



OPEN ACCESS

Wisdom of patients: predicting the quality of care using aggregated patient feedback

Alex Griffiths,¹ Meghan P Leaver²

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2017-006847>).

¹Centre for Analysis of Risk and Regulation, London School of Economics and Political Science, London, UK

²Department of Psychological and Behavioural Science, London School of Economics and Political Science, London, UK

Correspondence to

Dr Alex Griffiths, London School of Economics and Political Science, Centre for Analysis of Risk and Regulation, London WC2A 2AE, UK; a.griffiths@lse.ac.uk

Received 29 April 2017

Revised 12 August 2017

Accepted 23 August 2017

Published Online First

28 September 2017



- <http://dx.doi.org/10.1136/bmjqs-2016-006291>
- <http://dx.doi.org/10.1136/bmjqs-2017-007149>
- <http://dx.doi.org/10.1136/bmjqs-2017-007151>



To cite: Griffiths A, Leaver MP. *BMJ Qual Saf* 2018;**27**:110–118.

ABSTRACT

Background The Care Quality Commission (CQC) is responsible for ensuring the quality of healthcare in England. To that end, CQC has developed statistical surveillance tools that periodically aggregate large numbers of quantitative performance measures to identify risks to the quality of care and prioritise its limited inspection resource. These tools have, however, failed to successfully identify poor-quality providers. Facing continued budget cuts, CQC is now further reliant on an 'intelligence-driven', risk-based approach to prioritising inspections and a new effective tool is required.

Objective To determine whether the near real-time, automated collection and aggregation of multiple sources of patient feedback can provide a collective judgement that effectively identifies risks to the quality of care, and hence can be used to help prioritise inspections.

Methods Our Patient Voice Tracking System combines patient feedback from NHS Choices, Patient Opinion, Facebook and Twitter to form a near real-time collective judgement score for acute hospitals and trusts on any given date. The predictive ability of the collective judgement score is evaluated through a logistic regression analysis of the relationship between the collective judgement score on the start date of 456 hospital and trust-level inspections, and the subsequent inspection outcomes.

Results Aggregating patient feedback increases the volume and diversity of patient-centred insights into the quality of care. There is a positive association between the resulting collective judgement score and subsequent inspection outcomes (OR for being rated 'Inadequate' compared with 'Requires improvement' 0.35 (95% CI 0.16 to 0.76), Requires improvement/Good OR 0.23 (95% CI 0.12 to 0.44), and Good/Outstanding OR 0.13 (95% CI 0.02 to 0.84), with $p < 0.05$ for all).

Conclusions The collective judgement score can successfully identify a high-risk group of organisations for inspection, is available in near real time and is available at a more granular level than the majority of existing data sets. The collective judgement score could therefore be used to help prioritise inspections.

INTRODUCTION

The Care Quality Commission (CQC) is the independent regulator responsible for ensuring the quality of over 30 000 health and social care providers in England.¹ This includes 150 acute and

specialist trust—administrative groupings of one or more hospitals in a geographical region—that provide the majority of hospital-based care for the National Health Service (NHS).² Rather than adopt a random or cyclical approach to prioritising its limited inspection resource, CQC is legally required to operate an intelligence-led, risk-based approach.³ With its annual budget to be cut by £32 million (13%) by 2019,⁴ CQC has stated it plans to further rely on data to 'target our resources where the risk to the quality of care provided is greatest'.⁵

To support its risk-based approach, CQC has relied on a succession of statistical surveillance tools aggregating large volumes of trust-level performance data. Based on the Healthcare Commission's surveillance tool, 'Quality and Risk Profiles' (QRPs) aggregated hundreds of weighted, z-scored quantitative indicators and manually-coded qualitative comments, each mapped to one of CQC's 16 'essential standards' of care.⁶ QRPs automatically generated a risk score for each of the 16 'essential standards' and were updated simultaneously for all trusts nine times a year. Although comprehensive, QRPs faced heavy criticism for the quality and timeliness of information they contained, their complexity and their inability to achieve their core function—to effectively predict risks to the quality of care.^{7 8}

Following the critical Francis Inquiry into the high-profile failings at Mid-Stafordshire Trust, CQC overhauled its regulatory approach including its statistical surveillance tool.⁹ QRPs were replaced by the far simpler 'Intelligent Monitoring' (IM) tool. IM generated a single trust-level 'risk score' based on approximately 150 equally weighted performance indicators that were simultaneously updated for all trusts every 5 months.¹⁰ Two sources

of patient feedback data were included in the final version of IM, the proportion of 'Share Your Experience' comments submitted via CQC's website that were manually coded as 'negative',¹¹ and standardised scores from select questions on the annual inpatient survey. IM was unable to successfully identify high-risk trusts with its predictions proving wrong more often than not.¹² Like QRPs, concerns were expressed over the quality, timeliness and granularity of information IM contained.^{13–15} CQC has now begun developing its 'CQC Insight' surveillance tool to replace IM.¹⁵

One widely supported solution to identifying risks to the quality of care is to make use of abundant patient feedback in the NHS.^{8 16–20} In addition to providing a different perspective on quality to traditional metrics, such as readmission rates and staffing levels,^{21 22} patient feedback from online rating pages NHS Choices and Patient Opinion, and via Twitter and Facebook, is available in a more timely manner (near real time), and at a more granular (hospital) level, than the data typically used by CQC. Patient feedback therefore has the potential to help address the perennial concerns over the timeliness and granularity of data used to identify risks to the quality of care.

Timely and granular information, however, will be of little use if it does not reflect the quality of care. Existing research into the use of patient feedback to identify quality concerns has compared individual sources with quality measures at a fixed point in time and has shown encouraging, if mixed, results. In the UK, an association has been demonstrated between NHS Choices ratings, Inpatient Survey scores and select clinical outcome measures.²³ No association has been found between the sentiment of tweets mentioning NHS trusts and mortality rates or NHS Inpatient Survey scores; however, the study analysed the sentiment of all tweets whether they related to the quality of care or not.²⁴ In the USA, an association was found between Yelp ratings, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, and to a lesser extent with select clinical outcomes.²⁵ Furthermore, an association between standardised Facebook ratings and readmission rates was identified.²⁶ A weak association was found between 30-day readmission rates and the sentiment of care quality-related tweets, although no association was found between sentiment and the HCAHPS survey.²⁷

Individual sources of patient feedback being biased towards certain demographics were identified as a barrier to the more effective measurement of quality in each of these studies. One possible way to overcome the bias of individual sources of patient feedback is to combine them. In the UK, Twitter has 15.8 million active users, 7.9 million (50%) of whom are aged below 35, predominantly of higher socioeconomic status. Facebook has 37.5 million active users, 21 million (56%) of whom are aged 35 or over, with

users of a lower socioeconomic status over-represented.^{28 29} Demographics data are not available for NHS Choices; however, the wider site, which includes health information, received 583 million unique visits in 2015, nearly 10 visits for every member of the UK.³⁰ Combining these high-volume, disparate sources of patient feedback in near real time is technically challenging however, and has yet to be successfully operationalised.

The aim of our study is to determine whether the near real-time, automated collection and aggregation of patient feedback can provide a collective judgement that effectively identifies risks to the quality of care, and hence can be used to help prioritise CQC inspections. This study furthers the existing research in a number of ways. Unlike previous research, this combines multiple sources of patient feedback, it looks at the more granular, hospital-level feedback, and it measures the association between patient feedback and other quality measures at hundreds of points over more than 3 years, rather than a fixed point in time. Moreover, the patient feedback used is contemporary having occurred within 90 days of the start of the inspection, and covers a greater volume and diversity of trusts and hospitals as a result of increased engagement with social media by the NHS.

METHODS

Our analysis measures the statistical relationship between a time-limited collective judgement score (CJS) formed of patient feedback from multiple sources at the start of comprehensive CQC inspections, and the subsequent outcome of those inspections. The data sets and methods for this analysis are detailed below.

Dependent variable

In October 2013, CQC introduced a new comprehensive inspection regime and by March 2017 had visited all NHS hospitals and trusts in England.³¹ Under that regime, inspections are conducted by large teams of specialist inspectors, clinicians and 'experts by experience' who assess individual hospital services against five 'key questions': is the service 'safe', 'effective', 'caring', 'responsive to people's needs' and 'well led'?³² Based on their inspection teams' on-site visits, CQC then awards one of four possible ordinally ranked ratings for each core service within a hospital:

1. Outstanding
2. Good
3. Requires improvement
4. Inadequate

CQC then aggregates those service-level ratings using an unpublished, rules-based approach to assign hospital-level ratings, which, in turn, are further aggregated, using a similar rules-based approach, to generate an overall trust-level rating.

Table 1 Summary of the three sources of patient feedback used to form the collective judgement score

	NHS Choices	Facebook	Twitter
Time period data available	1 January 2013 to 12 March 2017	1 January 2013 to 12 March 2017	21 February 2016 to 12 March 2017
Total number of comments collected	76 493	69 427	1 303 085
Unique comments suitable for study	76 493	69 427	20 914
Unique comments suitable for study covering 1 March 2016 to 28 February 2017	20 270	19 572	19 771
CQC-rated hospitals with an account/page	245	204	13
CQC-rated trusts with an account/page	148	132	142
Mean sentiment score (from 1 to 5)	3.85	4.13	4.28

CQC, Care Quality Commission.

Up to 12 March 2017, CQC published 204 and 339 trust and hospital-level ratings, respectively. The majority of trusts and hospitals have been rated only once under the new inspection regime. Those inspected more than once were typically, although not always, rated poorly on their first inspection and have since been inspected again to check for improvements. A minority of inspections contained no overall rating. This occurred when CQC deemed there was insufficient evidence to produce a rating, or where the inspection team focused on a specific aspect of care and therefore could not judge the overall quality of care.

Independent variables

We obtained details of all NHS acute and specialist trusts and their hospitals, including the URL for their patient feedback submitted to NHS Choices, from the NHS Choices application programming interface (API) in February 2016. We then searched Google for each organisation’s Twitter and Facebook details. Where no page was found, the organisation’s own website was checked. The details for each data source and how they were combined are described below and in [table 1](#).

NHS Choices (including Patient Opinion)

NHS Choices is a government-run website that captures unsolicited feedback for all NHS trusts and hospitals. Users can leave free-text feedback and rate how likely they are to recommend the organisation to friends or family in need of similar care on a scale from 1 (extremely unlikely) to 5 (extremely likely) stars. Patient Opinion is a social enterprise which provides a similar function to NHS Choices including the ability to leave qualitative feedback and recommend the service on the same scale. Feedback posted to Patient Opinion is shared with, and appears on, the NHS Choices website. Comments left on either site are actively moderated before being displayed with references to individuals and speculation removed.^{33–35} All patient feedback posted between 1 January 2013 and 12 March 2017 inclusive was sourced via NHS Choices’ API.

Facebook

Facebook directly captures patient feedback in two ways. First, organisations can create official Facebook pages with a ‘Reviews’ section. Anyone accessing the page can leave a review, which must contain a rating from 1 to 5 stars with additional free-text feedback being optional. Second, organisations that users frequently search for automatically have an unofficial page generated that includes a similar review function to official pages. Comments for both page types are reactively moderated; once posted, comments cannot be amended and will only be removed if they breach Facebook’s Community Standards.³⁶ We sourced all reviews posted from 1 January 2013 to 12 March 2017 using a custom Python script.

Twitter

Starting in February 2016, we queried the Twitter API at regular intervals of no more than 7 days to ensure complete coverage from the time and volume-limited data source. Unlike NHS Choices and Facebook data, not all comments related to the quality of care. To ensure only original feedback was included, and that the results were not biased by the popularity of users, all retweets and tweets generated by trusts or hospitals were removed.

Table 2 The count and reliability of each category of tweet coded by both AG and MPL

Tweet classification	N	Inter-rater reliability (Cohen’s κ)	p Value
Patient experience and effectiveness	348	0.64	<0.001
Environment and facilities	67	0.67	<0.001
Timeliness and access	90	0.75	<0.001
Safety	13	0.45	<0.001
Not patient feedback concerning the quality of care	5351	0.67	<0.001
Total	5869		

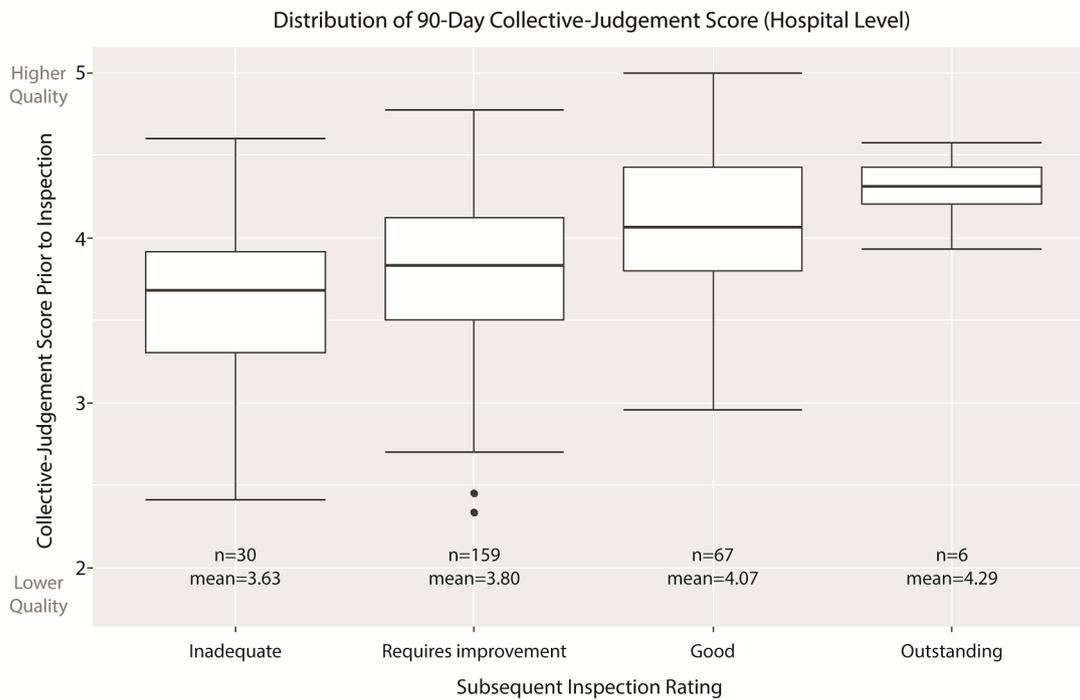


Figure 1 Box plot of collective judgement scores grouped by inspection rating.

After 6 months of data had been collected, a representative pilot data set of 1000 tweets was jointly coded by AG and MPL and the five classifications of tweet detailed in table 2 were agreed. Tweets that were clearly not original feedback from patients or their friends and family were classed as ‘Not patient feedback concerning the quality of care’ to prevent staff and promotional activities biasing the data set. A second representative sample comprising 10% (23 014) of all tweets, not retweeted or generated by the organisation in question, was then coded by AG, with just over 25% (5869) of AG’s coding second-coded by MPL to check reliability. Inter-rater reliability was ‘substantial’ ($\kappa=0.6-0.8$) for all classifications except for the rarest, ‘safety’, where it was ‘moderate’ ($\kappa=0.4-0.6$).³⁷ All differently coded tweets were then jointly reviewed and complete agreement was reached.

The next stage was to automate the classification of the tweets. We first oversampled the data to account for the class imbalance in care-quality related tweets^{38 39} and set aside 25% of the data set for model testing. We then developed a variety of models using fivefold cross-validation and a tuning grid of model-specific parameters to classify the tweets. The chosen model achieved overall precision, recall and F_1 scores of 0.97.⁴⁰

Tweets that were not classed as patient feedback concerning the quality of care were then excluded from the study and the sentiment of the remaining tweets was scored using the same 1 to 5-star scale as the NHS Choices and Facebook data. Initial attempts to use standard sentiment dictionaries such as SentiWordNet

proved ineffective, likely due to the healthcare-specific lexicon and non-prose style of tweets. Instead, the same modelling approach used for classifying the tweets was used to identify sentiment. The chosen model predicted the 1 to 5-star rating of scored comments with 79% accuracy, rising to 91% accuracy when estimating the sentiment to within 1 star, and was used to assign a star rating to all care-quality related tweets. Further details of the modelling are available in the online supplementary technical appendix.

Table 3 The regression coefficients and associated SEs, ORs and associated 95% CIs and p values for the hospital-level ordinal (cumulative) logistic regression model

	Beta (SE)	95% CI for OR			Pr(> z)
		2.5%	OR	97.5%	
Inadequate (intercept)	1.89 (1.47)				0.199
Requires improvement (intercept)	6.74 (1.31)				0.000
Good (intercept)	12.00 (4.05)				0.003
Inadequate (CJS)	-1.04 (0.40)	0.16	0.35	0.77	0.008
Requires improvement (CJS)	-1.47 (0.33)	0.12	0.23	0.44	0.000
Good (CJS)	-2.02 (0.94)	0.02	0.13	0.84	0.032

CJS, collective judgement score.

Table 4 A contingency table showing the number of hospitals and trusts that would have been inspected, and the outcome of those inspections, had the 50 trusts or 80 hospitals with the lowest 90-day collective judgement score been inspected

		Inadequate	Requires improvement	Good	Outstanding	Total
Hospital level	Higher priority (inspect)	14	56	10	0	80
	Lower priority (cannot inspect)	16	103	57	6	182
	Total	30	159	67	6	262
Trust level	Higher priority (inspect)	9	34	7	0	50
	Lower priority (cannot inspect)	13	83	41	9	146
	Total	22	117	48	9	196

Combining and analysing the data

With a date and score ranging from 1 (lowest) to 5 (highest) stars associated with every item of patient feedback, a CJS could be calculated for all trusts and hospitals on any given date where sufficient patient feedback was available. A 90-day moving average was chosen as a balance between the CJS reflecting the more recent patient feedback, and being too volatile and overly influenced by a small number of recent items of patient feedback leading to unnecessary inspections being triggered. The 90-day CJS on the start date of each trust and hospital-level inspection was then paired with the overall rating from the inspection. Of the 543 trusts and locations rated by CQC, 87 were then removed from the data set as their 90-day CJS comprised fewer than 10 items of patient feedback. The majority of organisations removed were specialist units within a trust.

As trust-level ratings are an aggregation of hospital-level ratings they are not independent, and combining them for a single model would violate the assumptions of the regression model. The relationships between the 90-day CJS at the start of an inspection and the subsequent trust and hospital-level inspection ratings were therefore assessed using two distinct ordinal logistic regression models both of which used random intercepts and random coefficients.

RESULTS

Figure 1 shows the distribution of the 90-day CJS on the first day of hospital-level inspections grouped by the subsequent overall rating. For each improved inspection rating, the mean and median CJS increases, indicating that on average patient feedback is better for hospitals the CQC subsequently awards a higher rating. The better the patient feedback in the 90 days prior to a CQC inspection, the greater the likelihood of a more positive overall rating.

The ordinal (cumulative) regression model detailed in table 3 describes the probability of a hospital being awarded a prescribed CQC rating given a particular CJS score and confirms the statistical significance of the association between increasing CJS and improved inspection ratings. The same significant relationship identified for hospitals is also present for trusts when considered in isolation. Further details about the

regression model can be found in the online supplementary technical appendix.

The overlapping CJS across all four rating categories show that, despite the tendency for hospitals awarded a higher rating to have a higher 90-day CJS, it is not a perfect predictor of the outcome of inspections. A small number of 'Inadequate' hospitals have a high CJS, as do a greater number of 'Requires improvement' hospitals. Conversely, all 'Outstanding' hospitals have an above average CJS, suggesting patients are better collective judges of organisations performing well than they are of organisations performing poorly.

As of 12 March 2017, CQC had published 55 trust-level and 88 hospital-level inspection reports that contained an overall rating and where the inspection began in 2016. With CQC adopting a more targeted approach, and facing budget cuts, the number of inspections it conducts is likely to fall. Were CQC to prioritise the 50 trusts or 80 hospitals with the lowest CJS in the data set the outcome would be as described in table 4.

For hospitals, 70 out of the 80 inspections would result in a rating of 'Inadequate' or 'Requires improvement'. Moreover, only 10 of the 67 'Good' hospitals and none of the 6 'Outstanding' hospitals would be unnecessarily burdened with inspection. Similarly, 43 out of the 50 inspections of trusts would result in a rating of 'Inadequate' or 'Requires improvement' and only 7 of the 48 'Good' trusts and none of the 9 'Outstanding' trusts would be unnecessarily burdened with inspection. This represents a precision rate (the proportion of high-risk organisations that were subsequently rated as 'Inadequate' or 'Requires improvement') of 87.5% and 86% for hospitals and trusts, respectively. The issue for CQC is that, even with a tool that was 100% precise, the constrained number of inspections would still mean that 109 out of the 189 hospitals that were rated 'Requires improvement' or 'Inadequate' would not be inspected. In the unlikely event CQC's capacity for inspection were to double, the CJS would remain an effective tool with precision rates of 81% for both hospital and trust-level inspections. This is a meaningful improvement on the 72% and 71% that would be achieved at hospital and trust levels, respectively, by chance alone.

DISCUSSION

Statement of principal findings

The study establishes that the near real-time, automated collection and aggregation of patient experiences from multiple sources, including social media, can provide a collective judgement that effectively identifies a high-risk group of organisations, and hence can be used to help prioritise inspections.

Policy implications

No combination of indicators, quantitative or qualitative, will ever perfectly predict the outcome of inspections. If they could, there would be little need for expensive inspections. Yet CQC does not have the resource to inspect each hospital every year. It therefore prioritises inspections as best it can with the data that are available. The CJS can support this data-driven approach in three ways.

First, while the collated patient feedback cannot perfectly identify poorly performing organisations, it can identify those organisations that are *most likely* to be performing poorly. Indeed, not a single organisation with a CJS below 2.95 was rated ‘Good’ or ‘Outstanding’. Moreover, 89 out of 130 (68%) ‘Good’ or ‘Outstanding’ organisations had an above average CJS at the start of their inspection, compared with 147 out of 328 (44%) ‘Requires improvement’ or ‘Inadequate’ organisations. Second, much of the patient feedback is available at hospital level and can therefore be used to support more focused inspections than the trust-level data currently used.

Third, unlike CQC’s statistical surveillance tools that comprised quarterly or annual performance measures that were slow to gather and process, and updated at most nine times a year, the CJS can be updated in

near real time with up-to-date patient feedback. One advantage of the more timely data is the ability to spot rapid changes, such as the extreme examples of Ealing Hospital (R1K04) and Newham General Hospital (R1HNN) illustrated in figure 2. While some patients must experience poor care for the CJS to prompt action, the number may be far fewer than when care is monitored solely by large quarterly data collections.

Not only does the more timely data have the potential to benefit regulators and by extension patients, whose experience make-up the signals they eventually act upon, hospitals benefit too. Healthcare settings are complex and unpredictable and the continuous identification of high-quality care (which is more common than low-quality care) can allow organisations to leverage information on good practice for mitigating, rather than reducing, poor-quality care. Such information enables organisations to proactively improve their resilience, and is consistent with a ‘Safety II’ approach to maintaining a high quality of care.⁴¹

Irrespective of these stated benefits, it would be unwise to rely solely on patient feedback to prioritise inspections. Doing so would disregard large volumes of potentially valuable quality information including mortality and readmission rates, waiting times and staffing levels. Although CQC’s IM tool has not been successful, other means of aggregating quantitative data have shown promise in the past⁴² and it is probable that such data considered in conjunction with patient feedback provide the best means to effectively identify risks to the quality of care and prompt further investigation.¹⁷

The results of this study raise two questions about patients’ collective judgement which require further

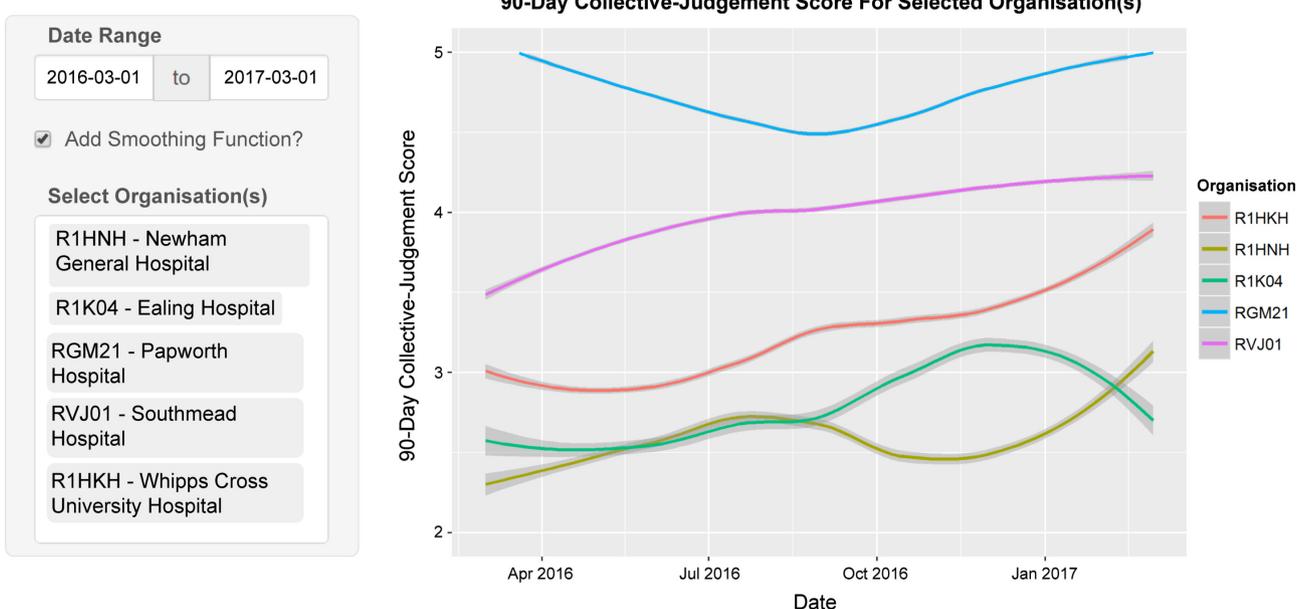


Figure 2 The Loess-smoothed 90-day collective judgement score for six NHS (National Health Service) organisations over a 1-year period taken from the Patient Voice Tracker System developed by the authors.

consideration. First, why is it that, collectively, patients are better able to identify organisations that CQC will rate well, than organisations they will rate poorly? One possibility is that the issues CQC found in hospitals and trusts they rated poorly but had a high CJS were less patient-facing, or impacted patients less directly, than others. An alternative explanation is that CQC ratings, and in particular the negative ratings, are unreliable. Concerns over the reliability of CQC's inspection outcomes are not new,⁴³ but the comprehensive inspections comprise large teams of specialist inspectors, clinicians and 'experts by experience', have received significant support from the sector.^{9 14 44}

The second question is how are patients, the vast majority of whom will have no clinical training, able to predict the outcome of comprehensive quality inspections? There are several possible non-exclusive explanations. A growing volume of research suggests that, in addition to the more service-orientated aspects of care such as dignity and respect, cleanliness and timeliness, patients are able to accurately assess some of the more clinical aspects of their care.⁴⁵⁻⁴⁷ This insight may be enhanced by the 'wisdom of crowds' phenomenon which states that, under the right circumstances, groups can be remarkably insightful, even if the majority of people within a group are not especially well informed.⁴⁸ While individuals seldom have all the necessary facts to make an accurate assessment, and are subject to numerous heuristics and biases, when their assessments are aggregated in the right way, the collective assessment is often highly accurate. Combining multiple sources increases the likelihood that patients' collective judgement will be accurate as it increases both the volume and diversity of feedback. It may also be the case that aspects of care that are more easily assessed reflect accurately on those that are not. The poor management which results in impolite staff and poor timekeeping may also impact clinical effectiveness. Finally, another possibility is that CQC inspectors focus on the satisfaction of patients at the expense of clinical measures that patients may be less able to accurately assess for themselves.

Regardless of the reason for the agreement between the CJS and CQC inspection ratings, it is probable that its adoption will result in some unwelcome behaviour. First, some degree of 'gaming' is inevitable.⁴⁹ While this cannot be eliminated entirely, the effects can be minimised: meta-data is available with all feedback and repeated comments from certain users, or multiple supposed patients at similar times or using similar language can be easily detected. Moreover, the sheer volume of comments means attempts by individuals to manually game the data will have minimal impact. Second, both organisations and patients may change their behaviour knowing the potential consequences of feedback. Some changes in behaviour, such as more courteous staff and better communicated appointments, may be positive; but there may also be

a pressure to provide unnecessary or ineffective treatments to placate newly empowered patients. With the collection of patient feedback well established in the NHS, the effect on patients and organisations of formally collating additional feedback may be limited.

Strengths and weaknesses of the study

Our study has a number of strengths. First, it combines multiple sources of patient feedback to increase the volume and diversity of patient feedback. Second, it accurately classifies tweets excluding those that are not patient feedback related to the quality of care. Third, the inspection ratings used as an independent variable are comprehensive and reflect the quality of care throughout at multiple points over more than 3 years, rather than a snapshot provided by other quality measures such as annual surveys. Fourth, it made use of the more granular hospital-level feedback and ratings, and therefore suffers less from the overaggregation of information concerning large, complex organisations.

Further to the potential challenges of gaming detailed above, our study has four constraints. First, Twitter data were only available from February 2016 onwards and so have only been included for inspections starting after this point; this is due to the time-limited nature of Twitter's API. Second, the class imbalance in the data, with 61% (277) of all organisations in the study being rated as 'Requires improvement' and only 3% (14) being rated as 'Outstanding', makes drawing robust conclusions harder than would be the case were the ratings more evenly distributed. Third, feedback via social media may be influenced by current affairs. Fourth, while combining multiple sources of patient feedback has reduced representative bias, some patient groups have undoubtedly remained under-represented.

Unanswered questions and future research

There is still a lot to be learnt about the use of aggregated patient feedback. It is possible that the predictive power of the CJS could be improved by weighting patient feedback by age or source, standardising for organisation type, shortening or lengthening the time period for the CJS, or adding additional data sources. Moreover, the quality of care is neither objective nor one dimensional.⁵⁰ Future research should aim to identify what feedback is most strongly associated with specific dimensions of quality in order to better target interventions. Identifying how patient feedback can be best combined with quantitative data, such as mortality rates and waiting times, in order to identify risk should be a priority for CQC.

CONCLUSION

Aggregating patient feedback increases the volume and diversity of patient-centred insights into the quality of care. The resulting collective judgement

can successfully identify a high-risk group of organisations for inspection, is available in near real time and is available at a more granular level than the majority of existing data sets. The near real-time, automated collection and aggregation of multiple sources of patient feedback should be used to help prioritise inspections.

Twitter @AlexG_aca

Contributors AG conceived the study, gathered the data and performed the analysis. MPL led on the safety aspects of the study and devised the systematic approach for coding the Twitter data. Both authors developed the argument of the study as it progressed and cowrote the article.

Funding We gratefully acknowledge the financial support provided for this project by the Economic and Social Research Council (Grant Ref: ES/N018869/1) under the Open Research Area Scheme (Project Title: QUAD — Quantification, Administrative Capacity and Democracy). The QUAD project is an international project co-funded by the Agence Nationale de la Recherche (ANR, France), Deutsche Forschungsgemeinschaft (DFG, Germany), Economic and Social Research Council (ESRC, UK), and the Nederlands Organisatie voor Wetenschappelijk Onderzoek (NWO, Netherlands).

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data are publicly available. Details of how to obtain the data are listed in the Technical Appendix.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- 1 CQC. About Us. 2015 <http://www.cqc.org.uk/content/about-us> (accessed 17 Apr 2015).
- 2 CQC. CQC Care Directory. 2017. <http://www.cqc.org.uk/content/how-get-and-re-use-cqc-information-and-data#directory>. (accessed 03 Apr 2017).
- 3 BERR. *Regulator's compliance code: statutory code of practice for regulators*. London, 2007.
- 4 CQC. *Care Quality Commission Annual Report and Accounts 2015/16*. UK: Williams Lea Group, 2016.
- 5 CQC. Shaping the Future: CQC's Strategy for 2016 to 2021 2016. http://www.cqc.org.uk/sites/default/files/20160523_strategy_16-21_strategy_final_web_01.pdf. (accessed 04 Apr 2017).
- 6 CQC. Quality and Risk Profiles: Statistical Guidance, Outcome-Based Risk Estimates in QRPs Produced for NHS Providers. 2013, 2014. http://www.cqc.org.uk/sites/default/files/documents/20130314_nhs_statistical_guidance_march_2013_for_publication.pdf (accessed 12 Jul 2014).
- 7 Walshe K, Phipps D. *Developing a Strategic Framework to Guide the Care Quality Commission's Programme of Evaluation*. London: CQC, 2013.
- 8 Francis R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: The Stationery Office*, 2013.
- 9 CQC. A new start: responses to our consultation on changes to the way CQC regulates, inspects and monitors care services. 2013 http://www.cqc.org.uk/sites/default/files/documents/cqc_newstartresponse_2013_14_tagged_sent_to_web.pdf (accessed 08 Jan 2013).
- 10 CQC. Intelligent Monitoring: NHS Acute Hospitals Indicators and Methodology - Guidance to Support the December 2014 Intelligent Monitoring Update: Care Quality Commission. 2014 http://www.cqc.org.uk/sites/default/files/20141127_intelligent_monitoring_indicators_methodology_v4.pdf
- 11 CQC. *Intelligent Monitoring: NHS Acute Hospitals Indicators and Methodology*. 2015: Care Quality Commission. http://www.cqc.org.uk/sites/default/files/20150526_acute_im_v5_indicators_methodology_guidance.pdf.
- 12 Griffiths A, Beaussier AL, Demeritt D, *et al*. Intelligent Monitoring? Assessing the ability of the Care Quality Commission's statistical surveillance tool to predict quality and prioritise NHS hospital inspections. *BMJ Qual Saf* 2017;26.
- 13 Providers NHS. Written Evidence from NHS Providers to the Public Accounts Committee Care Quality Commission Inquiry. 2015 <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/public-accounts-committee/care-quality-commission/written/23683.pdf> (accessed 13/11/2015).
- 14 NAO. Care Quality Commission: Capacity and Capability to Regulate the Quality and Safety of Health and Adult Social Care. 2015 <https://www.nao.org.uk/wp-content/uploads/2015/07/Capacity-and-capability-to-regulate-the-quality-and-safety-of-health-and-adult-social-care.pdf> (accessed 06 Nov 2015).
- 15 CQC. CQC Insight: Development and Information Sources: NHS Providers. 2016 <https://www.nhsproviders.org/media/2141/cqc-inspection-lisa-annaly-doo-coo-network-19-july-2016.pdf> (accessed 17/04/2017).
- 16 Department of Health. *The Power of Information: Putting All of Us in Control of the Health and Care Information We Need London*: The Stationary Office Ltd, 2012.
- 17 National Advisory Group on the Safety of Patients in England. *A Promise to Learn—a Commitment to Act: Improving the Safety of Patients in England*, 2013.
- 18 Donnelly L. *NHS will monitor facebook for comment about hospitals, says new CQC head*: The Telegraph, 2016. (accessed 13 Feb 2016).
- 19 Cambria E, Hussain A, Havasi C, *et al*. *Towards Crowd Validation of the UK National Health Service*. 10: WebSci, 2010.
- 20 Greaves F, Ramirez-Cano D, Millett C, *et al*. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. *BMJ Qual Saf* 2013;22:251–5.
- 21 Stewart M. Towards a global definition of patient centred care. *BMJ* 2001;322:444–5.
- 22 Weingart SN, Pagovich O, Sands DZ, *et al*. What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med* 2005;20:830–6.
- 23 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 Qual Saf* 2012;21:600–5.
- 24 Greaves F, Lavery AA, Cano DR, Ramirez Cano D, *et al*. Tweets about hospital quality: a mixed methods study. *BMJ Qual Saf* 2014;23:838–46.
- 25 Bardach NS, Asteria-Peñaloza R, Boscardin WJ, *et al*. The relationship between commercial website ratings and

- traditional hospital performance measures in the USA. *BMJ Qual Saf* 2013;22.
- 26 Glover M, Khalilzadeh O, Choy G, *et al.* Hospital Evaluations by Social Media: A Comparative Analysis of Facebook Ratings among Performance Outliers. *J Gen Intern Med* 2015;30:1440–6.
 - 27 Hawkins JB, Brownstein JS, Tuli G, *et al.* Measuring patient-perceived quality of care in US hospitals using Twitter. *BMJ Qual Saf* 2016;25.
 - 28 IPSOS. Ipsos Media CT Tech Tracker Q3 2016. 2016 https://www.ipsos-mori.com/Assets/Docs/Publications/Ipsos_Connect_TechTracker_Q3_2016.pdf (accessed 18 Apr 2017).
 - 29 IPSOS. Social Grade: A Classification Tool. 2009 https://www.ipsos-mori.com/DownloadPublication/1285_MediaCT_thoughtpiece_Social_Grade_July09_V3_WEB.pdf (accessed 18 Apr 2017).
 - 30 NHS Choices. NHS Choices Traffic in 2015. 2016 <http://blogs.nhs.uk/choices-blog/2016/01/06/nhs-choices-traffic-in-2015/> (accessed 18 Apr 2017).
 - 31 CQC. CQC Reveals Comprehensive Picture of the Quality of Hospital Care in England 2017. <http://www.cqc.org.uk/content/cqc-reveals-comprehensive-picture-quality-hospital-care-england> (accessed 16 Mar 2017).
 - 32 CQC. How CQC Regulates: NHS Acute Hospitals Provider Handbook: CQC. 2015 http://www.cqc.org.uk/sites/default/files/20140925_acute_hospital_provider_handbook_main_final.pdf (accessed 01 April 2014).
 - 33 Patient Opinion. How We Moderate. 2017 <https://www.patientopinion.org.uk/info/moderation> (accessed 17 Mar 2017).
 - 34 NHS Choices. Our Terms, Conditions and Policies. 2017 <http://www.nhs.uk/aboutNHSChoices/aboutnhschoices/termsandconditions/Pages/commentspolicy.aspx> (accessed 17 Mar 2017).
 - 35 Ziewitz M. Experience in action: Moderating care in web-based patient feedback. *Soc Sci Med* 2017;175:99–108.
 - 36 Facebook. Facebook Business: How Do I Remove Reviews? 2017 <https://www.facebook.com/business/help/community/question/?id=10207960226947347> (accessed 04 Apr 2016).
 - 37 McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012;22:276–82.
 - 38 Coelho LP, Richert W. *Building Machine Learning Systems with Python*. 2nd Revised Edition. Birmingham: edPackt Publishing, 2015.
 - 39 Chawla NV, Bowyer KW, Hall LO, *et al.* Smote: Synthetic Minority over-Sampling Technique. *Journal of Artificial Intelligence Research* 2002;16:321–57.
 - 40 Van Rijsbergen CJ. A THEORETICAL BASIS FOR THE USE OF CO-OCCURRENCE DATA IN INFORMATION RETRIEVAL. *Journal of Documentation* 1977;33:106–19.
 - 41 Braithwaite J, Wears RL, Hollnagel E. Resilient health care: turning patient safety on its head. *Int J Qual Health Care* 2015;27:418–20.
 - 42 Bardsley M, Spiegelhalter DJ, Blunt I, *et al.* Using routine intelligence to target inspection of healthcare providers in England. *Qual Saf Health Care* 2009;18:189–94.
 - 43 Walshe K, Addicott R, Boyd A, *et al.* Evaluating the Care Quality Commission's Acute Hospital Regulatory Model: Final Report. 2014 <https://www.cqc.org.uk/sites/default/files/CM071406%20Item%206%20Acute%20regulatory%20model%20evaluation%20report.pdf> (accessed 16 Dec 2014).
 - 44 Health Select Committee. *Accountability hearing with the care quality commission London*. London: Health Select Committee, House of Commons, 2013. <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmhealth/761/761.pdf>.
 - 45 Weissman JS, Schneider EC, Weingart SN, *et al.* Comparing Patient-Reported Hospital Adverse Events with Medical Record Review: Do Patients Know Something That Hospitals Do Not? Patient-Reported Adverse Events Versus Medical Record Review. *Annals of internal medicine* 2008;149:100–8.
 - 46 Weingart SN, Pagovich O, Sands DZ, *et al.* What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med* 2005;20:830–6.
 - 47 Ward JK, Armitage G. Can patients report patient safety incidents in a hospital setting? A systematic review. *BMJ Qual Saf* 2012;21:685–99.
 - 48 Surowiecki J. *The Wisdom of Crowds*. London: Little, Brown, 2004.
 - 49 Bevan G, Hood C. What's measured is what matters: targets and gaming in the english public health care system. *Public Adm* 2006;84:517–38.
 - 50 Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;260:1743–8.