"
- "We proactively respond to all social media conversations that mention us. We respond then try to facilitate a one-on-one email or phone conversation with the patient or caregiver to discuss their experience."
- "We utilize geocoding to narrow tweets to within .25 kilometers of each facility to capture any tweets that don't expressly mention [hospital network name], but are related to their health care experience at a facility."
- "If there is a serious complaint, we send those along to Patient Relations who may follow up outside of social media. We typically don't get into a back and forth around a negative comment, but rather let the person know we are sorry for their experience, and then direct them to patient relations."

(0.71 vs 0.69, p=0.001) and between HCAHPS rating in the highest versus lowest quartiles of hospitals ranked by the number of Twitter followers (0.72 vs 0.69, p<0.001). In addition, there was a significant difference in sentiment in the highest versus lowest quartiles of hospitals ranked on HCAHPS (0.30 vs 0.26, p=0.017). However, after adjusting for potential confounders (see Methods) through multivariate linear regression, we did not find any significant correlation between HCAHPS and any of these metrics.

We observed a correlation between 30-day readmission rates and sentiment. There is a weak negative correlation (r=−0.215, p<0.001), where higher sentiment scores are associated with lower 30-day

readmission rates (figure 2). In addition, there was a small but significant difference (p=0.014) between the 30-day readmission rates in the highest versus lowest quartiles of hospitals ranked on sentiment. Finally, after adjustment for hospital and Twitter characteristics using multivariate linear regression, there was still a small but significant association between higher sentiment scores and lower 30-day readmission rates (table 3; p=0.003).

DISCUSSION

Our findings indicate that patients use Twitter to provide feedback about the quality of care they receive at US hospitals. We found that approximately half of the hospitals in the USA have a presence on Twitter and that sentiment towards hospitals was, on average, positive. Of the 297 surveyed, half responded and all confirmed that they closely monitor social media and interact with users. We therefore conclude that the stakeholders of these hospitals see the value of capturing information on the quality of care in general, and patient experience in particular.
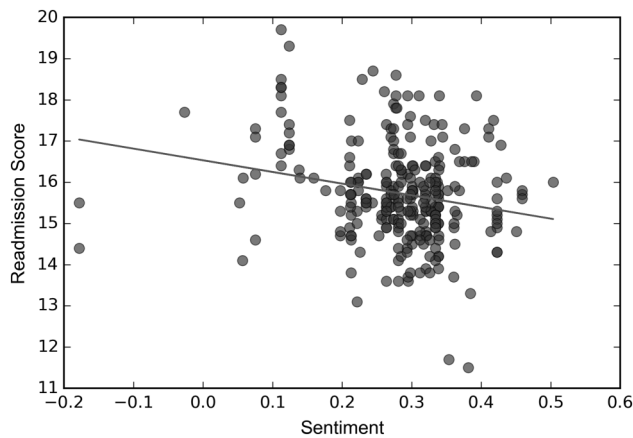
Figure 2 Sentiment correlated with 30-day hospital readmission rates. 30-day hospital readmission rates are plotted against average sentiment, for hospitals that have ≥50 patient experience tweets (n=297). This association displays a weak negative correlation (r=−0.215, p<0.001).

Surprisingly, we found only a weak association with one measure of hospital quality (30-day readmission), but not with an established standard of patient experience (HCAHPS). Taken together, our findings suggest that Twitter is a unique platform to engage with patients and to collect potentially untapped feedback —and possibly a useful measure for supplementing traditional approaches of assessing and improving quality of care.

Our findings on the extent of Twitter usage by hospitals are similar to what has been reported previously.[43] The generally positive sentiment on Twitter is consistent with other analyses that suggest a positive language bias on social media.[44] However, our analysis of Twitter sentiment, and exploring the association with conventional quality measures, is novel. There were some striking differences between the hospitals with the highest and lowest sentiment, with both large and public hospitals being over-represented in the lowest quartile. In addition, the number of tweets a hospital received influenced, in part, hospital sentiment; hospitals that received more tweets had, on average, higher sentiment. However, the number of tweets a hospital posted did not affect sentiment. Thus, having a more active online presence with a frequent posting behaviour is not sufficient to increase sentiment alone, although we did find that it increased

Table 3 Sentiment associated with 30-day readmission rates

Mean sentiment	30-day readmission rate	30-day readmission rate (adjusted score)	p Value
Lowest quartile	16.130	16.876	Ref
Second quartile	15.859	16.937	0.799
Third quartile	15.417	16.249	0.009
Highest quartile	15.534	16.163	0.009
p Value for trend		0.003	

the likelihood of receiving more patient experience tweets.

Twitter feedback is entirely unsolicited. As such, there was a wide range of patient experience topics discussed. These topics include those covered by the HCAHPS survey and previous research,[45] as well as some not typically addressed (eg, time, side effects, money and food concerns). It is not surprising that some topics tended to be more negative than others— for example, discussion of time, money or pain is not likely to be positive. Thus, from an individual hospital's perspective, it might not be useful to heavily weight the number of positive or negative tweets within one topic category at any one moment. However, monitoring these topics over time and detecting when sentiment goes above or below an established baseline could be useful.

We used both HCAHPS scores and 30-day hospital readmission rates as conventional measures of quality of care to compare against. Readmission rate was only one of several metrics we could have used to compare against HCAHPS; other measures such as mortality and Hospital Compare metrics could also be analysed. While there are conflicting studies on this association with readmission rates and it is disputed by some,[46-49] they have been used in this way before, including recent studies that showed correlation with ratings on Facebook[15] and Yelp.[18] We report associations between Twitter sentiment and readmission rates to evaluate the potential of this relationship and found a weak negative correlation, with higher-sentiment hospitals having a lower readmission rate. This association survived adjustment for potential confounders, with a small but significant downward trend for readmission rates as sentiment increases. Nonetheless, we acknowledge that the observed correlation was weak at best and probably influenced by confounding factors. Importantly, no association was observed between sentiment and HCAHPS score, after adjustment for hospital characteristics. This finding of only a weak association with a clinical metric, and no association with the more easily explainable alternative patient experience metric, suggests that Twitter sentiment must be treated cautiously in understanding quality. The use of Twitter data as we have in this analysis is in its infancy, and therefore development of methodologies to compare against traditional measures of patient experience is warranted. However, our findings suggest that there is new information here that hospital administrators may want to listen carefully to.

There were several limitations to our study. First, while the use of Twitter is becoming more pervasive in the older population, users under 30 years of age are the largest group, indicating there is a selection bias. Second, we only looked at tweets that explicitly included a hospital's Twitter handle. Broadening our criteria to include hospital names as keywords or

attempting to assign tweets to nearby hospitals given geospatial data could have potentially increased the number of patient experience tweets we identified. In addition, many hospitals within a larger network shared a Twitter account and, without additional follow-up, it is difficult to determine which hospital is being discussed. Like all surveys, our hospital questionnaire may have been subject to a potential response and selection bias. Owing to the cross-sectional design, while we have shown association between organisations that use Twitter and their interactions with patients, we cannot confirm any causal relationship. Further investigation of how these findings change over time would be helpful. Finally, while patient experience classification had relatively high agreement rates and inter-rater κ values, topic classification only had an overall agreement of 64.7%. In addition, some of the topics had a low κ value. This is probably an effect of using crowd-sourced curators without a high level of domain-specific training, which also explains why 77.3% of patient experience tweets were non-specifically labelled as 'General'. As for our automated approach, machine classification and sentiment analysis using NLP does not perform as well as human curation. With these caveats acknowledged, our approach enabled processing of an extremely large amount of data and illustrated that automated analysis of Twitter data can provide useful, unsolicited information to hospitals across a wide variety of patient experience topics.

Our findings have implications for various groups. Hospital administrators and clinicians should consider actively monitoring what their patients are saying on social media. Institutions that do not use Twitter should create accounts and analyse the data, while existing users might consider leveraging automated tools. Insight from key leaders at institutions will help to better understand gaps and potential opportunities. Regulators should continue to consider social media commentary as a supplemental source of data about care quality.[50] The information is plentiful and, although the techniques for processing and understanding these data are still being developed and improved, potentially important. We recommend a larger survey in the USA and globally with all relevant stakeholders, including patients and their families, to obtain a better understanding of the use and value of social media for patient interactions. The public should pay attention to what other people are tweeting and posting on social media, and systems to collect, aggregate and summarise this information for a public audience in real-time should be considered to complement data from traditional reporting platforms. To increase the utility of these data, we would recommend that each hospital manages their own unique Twitter identity, rather than share an account across a larger healthcare network.

CONCLUSIONS

We show that monitoring Twitter provides useful, unsolicited, and real-time data that might not be captured by traditional feedback mechanisms—Twitter sentiment only weakly correlates with readmission rates but not HCAHPS ratings, as would be expected. While many hospitals monitor their own Twitter feeds, we recommend that patients, researchers and policy makers also attempt to utilise this data stream to understand the experiences of healthcare consumers and the quality of care they receive.

Author affiliations
[1]Center for Biomedical Informatics, Harvard Medical School, Boston, USA
[2]Informatics Program, Boston Children's Hospital, Boston, USA
[3]Department of Pediatrics, Harvard Medical School, Boston, USA
[4]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA
[5]Computer Science, Virginia Tech, Blacksburg, Virginia, USA
[6]Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA
[7]Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[8]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
[9]Department of Primary Care and Public Health, Imperial College London, London, UK

REFERENCES

1 Committee on Quality of Health Care in America, Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st Century*. Washington DC, 2001.

2   Rozenblum R, Miller P, Pearson D, *et al*. Patient-centered healthcare, patient engagement and health information technology: the perfect storm. In: Grando MA, Rozenblum R, Bates DW, eds. *Information technology for patient empowerment in healthcare*. 1st edn. Berlin: Walter de Gruyter GmbH & Co. KG, 2015:3–22.

3   Zimlichman E. Using patient reported outcomes to drive patient-centered care. In: Grando MA, Rozenblum R, Bates DW, eds. *Information technology for patient empowerment in healthcare*. 1st edn. Berlin: Walter de Gruyter GmbH & Co. KG, 2015:241–56.

4   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.

5   US Government. The Patient Protection and Affordable Care Act. 2009.

6   Goldstein E, Farquhar M, Crofton C, *et al*. 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.

7   Darby C, Hays RD, Kletke P. Development and evaluation of the CAHPS hospital survey. *Health Serv Res* 2005;40(6 Pt 2):1973–6.

8   Zimlichman E, Rozenblum R, Millenson ML. The road to patient experience of care measurement: lessons from the United States. *Isr J Health Policy Res* 2013;2:35.

9   Duggan M, Nicole B. Cliff Lampe E, *et al*. Social Media Update 2014. Secondary Social Media Update 2014, 9 January 2015; 2015. http://www.pewinternet.org/2015/01/09/social-media-update-2014/

10 Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Health Affairs* 2009;28:361–8.

11 Hader AL, Brown ED. Patient privacy and social media. *AANA J* 2010;78:270–4.

12 Moses RE, McNeese LG, Feld LD, *et al*. Social media in the health-care setting: benefits but also a minefield of compliance and other legal issues. *Am J Gastroenterol* 2014;109:1128–32.

13 Huesch MD, Currid-Halkett E, Doctor JN. Public hospital quality report awareness: evidence from National and Californian Internet searches and social media mentions, 2012. *BMJ open* 2014;4:e004417.

14 Wong CA, Sap M, Schwartz A, *et al*. Twitter sentiment predicts Affordable Care Act marketplace enrollment. *J Med Internet Res* 2015;17:e51.

15 Glover M, Khalilzadeh O, Choy G, *et al*. Hospital evaluations by social media: a comparative analysis of Facebook ratings among performance qutliers. *J Gen Intern Med* 2015;30:1440–6.

16 Wallace BC, Paul MJ, Sarkar U, *et al*. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. *J Am Med Inform Assoc* 2014;21:1098–103.

17 Greaves F, Laverty AA, Cano DR, *et al*. Tweets about hospital quality: a mixed methods study. *BMJ QualSaf* 2014;23:838–46.

18 Verhoef LM, Van de Belt TH, Engelen LJ, *et al*. Social media and rating sites as tools to understanding quality of care: a scoping review. *J Med Internet Res* 2014;16:e56.

19 McCaughey D, Baumgardner C, Gaudes A, *et al*. Best practices in social media: utilizing a value matrix to assess social media's impact on health care. *Soc Sci Comput Rev* 2014;32:575–89.

20 Timian A, Rupcic S, Kachnowski S, *et al*. Do patients "like" good care? measuring hospital quality via Facebook. *Am J Med Qual* 2013;28:374–82.

21 Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? *BMJ Qual Saf* 2013;22:183–6.

22 Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. *N Engl J Med* 2009;360:2153–5, 57.

23 You are what you Tweet: Analyzing Twitter for public health. ICWSM; 2011.

24 S A, BA H. Predicting the Future with Social Media. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) 2010:492–9.

25 Coviello L, Sohn Y, Kramer AD, *et al*. Detecting emotional contagion in massive social networks. *PLoS ONE* 2014;9: e90315.

26 Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 2011;333:1878–81.

27 Haney C. Sentiment Analysis: Providing Categorical Insight into Unstructured Textual Data.

28 Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations Trends Info Retrieval* 2008;2:1–135.

29 A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse. AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs; 2006.

30 Hao Wang, Dogan Can, Abe Kazemzadeh, *et al*. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 115–120, Jeju, Republic of Korea, 8–14 July 2012*. Association for Computational Linguistics, 2012.

31 Tumasjan A, Sprenger TO, Sandner PG, *et al*. Predicting elections with twitter: what 140 characters reveal about political sentiment. *ICWSM* 2010;10:178–85.

32 Ceron A, Curini L, Iacus SM, *et al*. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media Soc* 2014;16:340–58.

33 Jansen BJ, Zhang M, Sobel K, *et al*. Twitter power: Tweets as electronic word of mouth. *J Am Soc Info Sci Technol* 2009;60:2169–88.

34 Ghiassi M, Skinner J, Zimbra D. Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems Appl* 2013;40: 6266–82.

35 Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behav Res Methods* 2014;46:1023–31.

36 Barr J. Get Better Results with Amazon Mechanical Turk Masters. Secondary Get Better Results with Amazon Mechanical Turk Masters, 2011. https://aws.amazon.com/blogs/aws/amazon-mechanical-turk-master-workers/

37 Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–3.

38 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.

39 Loria S. TextBlob: simplified text processing. Secondary TextBlob: Simplified Text Processing 2014. https://textblob.readthedocs.org/en/dev/index.html.

40 Center) CCLPR. pattern.en Documentation. Secondary pattern.en Documentation 2015. http://www.clips.ua.ac.be/pages/pattern-en

41 Medicare.gov. 30-day unplanned readmission and death measures. Secondary 30-day unplanned readmission and death

measures, 2015. http://www.medicare.gov/hospitalcompare/Data/30-day-measures.html?AspxAutoDetectCookieSupport=1

42 Centers for Medicare and Medicaid Services. *Frequently Asked Questions: CMS Publicly Reported Risk-Standardized Outcome and Payment Measures. Secondary Frequently Asked Questions: CMS Publicly Reported Risk-Standardized Outcome and Payment Measures*, 2015. http://www.qualitynet.org/dcs/ContentServer?cid=1228774724512&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page

43 Griffis MH, Kilaru SA, Werner MR, *et al*. Use of social media across US hospitals: descriptive analysis of adoption and utilization. *J Med Internet Res* 2014;16:e264.

44 Dodds PS, Clark EM, Desu S, *et al*. Human language reveals a universal positivity bias. *Proc Natl Acad Sci USA* 2015;112:2389–94.

45 Kleefstra SM, Zandbelt LC, de Haes HJ, *et al*. Trends in patient satisfaction in Dutch university medical centers:

room for improvement for all. *BMC Health Serv Res* 2015;15:112.

46 Tsai TC, Joynt KE, Orav EJ, *et al*. Variation in surgical-readmission rates and quality of hospital care. *N Engl J Med* 2013;369:1134–42.

47 Kansagara D, Englander H, Salanitro A, *et al*. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;306:1688–98.

48 Polanczyk CA, Newton C, Dec GW, *et al*. Quality of care and hospital readmission in congestive heart failure: an explicit review process. *J Card Fail* 2001;7:289–98.

49 Ashton CM, Kuykendall DH, Johnson ML, *et al*. The association between the quality of inpatient care and early readmission. *Ann Intern Med* 1995;122:415–21.

50 van de Belt T, Engelen L, Verhoef L, *et al*. Using patient experiences on Dutch social media to supervise health care services: exploratory study. *J Med Internet Res* 2015;17:e7.

BMJ Quality & Safety

Measuring patient-perceived quality of care in US hospitals using Twitter

Jared B Hawkins, John S Brownstein, Gaurav Tuli, Tessa Runels, Katherine Broecker, Elaine O Nsoesie, David J McIver, Ronen Rozenblum, Adam Wright, Florence T Bourgeois and Felix Greaves

Updated information and services can be found at:
http://qualitysafety.bmj.com/content/25/6/404

These include:

References

This article cites 35 articles, 8 of which you can access for free at:
http://qualitysafety.bmj.com/content/25/6/404#BIBL

Open Access

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections

Articles on similar topics can be found in the following collections

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/