Patient Perception, Preference and Participation

Making data more meaningful: Patients' views of the format and content of quality indicators comparing health care providers

Zoe Hildon *, Dominique Allwood, Nick Black

London School of Hygiene & Tropical Medicine, Department of Health Services Research & Policy, London, UK

ARTICLE INFO

Article history:
Received 26 July 2011
Received in revised form 9 February 2012
Accepted 11 February 2012

Keywords:
Data display
Format and content
Quality indicators
Patient reported outcome measures
(PROMs)
Elective surgery

ABSTRACT

Objective: Patient reported outcome measures (PROMs) are being used to inform national quality indicators for health care providers in England. Our objective was to explore patients' views of different formats and content of these data displays.

Methods: Six focus groups (N = 45) considered different formats (tables, bar charts, caterpillar and funnel plots) and content (uncertainty displays, volume of outcomes, color, icons, and ordering). A thematic analysis was carried out based on transcripts.

Results: Tables housing icons (star ratings) were the preferred display, having the most popular resonance and facilitating cognitive processing, appealing to most as the best format. Inferred meanings were overwhelmingly applied to unfamiliar formats (funnel plot) and content (uncertainty). Traffic light highlighting and using consistent and recognizable icons (five stars) helped understanding. Familiarity with the bar chart seemed to lend to false readings – error bars were not interpreted.

Conclusion: Aspects of familiarity, meaningfulness (whether and how displays resonated with participants), and cognitive or choice processes informed views of displays.

Practice implications: Tables with star ratings should be a primary display, ordered on statistical significance, including providers in the region plus England average. Caterpillar plots should be used to provide contextual information. More detailed numeric tables should also be accessible.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Over the past two decades the number of national clinical audits in England has grown. One of the latest focuses on common elective surgery procedures [1]. The National PROMs Programme uses patient reported outcome measures (PROMs), before and after surgery, to create quality indicators to compare competing health care providers. These data are intended to inform patient selection on where to have their operation. With the advent of greater available patient centered information there is also a need to ensure that these are easy to understand, positioned to enable the best choices, and that people will like using them.

To date, displays comparing the performance of healthcare providers are largely based on common sense. Nevertheless, the impact on comprehension, choice and preferences of varying the compositional format (e.g. bar charts vs. table) and the content housed within quantitative data displays (e.g. ordering) has been shown to be significant. To inform the scientific basis of the current study we conducted a systematic review of data displays used in simple (binary or categorical) decision making; these results are fully reported elsewhere [2]. However main findings are summarized as follows.

In terms of compositional formats tabular displays appeared easier to interpret than bar charts [3–5]. Although bar charts might be particularly appropriate for helping interpretation of complex multiple outcomes [6]. Consistently, studies showed a preference [7–9] for the bar chart which was viewed as the simpler format, which did not align with the evidence on comprehension. It was also apparent that more complex formats such as the funnel plot and caterpillar charts were not being examined.

In terms of content, icons appeared more user friendly than numbers, but could lead to over-estimation of risk [10–13]; this was also true for using words (e.g. probably) instead of statistics (e.g. 80%) [14–16]. Although words can suffer from inconsistency in interpretation [17]. Giving visual explanatory cues [3,18] or traffic lights to guide interpretation [19] and enhanced comprehension. It was also noted that simplicity should be attempted (‘less was more’). For instance, significantly more people chose a high quality hospital when only quality information was shown [20]. Rank ordering of data was also found to be important to
inform accuracy of choices [18,21]. Uncertainty was not widely understood, nor well represented [9,22].

We aim to build on these existing studies and to develop a set of practical recommendations for PROMs quality indicators informed by end-users; we therefore undertook to qualitatively observe and explain known aspects of quantitative presentational methods and to include format and content variations that have been so far neglected. We explored the following formats: The bar (Fig. 1) and caterpillar chart (Fig. 2), different tables (Fig. 3), and funnel plot (Fig. 4). Contents included: numeric, verbal and types of icons; color; representation of uncertainty; ways of ordering of information; and amount of information.

2. Methods

2.1. Study design

Six focus groups were undertaken across England. A range of displays was developed showing data on up to 100 healthcare...
Table 1
Demographic profile of focus group participants (N=45).

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex</th>
<th>Operation</th>
<th>Age (years)</th>
<th>Socio-economic status (IMD* quintiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>Hip</td>
<td>Knee</td>
</tr>
<tr>
<td>A (N=8)</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>B (N=7)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C (N=6)</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>D (N=8)</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>E (N=9)</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>F (N=7)</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>22</td>
<td>23</td>
<td>15</td>
<td>17</td>
</tr>
</tbody>
</table>

(A) and (B) London, (C) Birmingham, (D) Sheffield, (E) Liverpool, and (F) Bournemouth.

<sup>a</sup> Index of multiple deprivation, 1 least deprived to 5 most deprived, or – missing data.

<sup>b</sup> Varicose veins.

providers. These data were hypothetical, using fictitious provider names. Although derived from real PROMs data they showed less variation than would normally occur in order to keep discussion focused around displays themselves, rather than on the results. Different fictitious names were assigned to different providers on different charts to avoid feedback being influenced by preceding displays; e.g. learning which provider ranked statistically ‘best’ from an earlier displays rather than interpreting rank.

PowerPoint presentations were anchored in a topic guide that prompted discussion around how clear the different displays were perceived to be, and how they were interpreted; we asked where participants might chose to have elective surgery based on the different displays, and what they liked and did not like about them. Meetings lasted about an hour and a half and were held in hotels or community or conference centers. Before fieldwork began ethics approval was obtained (MREC and the LSHTM ethics committee) and implemented.

2.2. Participants

The first focus group was recruited through Arthritis Care and the Royal College of Surgeons Patient Liaison Group who helped identify people that had undergone or were planning knee surgery. Meanwhile, participants were also recruited from the Patient Outcomes in Surgery (POIS) Audit, a fore-runner of the National PROMs Programme. We chose to approach patients who were planning or had undergone elective surgery because they were likely to be more engaged in the study. In practice, we also agreed that a few participants bring their partners or carers, who represented the potential for joint decision making processes and related interpretation of the data displays we were examining.

People were invited to attend a focus group in their region between October and December 2010. From the POIS Audit, 40 patients per region were invited initially, and 40 more invitations were issued until the optimum number of responses (minimum of 13), including an adequate socio-demographic mix, was reached. We excluded those who needed help completing the questionnaires. In total, 376 invitations were issued and 76 agreed to be contacted to discuss taking part (20%). Selection was stratified by the surgical operation they had undergone, age, sex and index of multiple deprivation (IMD). This resulted in 45 patients attending the five group meetings. Eight were recruited through our external contacts, 31 from the POIS audit, and 6 were spouses or carers. Participant details are shown in Table 1.

2.3. Topics covered in meetings

Charts were based on the Oxford Hip Score (OHS) [23]. Participants were explained the scaling and told that ‘good’...
Table 3
Summary of descriptive codes and shared themes for compositional formats.

<table>
<thead>
<tr>
<th>Shared themes</th>
<th>Compositional themes: first order descriptive codes</th>
<th>Table with icons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Un/familiarity</td>
<td>Suited to the more numerate</td>
<td></td>
</tr>
<tr>
<td>Explanation</td>
<td>Known format</td>
<td></td>
</tr>
<tr>
<td>Less is more</td>
<td>Had to learn to read the format</td>
<td></td>
</tr>
<tr>
<td>Overabundance of data</td>
<td>Suggested more numerate</td>
<td></td>
</tr>
<tr>
<td>(group by region)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dis/enabling cognitive</td>
<td>Needed less explanation</td>
<td></td>
</tr>
<tr>
<td>processing</td>
<td>Lacked visual clarity</td>
<td></td>
</tr>
<tr>
<td>Inferred meaning</td>
<td>Needed less explanation</td>
<td></td>
</tr>
<tr>
<td>Popular resonance</td>
<td>Speculation around meaning of 'plots', and number of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>procedures</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.1. Compositional formats

3.1.1. Numerical table

It was sometimes noted that numerical tables lacked visual clarity; as such participants found them ‘busy’ or a ‘bit technical’. This format was also seen as suited to comparison of small numbers. To deal with the over abundance of (national) data, grouping by region was suggested: ‘I’d break them down into areas – who wants to wade through all of that?’

Yet, some participants found the numerical tables needed less explanation than other formats; although they were particularly suited to the more numerate: ‘I’ve spent all my life in analysis, programming […] I find that [graphics] really confusing. The numbers themselves were quite telling.’

3.1.2. Bar chart (with and without error bars)

Even though performance was seemingly harder to identify in the bar charts, these were liked because they were considered visually clear and facilitated appraisal at a glance, ‘You can see at a glance who did the best then, can’t you?’ Bar charts were also a known format, ‘That’s the old style of doing it like that. They still put the cricket scores up like that you see’; this might also account for inaccuracies in reading the graphs, people perhaps felt they did not need guidance on interpretation.

Familiarity with this format had another downside: ‘One reason I don’t care too much for them [bar charts] is because it takes me back to my school days […] Statistics was not my forte!’ Bar charts were also greeted with skepticism because – like the caterpillar and the funnel plot – they lacked popular appeal – ‘I just glaze over at a chart like that because I just hate facts and figures’. They were also seen not to give enough information: ‘We’re missing that information that we had before; it’s qualified information and you’ve done the qualification for us’.

3.1.3. Caterpillar plot

Caterpillar plots were seen as visually clearer and to as giving more information than the bar chart without CIs: ‘I find that one [the caterpillar] easy – so you can see that actually some hospitals, even their lower levels, are above average’.

However, participants noted that as well as lacking popular appeal – ‘It’s ever so pretty, but it wouldn’t mean nothing to the average person’ – they would have ‘to study it; look at it twice’ and learn to read them. Nevertheless, choices of ‘best’ provider were often quick and accurate with this format although participants function after the operation was around 42; and the average for England after surgery was 37. Post-operative OHS outcome (adjusted for the pre-operative OHS score and case-mix) was used throughout the presentation. Participants whose operation was not a hip replacement were asked to imagine that the outcome for their operation. Labels were neutralized to facilitate the personalizing of the indicators.

Several elements were kept constant throughout the presentation. Five categories of uncertainty were created and labeled: much better than average (dark green, based on 99.8% CI); better than average (light green, 95% CI), average (blue or white, not statistically significant); worse than average (yellow, 95% CI); much worse than average (amber, 99.8% CI). Thus, traffic light highlighting (except when testing color) was used to indicate where a provider's outcome had reached statistical certainty – i.e. the confidence intervals did not cross the mean for England. Also, the y axis was consistently labeled ‘worse’–‘better’ to indicate direction of scale. The displays were shown in the order and combinations listed in Table 2.

A few charts were added as the focus groups unfolded, based on participant suggestions (e.g. the chart showing regional data). Given the time available and the iterative nature of chart construction, each display was included in at least three focus groups. The order in which formats were shown was maintained. See Table 2 for summary and ordering of presentation.

2.4. Methods of analysis

The recordings were transcribed verbatim. Thematic analyses ordered by compositional format and content variation were undertaken independently by two authors (ZH, NB); firstly coding the descriptive content of the transcripts relative to the topics discussed and subsequently addressing shared themes and how these might inform data presentation. Inter-rater comparison of findings revealed a high level of agreement. Where differences occurred, a consensus was reached through discussion.

3. Results

Compositional formats will be considered first (Table 3), followed by content (Table 4) options. Descriptive codes are reported initially, in italics; themes are summarized in bold at the end of both sections, after all descriptive codes have been reported.
had benefited from assimilating the meaning of the traffic light scheme from other formats.

3.1.4. Funnel plot

The funnel plot was the focus of much heated discussion. It was an unknown format, seen as ‘something out of Star Wars’! More than any other format, it lacked popular appeal: ‘We never had nothing like this when I went to school – I’d like to see a proper chart’. An initial lack of understanding was married to a range of creative meanings, as participants grappled to make sense of this display.

Speculation around the number of procedures led to much over-interpretation of the data. Some viewed small numbers as better, representing ‘cozier’ hospitals, ‘You don’t want to be pushed through like sausages, do we?’ others that bigger volumes were better, regardless of their PROMs outcome.

There was also speculation around the meaning of the plots. For example, one participant initially saw the ‘dots’ as representing multiple outcomes for one hospital: ‘it’s actually showing that some hospitals can perform in the green and the yellow’; another wondered ‘are those little dots individual patients or operations or what?’

Positive feedback was that the funnel plot gave more information, ‘if you really wanted to go [into] a bit more detail that would provide it’; but this was related to the display being suited to the more numerate – ‘unless you’re a mathematician or a statistician, you just think what does it mean?’

Explanation increased the appeal of the funnel plot – which suited the (self-professed) more numerate; but it was also noted these participants had to learn to read them: ‘I’m getting to like that now […] But you’ve got to get your head round it to start with, haven’t you?’

3.1.5. Tables with icons

Tables with icons were seen as accessible to the average person, having popular appeal, ‘I’d want] that one with the stars […] Well, I’ll tell you why, it reminded me of looking at hotels and they have certain stars for what they provide’; particularly for those less numerate or some older people who made remarks such as ‘if I can understand it, then everybody can.’ It was by far the most liked of the displays, and needed less explanation than other formats, partly because it was viewed as visually clear, the ‘simpler’ format.

However, some people did have reservations; finding that it lacked credibility, e.g., ‘it’s a bit Gocompare; I don’t trust that much either.’ They commented that it lacked transparency – obscuring important differences between providers: ‘OK they’re average but some are more average than others, aren’t they?’ For the more skeptical this format was a starting point for giving basic information which can be followed up with other formats: ‘as first place to look, I think that is absolutely excellent.’

3.1.6. Overview of formats (Table 3)

Familiarity with a format was not necessarily a good thing; either connoting bad memories from the classroom or lending itself to false readings. Understanding of less familiar formats, such as caterpillar and funnel plots, grew quickly after explanation particularly for those who perceived themselves as numerate. Less was often more – or as participants commonly voiced ‘keep it simple’: numerical tables needed grouping to minimize cognitive overload. Caterpillar plots were seen as less cluttered and easier to interpret than bar charts and funnel plots.

For some the additional detail in the funnel plots was a bonus, but overwhelmingly led to inferred meanings.

The visually simpler table with icons was often the most preferred display, having the most popular resonance and being seen to enable cognitive processing. It appealed to everyone as either the only format they wanted, or a good starting point.

3.2. Content

3.2.1. How should uncertainty be displayed?

Representations of uncertainty were mostly new to this audience: ‘Who knows anything about confidence intervals […] if you put information like that into the general domain – you’re wasting your time’. With some exposure and explanation of confidence intervals (CIs) some participants easily grasped how to interpret uncertainty, and it appeared that the numbers facilitated interpretation: ‘For me Hospital A is the best, because it’s got a better confidence interval […] you’ve got far greater chance of actually getting between 37 and 40.’
A shared finding across all formats, except the tables housing icons (which in themselves were symbolic/explanatory), was that statistically accurate choice of hospital often relied on the highlighting, adapted from traffic light colors. Comments such as, ‘Green is the key! […] however you visualize it, we’re all going to look at the green’ were commonly voiced. Nonetheless, participants were aware that by relying on highlights they were deferring to our interpretation of the data: ‘I’m going to go to whatever I think you’re saying is the best, and that’s it!’ However, engaging with the concept of risk explicitly was viewed by some as unhelpfully pessimistic. Instead, uncertainty was seen as subjective and difficult to measure – ‘Everybody’s pain threshold’s completely different, so you are going to get uncertainties with any type of scoring.’

3.2.2. How many providers should be shown?

Many people only wanted data at a local level. Local providers were a common practical choice, ‘I would be thinking of my family, they would want to come.’ For some, choice of more regional providers could be swayed by having family elsewhere to stay with, the seriousness of the operation, fitness for travel and waiting times. In contrast, many also asked for enough providers to judge the range of performances: ‘you’d be better off with a choice of 20, wouldn’t you? And you could pick one that was above the line and was more local to you.’ Some participants wanted comparisons with national outcomes: ‘It may be the best in your area… but it might be right down low in the national charts.’

3.2.3. What colors should be used?

Traffic light colors were described as universally recognized – ‘everybody knows what green, yellow and red mean.’ Because of this it was noted that we could have used the full range of colors, as amber and red are more meaningful than yellow and amber. It was argued that red might cause undue alarm, it was also noted that red would help speed up processing and possibly accuracy of choice.

Alternative color schemes were not welcomed. The shaded blue funnel plot was likened to ‘a paint chart’ or ‘too much all the same’ whilst purple and green was disliked as ‘too modern’ or garish. For some the coloring was immaterial, but would matter for those who are color blind.

Using colors consistently was important, as this enabled understanding across formats: ‘If you keep your colors consistent all the way through the charts – that’s the most important thing because you begin to understand what the line really means then.’

3.2.4. What icons or words should be used?

Stars were described as universally recognized and their interpretation did not require the ability to read: ‘There’s going to be a lot of people that don’t read and write English and they’re still going to need to understand it.’

There was a strong consensus that star charts including grayed out possibilities (of five) was the clearest display: ‘That gives you an idea of what size the scale is, doesn’t it? You know its three stars out of five and not three stars out of ten.’

Thumb icons engendered only negative remarks. Thumbs were not easily recognizable symbols ‘I thought it was a boot. I thought it was a ham [laughter]! They look like hairy rabbits, don’t they?’ etc. They were also considered gimmicky, ‘You’re talking about peoples’ health, whereas that is rather trivializing.’

Words (these were ‘at average’ better; ‘worse’, etc.) were liked because they were perceived as needing no personal interpretation, ‘It just tells you straight up, doesn’t it, that?’ Using a mixture of icons and words was difficult to interpret; ‘It can be ‘OK bad’ or … ‘OK good’ and these would require checking against a key which would impede processing. It was also noted that icons and words provided less discrimination or accuracy than numbers.

3.2.5. In what order should providers be shown?

Views of ordering were influenced by participants’ use and knowledge of available providers. As regards choosing a provider, participants noted ‘if you know where the hospital is, that’s fine. But if you don’t… you could be missing out.’ As a consequence, alphabetical ordering was only useful if local providers were already known.

Rank ordering (in tables) according to absolute outcome rather than statistical significance caused some confusion, as a few people questioned why highlighted ‘green’ providers were not ranked top. These people selected on rank order regardless of statistical significance, suggesting they lacked understanding of how to interpret the data: ‘irrespective of how you look at it […] the people at the top are the better and the people down below are not so good.’ This confusion occurred in tables and bars where statistical ranking was more obvious and uncertainty not immediately relatable to the average for England.

3.2.6. Overview of content (Table 4)

Three themes dominated issues to do with comprehension. The first concerned cognitive processing, which was helped by highlighting and particular colors (red and amber rather than yellow and amber); use of simple, consistent and recognizable icons (stars); numbers to provide greater detail, words to convey our interpretation. The second was about requests for improving the meaningfulness of the data, including: enough providers to contextualize local ranking; ordering on statistical significance; and use of traffic light colors (including red). Third, inferred meanings were applied to the data, particularly with respect to trying to understand uncertainty, which was considered pessimistic and subjective. The use of red and amber was thought by some people to cause undue alarm while icons based on thumbs was seen as trivializing the issue.

In terms of familiarity, or indeed unfamiliarity, the concept and representations of uncertainty was new to most participants. Visual cues were easily recognizable, particularly star icons. As for making personal choices, people were conscious of being influenced by the highlights, though were happy to defer to our interpretation of the data. However, most simply wanted to choose from regional data. Acceptance of more distant providers was contingent on personal and institutional factors.

4. Discussion and conclusion

4.1. Discussion

Our findings are in line with the studies of health care plans or providers [3,9,18–22,24] identified from the systematic review underpinning this study. For instance, we also found that tabular displays outperform bar charts [3]. Of our new displays, the caterpillar plot was particularly user friendly, surpassing the bar chart in clarity and being seen to give more information. The funnel plot was suited mainly to the more numerate, and was sometimes misleading. As in previous studies [18,21] we saw some evidence that information was acquired in sequence or by relevancy; in our analyses this tended to differ by whether people grasped statistical significance/trusted the color codes, or not.

We also found that error bars on bar charts were not often understood [9,22], but in our study these were likely to be easily interpreted after explanation; echoing the usefulness of instructive aid [24]. As previous studies demonstrated [3,18,19] our analyses also suggested that visual cues and consistency made data more interpretable, although there was call for the most symbolic icons and color schemes. As for favoring non-numeric [3] and simpler [20,24] graphics, for our participants less was most definitely also more.
An important conclusion emerging from pooling earlier studies was the strong consistency in findings across population groups and topic areas [2]. Likewise, analyses across our focus groups showed views of format and content displays to be remarkably homogeneous. Within group differences hinged mainly on (self-identified) levels of numeracy or related professional background; as was also the case in the earlier identified studies.

In terms of the study's limitations, the socio-demographic characteristics of the participants reflected that of patients undergoing elective surgery as regards age and sex, thus results may not be generalizable to younger people. Also, people from the most deprived fifth of the population, likely to have lower levels of education, were under-represented. Finally, we asked about many different visual displays in a short amount of time and it was not possible to cover all displays in every meeting, therefore we were unlikely to have reached saturation.

4.2. Conclusion

The findings from our analyses add an explanatory element to what is known about data display methods, substantiating recommendations. In particular, we identified the role of familiarity, meaningfulness or whether and how displays resonated with participants, and cognitive or choice processes in informing views of the displays and decision-making.

Moreover, three practices were apparent, bridging these broader themes. The first was an unself-conscious effort of patients to grapple with the meanings of the data, including assigning creative meanings. Participants labeled unfamiliar images (e.g. ‘exploding tunnel plots’) or pointed out personal inferences, which sometimes bore little relation to the statistical reality. This process of banter, critique and at times exasperation related the unfamiliar with the familiar. The second was the pragmatic approach taken by participants who often spoke of the graphics in terms of their usefulness and need for tailoring, in terms of malleability to different purposes or specificity to a particular question. Lastly, a marked learning effect was observed, which was anchored in a willingness to draw on given explanations and visual cues, and to help each other to interpret the displays.

4.3. Practice implications

The data we gathered were tailored to informing preliminary recommendations for displaying PROMs to patients. Comprehension, choice and preferences relative to these initial recommendations need to be tested quantitatively.

Our recommendation is that the performance information on elective surgery for patients be primarily displayed by tables housing stars (showing a total possible of 5). This should be rank ordered by statistical significance and grouped by ‘region’. Traffic light colors should be used accompanied by explanations in plain English. To show trend data and provide further contextual information, caterpillar plots should be used. Numeric tables, providing further details, should also be available for those who want them.

Acknowledgments

We gratefully acknowledge Arthritis Care, the Royal College of Surgeon Patient Liaison Group, and all participants who generously gave their time to the project, as well as Elena Kombou for her help with the Focus Groups. We thank our funders, the Department of Health, and lastly Maria Boyce and Oliver Groene for comments.

References


