Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review

ZOE HILDON, DOMINIQUE ALLWOOD AND NICK BLACK

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

Address reprint requests to: Zoe Hildon, Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, 15–17 Tavistock Place, London WC1H 9SH, UK. Tel: +44-20 7927 2151; Mob: +7722360021; Fax: +44-20 7927 2701.
E-mail: zoe.hildon@lshtm.ac.uk

Accepted for publication 3 November 2011

Abstract

Purpose. Displays comparing the performance of healthcare providers are largely based on commonsense. To review the literature on the impact of compositional format and content of quantitative data displays on people’s comprehension, choice and preference.

Data sources. Ovid databases, expert recommendations and snowballing techniques.

Study selection. Evaluations of the impact of different formats (bar charts, tables and pictographs) and content (ordering, explanatory visual cues, etc.) of quantitative data displays meeting defined quality criteria.

Data extraction. Type of decision; decision-making domains; audiences; formats; content; methodology; findings.

Results of data synthesis. Most of the 30 studies used quantitative (n = 26) methods with patients or public groups (n = 28) rather than with professionals (n = 2). Bar charts were the most frequent format, followed by pictographs and tables. As regards format, tables and pictographs appeared better understood than bar charts despite the latter being preferred. Although accessible to less numerate and older populations, pictographs tended to lead to more risk avoidance. Tables appeared accessible to all. Aspects of content enhancing the impact of data displays included giving visual explanatory cues and contextual information while still attempting simplicity (‘less is more’); ordering data; consistency. Icons rather than numbers were more user-friendly but could lead to over-estimation of risk. Uncertainty was not widely understood, nor well represented.

Conclusions. Though heterogeneous and limited in scope, there is sufficient research evidence to inform the presentation of quantitative data that compares the performance of healthcare providers. The impact of new formats, such as funnel plots, needs to be evaluated.

Keywords: quality improvement, patient–provider communication/information, decision analysis

Purpose

Data displays of quality and risk in health care are being adopted largely on the basis of commonsense rather than on research on presentational methods. Yet, information presentation has been examined within several fields including psychology, market research and information systems/management. However, studies have mostly focused on framing effects, such as relative vs. absolute risk [1–3]. These have shown that people will be persuaded more readily by relative rather than absolute risk information [1, 3, 4] and by positively framed outcomes (e.g. survival rather than mortality data) or loss framing (e.g. years lost rather than gained) [1, 2, 4].

Another focus of presentational methods has been examining verbal (e.g. ‘probably’) compared with numeric (80%) terms for risk; it is widely acknowledged that words may better capture intuitive perceptions of risk, but can suffer from inconsistency in interpretation [5]. They may also lead to overestimations of risk [6, 7] or more risk-avoidant choices [7, 8].

Communication tools have also been examined in the context of health interventions, which have found that certain types of information seem to improve decision-making, e.g. individualized health risk scores [9]; as can data summation, e.g. verbal risk summaries [10] and the use of charts in the consultation process [11]. Studies on
information systems [12–14] comparing tables and graphs have proved inconclusive and seem to depend on the complexity of the task (graphs favoured in complex tasks).

An area that has received less attention and is the focus of this review is making data more meaningful and reducing the mental computational load so that automatic visual perception helps decision-making [14, 15]. This requires a distinction to be made between compositional formats and content. Format refers to the way data are housed and includes options such as bar charts, tables (numeric and non-numeric) and pictographs. Content includes aspects such as whether to rank order data, or whether to use explanatory visual cues (such as + or − symbols) to enhance statistics [15, 16].

When considering the impact of data displays, three decision-making domains have been identified in the literature [14]: people’s ‘comprehension’ (or interpretation) of the data; the way displays affect’s their ‘choice’ (either hypothetically or their behaviour in practice) and people’s ‘preference’ (or liking) for one display over another. Our aim was to review the research evidence on the impact of compositional format and content of quantitative displays across these three domains.

Methods

Data sources

Eight databases hosted by the Ovid interface were searched: Cochrane, EMBASE, Medline, Global Health, Health Management Information Consortium, ERIC, PsychInfo and PsychExtra. Recommendations for grey literature or for literature currently in press were requested from experts. Snowballing [17, 18] was used to ensure that literature which might have been missed by the electronic searches was captured; this process included following-up reference lists of review articles that met our inclusion criteria. We checked to ensure that in-text citations had reached saturation.

The search string combined general terms for metrics (MESH or thesaurus terms) or specific terms for metrics of interest (such as performance indicators, patient-reported outcomes) with decision-making terms; these were in turn joined to (MESH or thesauris and text) terms for data display methods. The search string was: (metrics OR decision-making terms) AND (data display terms).

Study selection

Screening was carried out on titles, abstracts and full-texts by Z.H. The key inclusion criterion was that studies examined quantitative data displays and the impact of different compositional format and/or content.

The following exclusion criteria were applied at all stages of screening: selection and testing of specific metrics, including testing framing effects; examining variations in verbal vs. numerical expressions; studies of decision-support software, for which an existing review [12] was used; exploring complex decision tasks, i.e. interpretation of statistical predictions, rather than (implicit or explicit) simple binary or categorical decisions, e.g. treatment.

Data extraction

The quality of studies was appraised using a tool based on two reviews of methods for rating quality across quantitative, qualitative and mixed methods [19, 20]. The Quality Appraisal across Study Methods (QASM) had five domains: reporting; appropriateness of study design; validity of data and rigour of analyses; reliability of findings; use of method-specific quality enhancing procedures (Supplementary material, Appendix S1). The domains were tailored for the four main methodologies: qualitative; quantitative descriptive; quantitative comparative and mixed methods. Each study was scored on a 10-point scale with results categorized as weak* (0–4), moderate** (5–7) and strong*** (8–10).

For each study, the following information was extracted by team members (Z.H., D.A.) using a template: the decision that the visual displays were intended to inform; the population groups they were tested on; compositional formats and content; study method; impacts considered (comprehension, choice, preference) and findings (results and conclusions). Discrepancies around interpretation of study findings were discussed and resolved.

The results of studies were synthesized according to a framework that considered separately the impact of compositional format and of content on comprehension, choice and preference. For quantitative studies, findings were considered if they reached statistical significance at the level of $P < 0.05$. When the level of significance was not reported, pertinent findings were extracted. For qualitative and mixed methods analyses, reported themes and study findings were also extracted.

Results of data synthesis

Search and screening results

Searching identified 1462 texts (after manual removal of duplicates) for screening (Fig. 1). Screening yielded 486 inclusions at title screen, 82 at abstract screen and 33 at full-text screen.

The majority of the studies ($n = 28$) used quantitative methods (25 comparative and 3 descriptive) and the rest used either qualitative ($n = 2$) or mixed methods ($n = 5$). Most studies were of moderate quality (58%), a third (33%) were strong, and three (9%) were weak (which were excluded from the synthesis). Of the studies rated as strong, seven used quantitative comparative designs, two used mixed methods and two were qualitative.

Description of included studies

Topics. The 30 studies considered a wide range of decisions. All were hypothetical but used realistic data. Only two studies focused on decision-making by professionals:
one asked clinicians whether to stop a trial early or not [21], and one required commissioners to identify poorly performing hospitals [22]. Two studies considered business strategy—selection of a restaurant site [23, 24]. A few asked the public or patients about selection of a healthcare plan, hospital or nursing home (n = 7); or other consumer decisions such as which tyres, toothpaste or cereal to chose (n = 6). But mostly they examined individual treatment and risk-related decisions, e.g. whether to take drug A or B, or consent to treatment, etc. (n = 13).

Audiences. Most studies tested displays on the public (n = 15). The rest were on convenience samples of students (n = 9), patients (n = 4), clinicians (n = 1) or commissioners (n = 1). Some studies considered the socio-demographic characteristics of their audiences: age [25–31], sex [32–35] and levels of numeracy [30, 31, 36].

Compositional formats. There were three principal formats considered: bar charts (n = 19), pictographs (n = 13) and (numerical and non-numerical) tables (n = 9). Consistent with the literature, a matrix-housing icons such as star ratings instead of or alongside a numerical statistic is categorized as a table while pictographs denote a visual representation of a proportion, e.g. 25 smiley faces to denote 25%. These formats incorporated either single or multiple outcomes. In graphs, multiple outcomes can be grouped (along the x axis of a bar chart) or panelled (two charts displayed side by side; Figs 2 and 3). Tables list either multiple indicators (referred to as scorecards) or only one (Fig. 4).

A few studies examined other formats: line graphs (n = 5), pie charts (n = 5) and scatter plots (n = 1) but they were rarely compared with other formats. In addition, comparisons were also made with narrative reports using statistics (n = 9).

Content. Five aspects of content were identified: ordering (n = 6); visual explanatory cues (n = 3); amount of content (n = 6); instructive aids (n = 4) and representation of uncertainty (n = 4).

Impact domains. Comprehension (n = 19) had been studied in terms of accuracy of understanding, numeracy, response...
Preference ($n = 14$) had been tested in terms of usability and formatting. Choice ($n = 19$) was tested in terms of confidence in choice, influences on choice, choice processes and making the (statistically) accurate choice. This last way of examining choice, clearly overlaps with, and is a mechanism for, testing understanding.

### Impact of compositional format

Literature examining compositional format is summarized in Table 1. Tables and pictographs seemed generally better understood than bar charts. The study on clinician’s decision-making found pictographs were the best understood; 82% made correct decisions with pictographs compared with 68% when using tables and only 43% for bar charts [21].

Nevertheless tables were (62%) preferred over the bar chart (23%) with no one liking the pictograph and some clinicians actually contemptuous of that format [21]. Despite preferring bar charts, Commissioners using this format were more likely to make incorrect choices when selecting supposedly under-performing hospitals than with scatter plots [22].

A study of the public’s use of quality information for selecting a nursing home also found that bar charts were less likely to be understood; 86% of participants correctly interpreted non-numeric tables (housing icons) compared with 47% for bar charts; 46% preferred tables and 22% preferred graphs [37].

In business strategizing, it was found that bar charts and line graphs were harder to interpret for people with poorer perception ability compared with tables or narrative reporting [23]. Only one study [26] found that vertical bar charts narrowly outperformed pictographs (same error rate of 1.1% but slower to process) and were better understood than pie charts and horizontal bar charts, but this study did not test tables.

Comprehension of pictographs was found to be generally good [27, 28, 31, 38]. For instance, in a task for estimating chance of survival using this format 86% of responses were correct [38]. In older people correct responses ranged from 70 to 98% using pictographs, higher than numeric statements (38–78%) [27]. Apart from simplifying data for older or less numerate populations, preference for pictographs was also found to be shaped by numeracy [33, 34]. However, one study found both pictographs and tables were liked—rated as effective, trustworthy and scientific irrespective of level of numeracy [31].

Pictographs have also been found to increase people’s perception of risk. Younger, less educated people perceived their life-time risk of breast cancer to be higher when conveyed with pictographs (human figures) than bar charts [34]; only 28% perceived their risk to be low with bar charts compared with 49% with pictographs [35].

Similarly, several studies found that displaying risk reduction using a pictograph rather than by numbers significantly increased risk-avoidant choices [39, 40, 41]. Although it has been suggested this effect might be attributable to framing [39], a later study found the effect to be independent of framing [42].

Even though bar charts seemed the least understood, several studies have found people tend to prefer them to other formats; one study explicitly identified this paradox [22]. Bar charts seemed preferred particularly for health risk information. One such study found 60% of people liked the bar chart compared with 20% the pictographs [33]. In another, bar charts were favoured over pictographs for multiple-risk estimates (22 vs. 60%), although the opposite was true when showing single estimates (46 vs. 28%) [35]. A study of diabetes treatment options reported even greater support for bar charts (97%) [43].

### Impact of content

**Ordering.** Literature examining content derivatives is summarized in Table 2. In presenting consumer product choices, comparing by rank order improves comprehension and choice. Ranking health plans in tables improved comprehension, particularly in older groups, made options easier to identify and significantly reduced poor decisions from 46 to 30% [15]. Twice as many non-quality maximizing choices...
occurred in an unordered matrix as with ranked information [16].

As regards acquisition of information, one study noted that people acquire information in line with the way it is arranged and adopt an approach to processing it that minimizes cognitive effort and errors, aiding decision quality [24].

In tables, attention seems affected by the importance assigned to different factors, whereas in bar charts it is by the height of the bar. However, one study concluded that consumers (of breakfast cereals) appeared to process information in the fashion which is visually easiest from the table they were given [32]. Another noted that choice (of

Table 1 Impact of compositional format on comprehension, preference and choice

<table>
<thead>
<tr>
<th>Comprehension</th>
<th>Choice</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar chart</td>
<td>Bar charts not so easy to understand</td>
<td>Bar charts could lead to inaccurate choices</td>
</tr>
<tr>
<td>Brown [23]**</td>
<td>Elting et al. [21]***</td>
<td>Marshall et al. [22]**</td>
</tr>
<tr>
<td>Elting et al. [21]***</td>
<td>Edwards et al. [43]***</td>
<td></td>
</tr>
<tr>
<td>Gerteis et al. [37]***</td>
<td>Marshall et al. [22]**</td>
<td></td>
</tr>
<tr>
<td>Marshall et al. [22]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical bar charts led to best understanding compared with horizontal or pie charts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feldman-Stewart et al. [26]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictograph</td>
<td>Pictographs easy to understand</td>
<td>Pictographs led to more risk-avoidant choices compared with numbers</td>
</tr>
<tr>
<td>Elting et al. [21]***</td>
<td>Edwards et al. [43]***</td>
<td></td>
</tr>
<tr>
<td>Fuller et al. [27]<strong>, [28]</strong></td>
<td>Marshall et al. [22]**</td>
<td></td>
</tr>
<tr>
<td>Hawley et al. [31]***†</td>
<td>Fortin et al. [33]***</td>
<td></td>
</tr>
<tr>
<td>Price et al. [38]**</td>
<td>Marshall et al. [22]**</td>
<td></td>
</tr>
<tr>
<td>Pictographs could lead to overestimations of harm compared with bar charts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schapira et al. [34]***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schapira et al. [35]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference for pictographs was shaped by numeracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schapira et al. [35]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schapira et al. [35]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table (numeric–non-numeric)</td>
<td>Tables housing icons, words or numbers were sometimes easier to understand than bar charts</td>
<td>Tables led to more accurate choices than graphs for people with poorer perception ability</td>
</tr>
<tr>
<td>Brown [23]**</td>
<td>Elting et al. [21]***</td>
<td></td>
</tr>
<tr>
<td>Gerteis et al. [37]***</td>
<td>Marshall et al. [22]**</td>
<td></td>
</tr>
<tr>
<td>Elting et al. [21]***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quality rating: Weak* (0–4), Moderate** (5–7) and Strong*** (8–10).
<table>
<thead>
<tr>
<th>Comprehension</th>
<th>Choice</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordering</td>
<td>Ordering influences choice, helping people to make ‘right’ choices, or can be used to ‘nudge’ a selection</td>
<td>Not examined</td>
</tr>
<tr>
<td>Information is acquired in sequence, either by rank or by relevancy of information, which seems to make ordered information easier/faster to understand</td>
<td>Hibbard et al. [15]<strong>, Hibbard et al. [16]</strong></td>
<td></td>
</tr>
<tr>
<td>Betman and Kakkar [32]**</td>
<td>Hibbard et al. [24]**</td>
<td></td>
</tr>
<tr>
<td>Hibbard et al. [15]**</td>
<td>Fasolo et al. [30]**</td>
<td></td>
</tr>
<tr>
<td>Hibbard et al. [16]**</td>
<td>Fasolo et al. [30]**</td>
<td></td>
</tr>
<tr>
<td>Jarvenpaa et al. [24]**</td>
<td>Fasolo et al. [30]**</td>
<td></td>
</tr>
<tr>
<td>Although recall is brand led even if information is presented by attribute</td>
<td>Mazur and Merz [45]**</td>
<td></td>
</tr>
<tr>
<td>Biehal et al. [44]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual explanatory cues</td>
<td>Such visual cues also helped towards making the ‘right’ choice</td>
<td>People liked being helped with the process of bringing information together, and liked consistency</td>
</tr>
<tr>
<td>Using visual cues (icons, words) helped to clarify the meaning of data</td>
<td>Hibiard et al. [16]**</td>
<td>Fasolo et al. [30]**</td>
</tr>
<tr>
<td>Hibbard et al. [16]**</td>
<td>Gerteis et al. [37]**</td>
<td>Gerteis et al. [37]**</td>
</tr>
<tr>
<td>Gerteis et al. [37]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount of content</td>
<td>Simpler graphics were more understandable</td>
<td>Preference was for simpler graphics; but also showing information in a variety of ways, to match knowledge/use</td>
</tr>
<tr>
<td>Simpler graphics were more understandable</td>
<td>Simpler graphics encouraged the accurate choice</td>
<td></td>
</tr>
<tr>
<td>Edwards et al. [43]**</td>
<td>Peters et al. [36]**</td>
<td>Edwards et al. [43]**</td>
</tr>
<tr>
<td>Feldman-Stewart et al. [25]**</td>
<td>Uhrig et al. [29]**</td>
<td>Fasolo et al. [30]**</td>
</tr>
<tr>
<td>Peters et al. [36]**</td>
<td></td>
<td>Schapira et al. [34]**</td>
</tr>
<tr>
<td>Uhrig et al. [29]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructive aids</td>
<td>Instructive aids such as practice exercises or worksheets helped comprehension</td>
<td>Not examined</td>
</tr>
<tr>
<td>Instructive aids such as practice exercises or worksheets helped comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armstrong et al. [47]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rakow et al. [46]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uhrig et al. [29]**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Representation of uncertainty

Confidence intervals did not increase understanding of uncertainty

Marshall et al. [22]**
Mazor et al. [48]**

Missing data was not understood and negatively interpreted

Fasolo et al. [30]**
Gerteis et al. [37]**

Using confidence intervals did not influence making the accurate choices

Marshall et al. [22]**

Not examined

Quality rating: Weak* (0–4), Moderate** (5–7) and Strong*** (8–10).
issue, missing data, rather than being viewed as an ‘unknown’, tended to be negatively interpreted [30, 37].

Conclusion

Main findings

As regards compositional format, tables and pictographs were judged to be more accurate decision aids than bar charts. It was also noted that over-estimation and avoidance of risk can occur with the use of icons. This data display method may be more appropriate for older and less numerate populations. Tables appeared accessible to all though for complex data, bar charts may be more appropriate.

Studies of decision-support software have indicated that numerical tables are more effective than graphs unless the task demand is high [49] but this debate has not been resolved [12, 13, 50]. When the decision is a simple binary or categorical choice, formats other than bar charts may be the most appropriate. One reason for the bar chart’s underperformance may be due to a lack of understanding and poor representation of uncertainty (confidence intervals) [22].

With bar charts, choice may be driven by rank ordering [15, 16, 24, 32], rather than whether options are within or outside of normal variation. No studies have tested the interaction between representing uncertainty and ordering, and only one study has considered ordering effects in bar charts [24]. Because of this potential caveat, other graphical formats such as funnel or scatter plots with clearly identified control limits are likely to be more appropriate for comprehension and decision accuracy in simple decision tasks.

As regards content several ways of improving displays emerged. Including consistency in symbols, colours and metrics; careful selection of visual cues to help clarify the meaning of data; yet avoidance of over-elaboration; clear instructions or even practice tasks, on how to interpret displays (particularly for less familiar formats); ordering by rank and relevance of information; clear accounts of uncertainty, particularly with bar charts—especially if data are contingent on significant associations.

Strengths and limitations

While there were a substantial number of studies of moderate or strong methodological quality, the heterogeneity of included studies precluded quantitative synthesis. Instead we sought to identify important themes in a narrative review. We were limited by the tendency to test popular formats, which ignored other displays (e.g. funnel and caterpillar plots) useful for comparing performance of providers. All the studies tested hypothetical decisions, which are known to differ from actual behaviour change.

Implications

Despite the heterogeneity of the studies, we were able to synthesize research on a wide range of topics and using various methodologies. Common findings emerged with considerable consistency. This suggests that the nature of data processing is less dependent on the type of information displayed and more dependent on ‘how’ it is displayed.

In terms of the framework that was applied (comprehension, choice, preference), there is some overlap between the first two constructs. All three are multi-faceted, hinging on specific contexts, ability and frame of reference. Comprehension and choice should be prioritized over preference. Whilst this may not favour the most popular displays at the outset, people are willing to alter their preference if given a good reason [51].

Future research to improve our understanding of how best to convey data that compares performance should include less used formats, such as funnel or scatter plots; and should examine visual displays both with public and professional audiences, something that none of the studies that met our selection criteria attempted.

Supplementary material

Supplementary material is available at INTQHC Journal online.

Acknowledgements

We thank: the Department of Health for funding; Judith Green and Andrew Hutchings for help in developing and refining QASM; library staff at the London School of Hygiene and Tropical Medicine for their support with sourcing literature; Jenny Neuberger and Jan van der Meulen for comments; Elena Kombou for database maintenance and contributions to extractions and Rebecca Wright and Tim Rakow for their comments and recommendations.

Funding

This study was funded by the Department of Health for England.

References


