

Patient experience feedback: we need to engage with the issues of using Big Data methods to capture the human voice

 blogs.lse.ac.uk/impactofsocialsciences/2017/03/20/patient-experience-feedback-we-need-to-engage-with-the-issues-of-using-big-d

3/20/2017

The NHS regularly asks its patients to complete surveys reporting on the quality of care they have received. These surveys include opportunities for patients to submit feedback in their own words. **Carol Rivas** describes how computational and digital methods can be used to analyse and report patient feedback in an efficient and timely manner. However, it is important to recognise and resolve some of the issues arising from use of these methods; from the need to avoid reducing patients' voices to data points, to the risks of outputs being viewed as incontestable "truths".



This post is part of our [digital methodologies](#) series, arising from [a stream at the ISA Logic and Methodologies conference](#), September 2016, organised by Mark Carrigan, Christian Bokhove, Sarah Lewthwaite and Richard Wiggins.

The NHS is said to have the most comprehensive patient experience data collection in the world. Much of this comes from annual surveys, responses to which suggest that as a nation we are generally [pretty satisfied with the health care we receive](#) at an individual level. This might seem like good news. But it turns out that what people say in answer to the closed questions in patient experience surveys (the type that are answered by multiple choice questions, drop-down menus or measurements) does not always agree with the comments those same patients make when answering associated freetext questions (the type that ask "Have you any other comments? Please write them in this box", for example). The reason, of course, is that these comments are free of the constraints and specificities of a closed question. As such they often provide a more complete picture, and rich detail on care failings, and the way they are written can make them more persuasive vehicles for change than simple yes/no answers.

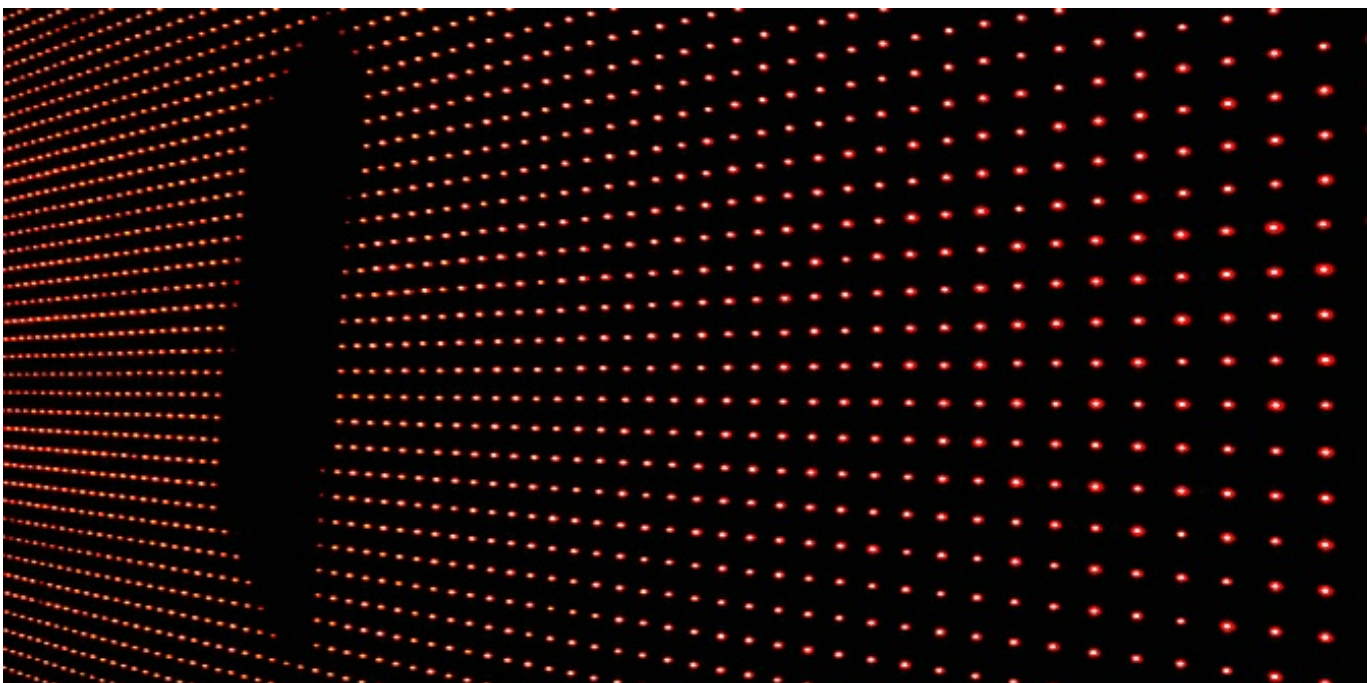


Image credit: [LEDs](#) by Shaun Dunmall. This work is licensed under a [CC BY-SA 2.0](#) license.

The challenge of large-volume qualitative data on healthcare

Unfortunately it has proved a challenge for health service staff to make sense of freetext feedback from patients in any systematic way. So while it is important to ask patients about the care they receive in order to change healthcare practice, we [cannot be certain how effective this approach actually is](#).

As a way of tackling the problem, [we are applying a two-stage approach](#). First we are analysing the patient experience survey freetext comments using [knowledge engineering](#), a type of text mining. This involves constructing [gazetteers](#) and writing “rules” or algorithms so words in the freetext comments can be tagged and grouped according to healthcare topics or themes chosen by patients and healthcare staff. We are then outputting the results as web-based summary visualisations that can be mined down to the original data. Our summaries have been designed to enable frontline staff to make [better, more timely use of survey freetext](#) responses to drive healthcare improvements. Each year new survey data could be put through our system and be reported to frontline staff within hours.

We have begun with the Cancer Patient Experience Survey, since this is widely acknowledged as [the most successful national patient experience survey](#) in enabling service improvement through summaries of responses to its closed questions. Each year CPES is sent out over a three-month period to all patients treated for cancer as in-patients or day cases in NHS trusts in England. The 70,000+ freetext comments it generates annually, which are our focus, are provided to NHS trusts as unstructured raw data.

While [strictly speaking these are not “Big Data”](#), their volume is too great for useful and systematic analysis or reporting without using computational, Big Data-style approaches such as ours. For example, when healthcare staff in London tried to analyse the 2013 London CPES freetext comments on top of their other duties, it took them 18 months to do so and the analysis was not published until [two years after the data had been released](#). Services were likely to have changed from the time the data were collected, reducing the impact and direct usefulness to local service improvement of any analyses. Clearly, then, our approach fulfils a need for staff wishing to use CPES.

Using Big and Small Data methods

It is important to remember that these data represent feedback from people living through often serious health events. Computational approaches risk removing the patient from the data. When we began our study, our patient advisory group was concerned that our results would be mechanistic and reductionist, with survey respondents and the people and healthcare processes they wrote about reduced to mere numbers, summary generalisations and data points developed from our own perspectives. So, when developing our process, we interlaced our computational (Big Data-style) work with Small Data inputs, to keep the patient’s voice present throughout, in what Halford and Savage call a [symphonic approach](#).

What do we mean by “Small Data”? Well, in spring 2016, for example, we invited patients, carers and healthcare professionals to complete an online survey which prompted them to brainstorm terms and phrases they might use when writing about their experiences as a patient or potential patient. From this we discovered not only terms and phrases for our knowledge engineering, but also that patients and professionals had very different perspectives on what good care meant. This divergence of opinion was also apparent when we held discussion workshops in autumn 2016 and is something we have tried to take into account in the presentation of our outputs. Ultimately we want these to drive healthcare improvements that matter to the patients themselves.

Data quality

Several other issues affect our final outputs. In a second survey sent to patients and healthcare professionals, asking them what they thought were the benefits and drawbacks of dashboard data visualisations, data quality was considered the most critical issue; something we discussed in our workshops.

There is a danger our outputs, being based on a national survey, will be seen as absolute and incontestable “truth” by some users of our system, when in fact they can only provide a rough indication of care quality. There are many

reasons for this, not least the fact that survey data themselves are typically flawed; for example, ethnic minority groups are not well represented in CPES and there are other [sampling and reporting biases](#). Biases also arise because we are [channelling and filtering the data](#) in particular ways; for example, we are making conscious choices to select for display on a dashboard only those parts of the data we consider to be significant to healthcare improvement.

As another issue, some trusts may have many comments on a particular aspect of their care, such as waiting times, and other trusts may only have one or two, but this does not necessarily mean that one is better than the other.

There may also be errors in the original data entry; Quality Health, commissioned by the NHS to run CPES, had to [publish a revision](#) shortly after release of the latest (2015) data because some responses had been assigned to the wrong tumour group. There may be other errors that have not yet been spotted.

Issues with data quality will not necessarily be apparent to end users who need to know not just the limits of the data set, but also the limits of which questions can be asked of it and what interpretations are appropriate.

There are two ways we are tackling this. The first is simply to make the limitations clear in an article that accompanies our visualisations. The second is to develop indicators within our data displays that alert the user to the relative strength or weakness of that particular bit of data.

Good design principles limit the number of themes we can show together on a webpage and users will be able to choose which ones to look at. In doing so they may find spurious patterns in the data; for example, if staffing levels and hospital food quality both appear to be rated poorly by patients, the service provider may think they are linked when the correlation may have simply occurred by chance. Such issues may lead us to limit some of the individualisations that can be made.

Opening up the debate

Many of the issues we have encountered have been specified by others as the reason the health sciences lag behind other disciplines in their use of computational Big Data approaches. We hope we have gone some way to bridge the chasm between person-centred care and the depersonalised approach of computing, using our combination of Big and Small Data approaches. Our aim has been to produce actionable insights into healthcare that lead to its improvement in ways that are meaningful to both patients and healthcare staff. But we also hope to open up a debate on how to deal with the challenges of data dissemination in this area, as more and more of these patient experience data become reconstructed into what [Boyd and Crawford](#) call a “garden hose” trickle of digestible bytes designed for professional and public consumption.

This blog post is based on the work of the PRESENT project. PRESENT is funded by the National Institute for Health Research Delivery Research Programme (project number 14/156/15). The views and opinions expressed herein are those of the author and do not necessarily reflect those of the HS&DR, NIHR, NHS or the Department of Health.

Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

About the author

Carol Rivas is a Senior Research Fellow in Health Sciences at the University of Southampton, interested in patient experience, health informatics and digital technologies. She leads on the study described in this post. Her ORCID is [0000-0002-0316-8090](#).

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.

