DISCERN: an instrument for judging the quality of written consumer health information on treatment choices

Deborah Charnock, Sasha Shepperd, Gill Needham, Robert Gann

Abstract

Objective—To develop a short instrument, called DISCERN, which will enable patients and information providers to judge the quality of written information about treatment choices. DISCERN will also facilitate the production of new, high quality, evidence-based consumer health information.

Design—An expert panel, representing a range of expertise in consumer health information, generated criteria from a random sample of information for three medical conditions with varying degrees of evidence: myocardial infarction, endometriosis, and chronic fatigue syndrome. A draft instrument, based on this analysis, was tested by the panel on a random sample of new material for the same three conditions. The panel re-drafted the instrument to take account of the results of the test. The DISCERN instrument was finally tested by a national sample of 15 information providers and 13 self help group members on a random sample of leaflets from 19 major national self help organisations. Participants also completed an 8 item questionnaire concerning the face and content validity of the instrument.

Results—Chance corrected agreement (weighted κ) for the overall quality rating was κ=0.53 (95% CI κ=0.48 to κ=0.59) among the expert panel, κ=0.40 (95% CI κ=0.36 to κ=0.43) among information providers, and κ=0.23 (95% CI κ=0.19 to κ=0.27) among self help group members. Higher agreement levels were associated with experience of using the instrument and with professional knowledge of consumer health information. Levels of agreement varied across individual items on the instrument, reflecting the need for subjectivity in rating certain criteria. The trends in levels of agreement were similar among all groups. The final instrument consisted of 15 questions plus an overall quality rating. Responses to the questionnaire after the final testing revealed the instrument to have good face and content validity and to be generally applicable.

Conclusions—DISCERN is a reliable and valid instrument for judging the quality of written consumer health information. While some subjectivity is required for rating certain criteria, the findings demonstrate that the instrument can be applied by experienced users and providers of health information to discriminate between publications of high and low quality. The instrument will also be of benefit to patients, though its use will be improved by training.
news reports using The Index of Scientific Quality (ISQ) developed by Oxman and colleagues.\textsuperscript{15} There is, however, no systematic method for judging the quality of literature written specifically for patients. We have developed a short instrument, called DISCERN, which features key questions for the critical appraisal of written consumer health information about treatment choices. The DISCERN questionnaire will: (1) enable information providers and patients to judge the quality of written information about treatment choices; (2) facilitate the production of high quality evidence-based consumer health information by setting standards, and by providing a reference point for authors.

**Methods**

The DISCERN instrument was developed in two stages: (1) an expert panel developed and tested the instrument; (2) a national sample of health information providers and self help group members tested the instrument on a range of consumer health information on treatment choices.

**Development of the DISCERN instrument by the expert panel**

**Participants**

We recruited an expert panel representing a range of expertise\textsuperscript{16} in areas relevant to the production and use of consumer health information (see table 1).

**Materials**

We collected consumer health information for three conditions where the evidence-base for treatment choices differs. These conditions were myocardial infarction, endometriosis, and chronic fatigue syndrome. The information collected consisted of books available in public libraries and bookstores, leaflets produced by professional organisations and national self help groups identified through Helpbox\textsuperscript{17} and the Voluntary Agencies Directory,\textsuperscript{18} and articles published in magazines or newspapers listed in the Popular Medical Index\textsuperscript{19} for the preceding 12 months. Publications were included if they were written in English, were written for patients or health consumers and referred to a treatment or treatments (including "no cure or treatment available").

**First analysis by the expert panel**

A random sample of 3 books, 16 book chapters, 48 leaflets, and 16 periodical articles was generated for the expert panel’s initial analysis. This was considered the maximum the panel would be able to read in the time available. We asked the panel to analyse the sample of information independently using their own experience and expertise. They were asked to list one overall set of criteria or questions that best represented the ways in which they made judgements about quality, and which could be applied consistently to rate new material. The panel had four weeks to complete the exercise. The three clinicians and three self help group members constituted a “specialist” group and only analysed information concerning their own area of interest. DC and SS independently sorted the panel’s criteria under common themes.\textsuperscript{16 20}

**First meeting of the expert panel**

The panel met to discuss the results of the initial analysis and to draft the instrument. The meeting was chaired by a member of the project team (RG) experienced in facilitation of consumer health groups and in consensus techniques.

**Second meeting of the expert panel**

The panel independently tested the draft instrument with a new random sample of information about treatment for the same three conditions. The test sample consisted of 21 publications (one book, two book chapters, two leaflets, and two periodical articles for each condition). Clinicians and self help group members again only rated material in their area of interest. Panel members had four weeks to complete and return their ratings. We analysed the data using a measure of inter-rater agreement (see statistical analysis).

**Evaluation of the DISCERN instrument by a national sample of health information providers and self help group members**

**Participants**

We recruited 15 information providers via a national network of contacts in public sector health information services through the Help for Health Trust at Winchester. We defined eligible participants as "professionals responsible either for providing health advice directly to the public, or for dealing regularly with consumer health information (for example, health librarians or resource centre managers)". Fifteen self help group members were recruited through group leaders of local branches of the organisations represented on the expert panel. Two of these participants were unable to complete the rating exercise by the deadline and were subsequently dropped from the pilot.

**Materials**

We obtained a sample of leaflets from each of 21 major producers of consumer health information (as identified by RG). Two organisations were subsequently excluded because...
none of the leaflets they supplied referred to treatment choices. We selected one leaflet at random from the sample for each of the remaining 19 organisations.

**National pilot of the instrument**

We sent each participant the sample of 19 leaflets plus copies of the DISCERN instrument and instructions for rating the information. The order of the leaflets was randomised. Participants were asked to use DISCERN to rate independently the leaflets in the order in which they were presented and not to consult or use additional sources of information when making their ratings. Participants were also asked to complete an 8 item questionnaire assessing the face and content validity and general applicability of the instrument, and to provide demographic details. Participants were assured that all information would be treated in confidence. They had five weeks to complete and return their ratings.

**STATISTICAL ANALYSIS**

We tested the reliability of the instrument at each phase by calculating agreement between raters for each DISCERN item using \( k \) with quadratic weights, a chance corrected measure of agreement. \(^22\) Weighted \( k \) is appropriate for the analysis of data in ordered categories, such as the 5 point Likert scale used to rate each DISCERN item, because it does not treat all disagreements equally. Different weights are given to disagreements between raters according to the magnitude of the discrepancy. Ratings representing a difference of only one category are considered less serious than those where the discrepancy is greater. \(^22\) \(^23\) In the case of multiple raters, weighted \( k \) is calculated by generating a \( k \) score for each possible pair of raters for each item being rated. An overall \( k \) score is then generated by calculating the average of these individual \( k \) with an appropriate overall standard error. \(^21\)

<table>
<thead>
<tr>
<th>Question</th>
<th>Expert panel</th>
<th>Information group</th>
<th>Self help group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \kappa )</td>
<td>95% CI</td>
<td>( \kappa )</td>
</tr>
<tr>
<td>1 Explicit aims</td>
<td>0.31</td>
<td>0.26, 0.36</td>
<td>0.38</td>
</tr>
<tr>
<td>2* Aims achieved</td>
<td>0.34</td>
<td>0.29, 0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>3 Relevance to patients</td>
<td>0.59</td>
<td>0.54, 0.64</td>
<td>0.21</td>
</tr>
<tr>
<td>4 Sources of information</td>
<td>0.36</td>
<td>0.30, 0.41</td>
<td>0.50</td>
</tr>
<tr>
<td>5 Currency (date) of information</td>
<td>0.39</td>
<td>0.34, 0.43</td>
<td>0.59</td>
</tr>
<tr>
<td>6 Bias and balance</td>
<td>0.46</td>
<td>0.41, 0.51</td>
<td>0.18</td>
</tr>
<tr>
<td>7 Additional sources of information</td>
<td>0.70</td>
<td>0.65, 0.76</td>
<td>0.63</td>
</tr>
<tr>
<td>8 Reference to areas of uncertainty</td>
<td>0.41</td>
<td>0.35, 0.46</td>
<td>0.24</td>
</tr>
<tr>
<td>9 How treatment works</td>
<td>0.48</td>
<td>0.43, 0.54</td>
<td>0.28</td>
</tr>
<tr>
<td>10 Benefits of treatment</td>
<td>0.47</td>
<td>0.42, 0.52</td>
<td>0.31</td>
</tr>
<tr>
<td>11 Risks of treatment</td>
<td>0.45</td>
<td>0.40, 0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>12 No treatment options</td>
<td>0.39</td>
<td>0.34, 0.44</td>
<td>0.31</td>
</tr>
<tr>
<td>and</td>
<td></td>
<td>0.23</td>
<td>0.18, 0.28</td>
</tr>
<tr>
<td>13 Quality of life</td>
<td>0.31</td>
<td>0.26, 0.36</td>
<td>0.32</td>
</tr>
<tr>
<td>14 Other treatment options</td>
<td>0.40</td>
<td>0.35, 0.45</td>
<td>0.36</td>
</tr>
<tr>
<td>15 Shared decision making</td>
<td>0.40</td>
<td>0.35, 0.45</td>
<td>0.39</td>
</tr>
<tr>
<td>16 Overall quality</td>
<td>0.53</td>
<td>0.48, 0.59</td>
<td>0.40</td>
</tr>
</tbody>
</table>

*If the answer to Question 1 was No, raters were instructed not to answer Question 2. The analysis for Question 2 was confined to 8 publications where complete data were available for both national pilot groups. \(*\)No value could be calculated for the self help group data, because the dataset consisted almost entirely of mid-range ratings (that is, 3 or 4). While there appeared to be very good agreement, there was insufficient variability across publications for the calculation of \( k \). The value of 0.13 in the Information group also appears low considering the pattern of data was very similar. It is possible that \( k \) values are not representative of agreement for those questions where all the literature being rated is considered equally good or equally poor by the majority of raters. \( \dagger \)Several raters in both final pilot groups reported that Questions 12 and 13 were not applicable to a publication about water births; this publication was therefore excluded from the analyses for these questions. \( \dagger \)Two questions on an earlier draft of the instrument were merged to form Question 12 on the final instrument.
Table 3
Summary of positive responses to questions concerning the validity and applicability of the instrument

<table>
<thead>
<tr>
<th>Question</th>
<th>Information group (n=15) Number (%)</th>
<th>Self help group (n=13) Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Clear and understandable questions</td>
<td>14 (93.3)</td>
<td>8 (61.5)</td>
</tr>
<tr>
<td>2 Clear and understandable instructions*</td>
<td>15 (100)</td>
<td>12 (92.3)</td>
</tr>
<tr>
<td>3 All questions necessary</td>
<td>13 (86.7)</td>
<td>9 (69.2)</td>
</tr>
<tr>
<td>4 All relevant topics included</td>
<td>11 (73.3)</td>
<td>11 (84.6)</td>
</tr>
<tr>
<td>5 All questions of equal importance when rating overall quality</td>
<td>5 (33.3)</td>
<td>8 (61.5)</td>
</tr>
<tr>
<td>6 Discriminates between publications of high and low quality</td>
<td>12 (80.0)</td>
<td>11 (84.6)</td>
</tr>
<tr>
<td>7 Applicable to a wide variety of patient publications</td>
<td>11 (73.3)</td>
<td>10 (76.9)</td>
</tr>
<tr>
<td>8 Useful for patients and information providers</td>
<td>14 (93.3)</td>
<td>11 (84.6)</td>
</tr>
</tbody>
</table>

* Full instructions are provided in the handbook. Please contact the authors for more information.

1 = No (that is, the criterion is not fulfilled by the publication) to 5 = Yes (that is, the criterion is fulfilled by the publication). A section was also included for comments. The panel recommended that the prompt questions should be developed into hints in the form of educational statements to guide the user.

Panel's Test of the Draft Instrument

An overall $\kappa$ score for each of the 20 DISCERN items was generated by calculating the average of individual $\kappa$ scores for each pair of the 11 panel members (nine non-specialist and two specialist panel members) for each of the 21 publications rated. Ten questions achieved an acceptable level of reliability, including the rating for overall quality ($\kappa=0.53$, 95% CI $\kappa=0.48$ to $\kappa=0.59$). A sensitivity analysis revealed that inter-rater agreement for each question was similar regardless of whether or not the specialists' ratings were included. The same result was found when data for each of the three conditions were analysed separately. For the majority of questions, levels of agreement were similar across the three conditions. Cases where agreement levels differed across conditions were because of the shortcomings of the draft instrument. For example, the endometriosis and myocardial infarction literature produced low agreement scores on questions about dates and sources because of incomplete guidance for rating the degree to which the publication fulfilled these criteria. In contrast, the chronic fatigue syndrome literature produced good agreement on these questions largely because these features were definitely absent from most chronic fatigue syndrome publications.

Second Meeting of the Expert Panel

During the second meeting of the expert panel the instrument was re-drafted to take account of the results of the panel's test. Modifications consisted of merging, dividing or re-wording questions. Two questions about the presentation of information were dropped from the instrument. One question asked if the information was easy to understand, and the other if it was attractive and interesting. These questions yielded the lowest $\kappa$ values despite detailed hints, and it was not possible to gain consensus on these issues during panel debate. The re-drafted instrument consisted of 15 questions plus an overall quality rating.

Discussion

The development of a critical appraisal tool for users of consumer health information that meets acceptable levels of reliability and validity is long overdue. There is a vast literature describing ways to improve the quality of written consumer health information, including many useful recommendations and guidelines. Most of this work has been targeted at producers of consumer health information and has been concerned with aspects of presentation and style such as readability. However, information that is well written and comprehensible is not necessarily accurate or informative. While there is some evidence suggesting the best ways of conveying information to patients, there has not been a robust analysis of what constitutes good quality information. Previous methods for developing and evaluating consumer health information have mainly been confined to qualitative techniques such as non-standardised questionnaires and focus groups, and have usually assessed the impact of a single publication on the knowledge, satisfaction or compliance of a specific target group. The tendency to focus on the presentation of written information and to confine evaluations to qualitative feedback from patients has hindered the development of a general set of quality guidelines about the content of written information about treatment choices. DISCERN is the first standardised quality index of consumer health information that can be used by producers, health professionals, and patients to appraise written information on treatment choices.

The combination of qualitative methods and a statistical measure of inter-rater agreement among groups representing a diverse range of experience in consumer health information was crucial to the development of DISCERN. The use of agreement statistics not only provided a demonstration of the reliability of the final instrument, but also guided the development of the instrument by focusing panel discussion. The $\kappa$ scores produced by the panel's test of the instrument highlighted areas where consensus
was difficult to achieve. The panel debated these weaknesses and refined the instrument. The exclusion of questions concerning the presentation of information demonstrates the importance of this combined methodology. Although members of the expert panel listed aspects of presentation as important quality criteria in their initial analysis, the analysis and panel debate revealed that it was impossible to identify simple “gold standards” for these features that could be used reliably by all participants. The visual appeal of written information can be highly subjective and it may not always be possible to present information in a way that is acceptable to all readers. The methodology used to develop DISCERN has enabled us to identify an agreed set of standards for the content of written information on treatment choices that can be consistently understood and applied by a wide range of users.

Although it can be misleading to compare values of $\kappa$ from different studies, as the prevalence of the categories may differ, the only recent study using a similar analysis to DISCERN was that conducted by Oxman and colleagues. In Oxman’s study, levels of inter-rater agreement on individual items of an 8 item quality assessment tool ranged from 0.21 to 0.69, with an agreement score for overall quality of 0.62. The raters in Oxman’s study were similar to the expert panel for DISCERN in that they were skilled in the application of their tool through repeated use and discussion. Oxman et al predicted that lower levels of agreement would be found among groups of untrained raters. The DISCERN instrument was tested by three groups with varying degrees of experience in critically appraising consumer health information. As would be expected, levels of agreement were consistently higher among the expert panel and the information providers compared with the self help group members. Although the information providers were less “trained” in the use of DISCERN than the expert panel, they were likely to have greater familiarity with consumer health information and concepts relating to treatment choices than self help group members.

The chance corrected agreement for the DISCERN overall quality rating reached acceptable levels among the expert panel ($\kappa=0.53$, 95% CI $\kappa=0.48$ to $\kappa=0.59$) and among the information provider group in the final test ($\kappa=0.40$, 95% CI $\kappa=0.36$ to $\kappa=0.43$), indicating that the instrument can be used successfully to discriminate between publications of high and low quality among professional groups. The overall quality rating also reflected a “fair” level of agreement ($\kappa=0.23$, 95% CI $\kappa=0.19$ to $\kappa=0.27$) among the self help group, who were “untrained” and mostly rating material concerned with health conditions other than their own. These findings indicate that the instrument can also be of use to patients, though its use may be improved by training. It is likely that effective use of DISCERN could be reached in all groups of users after training, and we have developed a handbook to support the instrument. It is important to note that use of the DISCERN instrument was not dependent on specialist knowledge of a health condition or treatment and was consistent across conditions where the evidence base differs.

Although there were differences between groups in the strength of agreement scores, the trends in scores were similar in all three groups. Higher levels of agreement were associated with the assessment of more objective criteria where there are specific indicators, such as whether a publication is current or includes additional sources of information. Questions where inter-rater agreement tended to be lowest were those where subjective judgement is required, such as the extent to which a publication is biased or refers to areas of uncertainty. Oxman et al also found that rating certain topic areas inevitably entailed some subjectivity. However, no DISCERN question was consistently identified as being of primary importance and the instrument should be used in its entirety to make judgements about the overall quality of a publication.

Patients can only make informed choices if they have access to good quality information. Members of all groups involved in the development of DISCERN reported that using the instrument led to the adoption of good practice in selecting and producing consumer health literature. The critical appraisal skills acquired from using DISCERN caused many participants to challenge their assumption that some information is always better than none. The rapid growth of medical information publicly available, particularly on the world wide web, has led to calls for a set of basic quality standards to ensure that professionals and patients know which information to use and which to discard. The DISCERN instrument will improve the standard of information available to the public in all settings by providing a framework for the production, evaluation, and screening of written consumer health information. Furthermore, by improving standards of consumer health information, DISCERN will play a vital part in the delivery of effective

### KEY POINTS

- Patients require good quality evidence-based information so they can take an active role in decisions about their health care.
- Currently patients and providers of health information have no systematic way of judging the quality of information on treatment choices.
- Written and electronic information for patients is growing at an unprecedented rate.
- The DISCERN instrument has been developed to enable patients and information providers to judge the quality of written information about treatment choices; and to facilitate the production of high quality evidence-based patient information.
healthcare by enabling patients to make informed treatment choices based on good evidence.

Our thanks to Mike Bradburn, Medical Statistician, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford for statistical advice; Professor Stephen Walter, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5 for providing the PC-Agree Program for the calculation of weighted x; members of the expert panel and national pilot groups; and staff at the Popular Medical Index, Lister Hospital, Stevenage for the provision of magazine and newspaper articles.

Funding: this work was supported by a grant from the British Library and NHS Executive Anglia and Oxford.

Conflicts of interest: none.

Appendix

DISCERN
An instrument for judging the quality of written consumer health information on treatment choices Funded by the British Library

For further information please contact: Sasha Shepperd University of Oxford Division of Public Health and Primary Care Institute of Health Sciences Old Road Headington Oxford OX3 7LF

Section 1

IS THE PUBLICATION RELIABLE?

1. Are the aims clear?

No 2 Partially 3 4 Yes 5

Hint: Look for a clear indication at the beginning of the publication of what it is about (what topics are meant to be excluded) who might find it useful. If the answer to Question 1 is 'No', go directly to Question 3

2. Does it achieve its aims?

No 2 Partially 3 4 Yes 5

Hint: Consider whether the publication provides the information it aimed to as outlined in Question 1

3. Is it relevant?

No 2 Partially 3 4 Yes 5

Hint: Consider whether the publication addresses the questions that readers might ask. Recommendations and suggestions concerning treatment choices are realistic or appropriate

4. Is it clear what sources of information were used to compile the publication (other than the author or producer)?

No 2 Partially 3 4 Yes 5

Hint: Check whether the main claims or statements made about treatment choices are accompanied by a reference to the sources used as evidence (e.g. a research study or expert opinion). Look for a means of checking the sources used such as a bibliography/reference list or the addresses of the experts or organisations quoted.

Rating note: In order to score a full '5' the publication should fulfill both hints. Lists of additional sources of support and information (Q.7) are not necessarily sources of evidence for the current publication

5. Is it clear when the information used or reported in the publication was produced?

No 2 Partially 3 4 Yes 5

Hint: Look for dates of the main sources of information used to compile the publication. Date of any revisions of the publication (but not dates of reprinting) date of publication (copyright date)

Rating note: The hints are placed in order of importance - in order to score a full '5' the dates relating to the first hint should be found

6. Is it balanced and unbiased?

No 2 Partially 3 4 Yes 5

Hint: Look for a clear indication of whether the publication is written from a personal or objective point of view: evidence that a range of sources of information was used to compile the publication (e.g. more than one research study or expert). Evidence of an external assessment of the publication. Be wary if the publication focuses on the advantages or disadvantages of one particular treatment choice without reference to other possible choices. The publication relies primarily on evidence from single cases (which may not be typical of people with this condition or of responses to a particular treatment). The information is presented in a sensational, emotive or alarmist way

7. Does it provide details of additional sources of support and information?

No 2 Partially 3 4 Yes 5

Hint: Look for suggestions for further reading or for details of other organisations providing advice and information about the condition and treatment choices

8. Does it refer to areas of uncertainty?

No 2 Partially 3 4 Yes 5

Hint: Look for discussion of the gaps in knowledge or differences in expert opinion concerning treatment choices. Be wary if the publication implies that a treatment choice affects everyone in the same way (e.g. a 100% success rate with a particular treatment)

Section 2

HOW GOOD IS THE QUALITY OF INFORMATION ON TREATMENT CHOICES?

N.B. The questions apply to the treatment (or treatments) described in the publication. Self-care is considered a form of treatment throughout this section.

9. Does it describe how each treatment works?

No 2 Partially 3 4 Yes 5

Hint: Look for a description of how a treatment acts on the body to achieve its effect

10. Does it describe the benefits of each treatment?

No 2 Partially 3 4 Yes 5

Hint: Benefits can include controlling or getting rid of symptoms, preventing recurrence of the condition and eliminating the condition - both short-term and long-term

11. Does it describe the risks of each treatment?

No 2 Partially Yes 5

Hint: Risks can include side effects, complications and adverse reactions to treatment - both short-term and long-term
Judging the quality of written consumer health information on treatment choices

12. Does it describe what would happen if no treatment is used?

No | Partially | Yes
---|---|---
1 | 2 | 3 | 4 | 5

Hint: Look for a description of the risks and benefits of postponing treatment, of watchful waiting (i.e. monitoring how the condition progresses without treatment) or of permanently forgoing treatment.

13. Does it describe how the treatment choices affect overall quality of life?

No | Partially | Yes
---|---|---
1 | 2 | 3 | 4 | 5

Hint: Look for a description of the effects of the treatment choices on day-to-day activity.* description of the effects of the treatment choices on relationships with family, friends and carers.

14. Is it clear that there may be more than one possible treatment choice?

No | Partially | Yes
---|---|---
1 | 2 | 3 | 4 | 5

Hint: Look for a description of who is most likely to benefit from each treatment choice mentioned, and under what circumstances suggestions of alternatives to consider or investigate further (including choices not fully described in the publication) before deciding whether to select or reject a particular treatment choice.

15. Does it provide support for shared decision-making?

No | Partially | Yes
---|---|---
1 | 2 | 3 | 4 | 5

Hint: Look for suggestions of things to discuss with family, friends, doctors or other health professionals concerning treatment choices.

Section 3

OVERALL RATING OF THE PUBLICATION

16. Based on the answers to all of the above questions, rate the overall quality of the publication as a source of information about treatment choices:

Low | Moderate | High
---|---|---
Serious or extensive shortcomings | Potentially important but not serious shortcomings | Minimal shortcomings

1 | 2 | 3 | 4 | 5

Copyright British Library and the University of Oxford 1997
