

## Developing a quality criteria framework for patient decision aids: online international Delphi consensus process

Glyn Elwyn, Annette O'Connor, Dawn Stacey, Robert Volk, Adrian Edwards, Angela Coulter, Richard Thomson, Alexandra Barratt, Michael Barry, Steven Bernstein, Phyllis Butow, Aileen Clarke, Vikki Entwistle, Deb Feldman-Stewart, Margaret Holmes-Rovner, Hilary Llewellyn-Thomas, Nora Mounjid, Al Mulley, Cornelia Ruland, Karen Sepucha, Alan Sykes, Tim Whelan, on behalf of the International Patient Decision Aids Standards (IPDAS) Collaboration

### Abstract

**Objective** To develop a set of quality criteria for patient decision support technologies (decision aids).

**Design and setting** Two stage web based Delphi process using online rating process to enable international collaboration.

**Participants** Individuals from four stakeholder groups (researchers, practitioners, patients, policy makers) representing 14 countries reviewed evidence summaries and rated the importance of 80 criteria in 12 quality domains on a 1 to 9 scale. Second round participants received feedback from the first round and repeated their assessment of the 80 criteria plus three new ones.

**Main outcome measure** Aggregate ratings for each criterion calculated using medians weighted to compensate for different numbers in stakeholder groups; criteria rated between 7 and 9 were retained.

**Results** 212 nominated people were invited to participate. Of those invited, 122 participated in the first round (77 researchers, 21 patients, 10 practitioners, 14 policy makers); 104/122 (85%) participated in the second round. 74 of 83 criteria were retained in the following domains: systematic development process (9/9 criteria); providing information about options (13/13); presenting probabilities (11/13); clarifying and expressing values (3/3); using patient stories (2/5); guiding/coaching (3/5); disclosing conflicts of interest (5/5); providing internet access (6/6); balanced presentation of options (3/3); using plain language (4/6); basing information on up to date evidence (7/7); and establishing effectiveness (8/8).

**Conclusions** Criteria were given the highest ratings where evidence existed, and these were retained. Gaps in research were highlighted. Developers, users, and purchasers of patient decision aids now have a checklist for appraising quality. An instrument for measuring quality of decision aids is being developed.

### Introduction

Clinical guidelines often recommend that healthcare professionals should involve patients in decisions about screening, treatment, and other interventions,<sup>1-3</sup> to help them to arrive at informed choices. Patient decision aids are designed to support patients in this process<sup>4</sup>; they are intended to supplement rather

than replace patient-practitioner interaction. They may be leaflets, interactive media, or video or audio tapes. Patients may use them to prepare for talking with a clinician, or a clinician may provide them at the time of a visit to facilitate decision making. At a minimum, patient decision aids provide information about the options and their associated relevant outcomes. These technologies also help patients to personalise this information, to understand that they can be involved in choosing among the various options, to appreciate the scientific uncertainties inherent in that choice, to clarify the personal value or desirability of potential benefits relative to potential harms, to communicate their values to their practitioners, and to gain skills in the steps of collaborative decision making.<sup>1</sup>

Evidence describing the effectiveness and feasibility of patient decision aids is substantial.<sup>5-8</sup> Trials indicate that decision aids are superior to standard counselling in improving patients' knowledge and realistic expectations about the results of treatments and other procedures. In most studies, outcomes such as perceived involvement, agreement between values and choice, and decisional conflict have changed in a desirable positive direction.<sup>5</sup> Decision aids can also affect the uptake of options, reducing the use of some procedures (such as fewer mastectomies in favour of breast conservation surgery<sup>9</sup> or a reduction in hysterectomy rates<sup>10</sup>), and increasing the use of others (such as the uptake of colon cancer screening<sup>5</sup>). These effects are desirable when decision aids are unbiased and the motivation is to rectify variations in practice due to poor comprehension or disregarding of patients' preferences. However, concerns will emerge if decision aids affect uptake rates because of bias or inaccuracy.

By 1999, approximately 15 patient decision aids had been developed in academic institutions. More than 500 now exist, produced largely by a mix of not for profit and commercial organisations. Many are easily available on the internet.<sup>11</sup> However, their quality varies; some do not cite their evidence sources, and others have presentational biases. Furthermore, debate exists about underlying concepts<sup>12</sup> and about the lack of agreed quality criteria for these tools. Because patient decision aids can have an important influence on choices made,<sup>13</sup> developers need to have followed recognised methods, avoid bias, and cite valid evidence sources.



An extra table, a statistical appendix, and a checklist are on [bmj.com](http://bmj.com)

Acting on this need, the International Patient Decision Aids Standards (IPDAS) Collaboration was established at the 2nd International Shared Decision Making conference (Swansea, 2003). The proposal to generate a quality framework was also supported at the Society for Medical Decision Making (Chicago, 2003) and Society for Information Therapy (Utah, 2003). We reviewed existing checklists for assessing the quality of randomised trials (CONSORT),<sup>14</sup> meta-analyses (QUOROM),<sup>15</sup> practice guidelines (AGREE),<sup>16</sup> and general patient information.<sup>17–19</sup> However, patient decision aids differ from scientific studies and practice guidelines and aim to do more than provide general information for patients. They are interventions that recognise the need for both patients and professionals to consider, at the individual level, the impact of uncertainties surrounding many healthcare decisions; they communicate risk probabilities and use methods to clarify values and guide deliberation. Moreover, they use powerful and potentially misleading strategies that have not been used in other quality checklists. Aware of best practice regarding the development of quality criteria,<sup>20</sup> IPDAS adapted an approach used for appraising clinical guidelines (AGREE collaboration)<sup>16</sup> and established an international collaboration of different stakeholder groups. The aim was to achieve an international consensus based framework of quality criteria for patient decision aids that would act as a checklist for developers and users.

## Methods

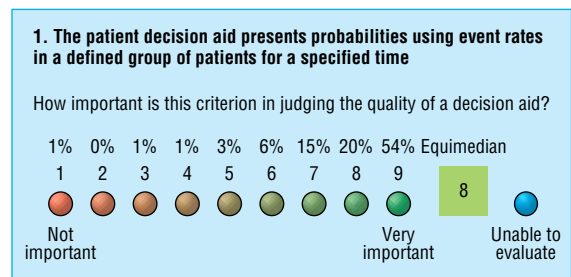
Substantive research evidence about the overall effectiveness of patient decision aids exists,<sup>5</sup> but little information is available about which components and processes are most influential for improving “decision quality.”<sup>13</sup> IPDAS therefore decided that criteria should be developed by using a recognised consensus based approach,<sup>21–22</sup> capable of integrating empirical evidence where it exists and also the views of experts and a range of other people, such as informed stakeholders, patients, health professionals, policy makers, and potential purchasers of the tools. We therefore established a Delphi consensus process.

### Delphi consensus technique and study management

Considerable experience in using Delphi consensus techniques exists,<sup>21–23–24</sup> but few researchers have used the methods to develop quality criteria among different stakeholder groups on an international basis.<sup>25–26</sup> To manage the task, we convened four groups: a strategic steering group oversaw the project, an evidence review group prepared background documents, a methods group specified a two stage web based criteria rating process (adapting the RAND appropriateness rating system<sup>27</sup>), and a stakeholder selection group identified the people who would be invited to serve as raters. The process comprised the following steps.

**Defining quality domains**—Delegates at the 2003 conference identified an initial list of quality areas from previous work<sup>5</sup> and expanded it into 12 broad quality domains (see extra table on [bmj.com](http://bmj.com)). Members of the shared decision making electronic listserve, composed of 181 interested academics and practitioners, discussed the validity of the 12 domains. Next, we used these 12 broad quality domains to specify which background evidence reports were needed.

**Developing background evidence reports**—Twelve panels (a total of 50 international experts) prepared “background evidence reports” for each quality domain.<sup>28</sup> Each report included definitions of key concepts; theoretical links between the domain and decision quality<sup>13</sup>; and evidence to support the inclusion or exclusion of suggested domain criteria, including fundamental



**Fig 1** Example of a criterion (second round). Quality domain=presenting probabilities

studies<sup>28</sup> and results from the systematic review of 34 randomised trials.<sup>5</sup> From these reports (available online<sup>28</sup>), we drafted quality criteria.

**Producing quality criteria**—We subjected the quality criteria to iterative consultation about comprehensiveness and subsequent editing by the steering group, the methods group, the evidence review groups panels,<sup>28</sup> and finally by a plain language expert, before a pilot rating exercise. We established a final set of 80 quality criteria.

**Establishing participant stakeholder groups**—We considered four stakeholder groups to be relevant: patients, health practitioners, policy makers, and decision aid developers and researchers. The collaboration decided that the framework would represent views among stakeholder groups equally, on the basis of the view that decision aids should reflect a balance, if possible, between positions taken by patients, researchers, clinicians, and society at large over the attribution of priorities and choices. We based statistical analyses (see below) on this intent. Potential participants were nominated by the IPDAS Collaboration, by the Cochrane Collaboration Consumers Group, and by word of mouth among the related networks. The inclusion criteria were familiarity or awareness of patient decision aids and an ability to provide ratings within a specified time window. The researcher and developer group was over-represented to allow wide participation, and we weighted the group ratings to ensure equal contribution (see analysis).

**Rating quality criteria**—We invited nominated participants by email to complete a two stage rating process, gave them access to a password protected website,<sup>28</sup> and asked them to complete a short demographic questionnaire. For each quality domain, we asked participants to read a short summary of the background reports (full text was available by URL) and then rate the importance of quality criteria on a scale from 1=not important to 9=very important. Raters could also choose to add free text comments (fig 1). We sent two email reminders in each round. At the second round, we presented raters with a summary of the results for each domain and the first round ratings for each of the criteria (fig 1).

### Analysis of ratings

After the first round, we calculated aggregate ratings and summarised comments. To ensure equal weighting for each stakeholder group in the overall rating, we obtained a weighted median by calculating a separate empirical cumulative distribution function for each group. We estimated the empirical cumulative distribution function for a population with equal numbers in each stakeholder group by taking an equally weighted sum of these functions. We calculated the median of this distribution (equimedian) by finding the value for which this function equals 0.5. For further details on calculation of an equimedian, see statistical appendix on [bmj.com](http://bmj.com). We based thresholds for retaining quality criteria in the framework on the

## Participants' characteristics and responses by stakeholder group and country

|                     | No of participants invited | Round 1 |                   | Round 2 |                   |
|---------------------|----------------------------|---------|-------------------|---------|-------------------|
|                     |                            | No      | Response rate (%) | No      | Response rate (%) |
| <b>Stakeholders</b> |                            |         |                   |         |                   |
| Patients            | 44                         | 21      | 48                | 19      | 91                |
| Practitioners       | 18                         | 10      | 56                | 9       | 90                |
| Policy makers       | 25                         | 14      | 56                | 9       | 64                |
| Researchers         | 125                        | 77      | 62                | 67      | 87                |
| Total               | 212                        | 122     | 58                | 104     | 85                |
| <b>Country</b>      |                            |         |                   |         |                   |
| Australia           | 18                         | 12      | 67                | 11      | 92                |
| UK                  | 44                         | 18      | 41                | 16      | 89                |
| USA                 | 65                         | 45      | 69                | 40      | 89                |
| Canada              | 50                         | 30      | 60                | 24      | 80                |
| China               | 1                          | 0       | 0                 | 0       | 0                 |
| Finland             | 2                          | 0       | 0                 | 0       | 0                 |
| France              | 7                          | 5       | 71                | 4       | 80                |
| Germany             | 2                          | 1       | 50                | 1       | 100               |
| Israel              | 1                          | 1       | 0                 | 1       | 100               |
| Italy               | 1                          | 1       | 0                 | 1       | 100               |
| Kenya               | 1                          | 0       | 0                 | 0       | 0                 |
| Netherlands         | 8                          | 4       | 50                | 2       | 50                |
| New Zealand         | 1                          | 1       | 100               | 1       | 100               |
| Norway              | 5                          | 1       | 20                | 1       | 100               |
| Spain               | 1                          | 1       | 100               | 1       | 100               |
| Switzerland         | 3                          | 1       | 33                | 1       | 100               |
| Uruguay             | 2                          | 1       | 50                | 0       | 0                 |
| Total               | 212                        | 122     | 58                | 104     | 85                |

overall equimedian and the level of disagreement among participants at the second round. We considered that participants “disagreed” if 30% or more of the ratings were in the lower third (ratings 1-3) and 30% or more of the ratings were in the upper third (ratings 7-9). We regarded quality criteria with an overall equimedian rating of 7 to 9 (without disagreement) as “important” and included them. We considered criteria rated as 4 to 6 (without disagreement) to be “equivocal.” We regarded criteria rated with an equimedian of 1 to 3 as “not important.” We also considered criteria that exhibited disagreement to be not important. We based these thresholds on values used in other settings.<sup>21</sup> We used non-parametric analyses of variances to calculate the potential impact of differences between stakeholder groups' ratings.<sup>29</sup> We did sensitivity analyses to establish whether the exclusion of stakeholder groups had an effect on the overall results.

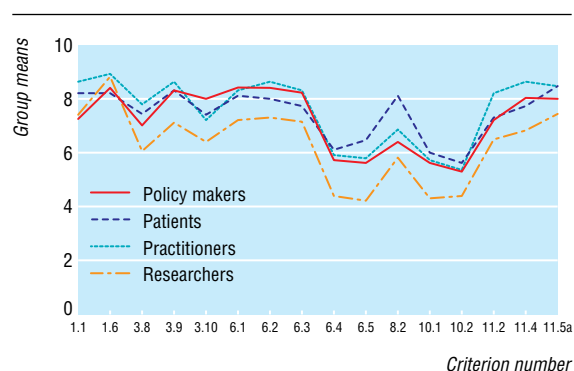
## Results

Although listserv participants debated whether the domain labelled “patient stories” should be included, we retained all 12 quality domains suggested for criterion development. Listserv participants made no additions. We included the following quality domains in the final quality criterion framework: (1) systematic development process; (2) providing information about options; (3) presenting probabilities; (4) clarifying and expressing values; (5) using patient stories; (6) guiding or coaching in deliberation and communication; (7) disclosing conflicts of interest; (8) delivering patient decision aids on the internet; (9) balancing the presentation of options; (10) using plain language; (11) basing information on up to date scientific evidence; and (12) establishing effectiveness.

The table describes the participants in the first and second rounds. We invited 212 people to the Delphi process (125 researchers/researcher practitioners, 44 patients, 25 policy makers,

and 18 health professionals). Of those invited, 122 provided ratings at the first round; 104/122 (85%) participants completed both rating rounds. Participants were from 14 countries, although most were from the United States (65), Canada (50), the United Kingdom (44), and Australia (18).

The free text comments prompted the addition of three new criteria for the second round (1.8b, 8.6, and 11.5b). The extra table on *bmj.com* reports the equimedian ratings achieved for each criterion after the second round. Of the 83 criteria, 41 were given an overall equimedian rating of 9, 28 a rating of 8, and 7 a rating of 7. Eight criteria (3.11, 3.12, 5.1, 5.3, 5.4, 6.4, 6.5, and 10.2) had equivocal ratings (rated 4 to 6 without disagreement). Of the 13 criteria in the domain “presenting probabilities,” two had equimedian ratings of 5 and 6 (3.11 “describe how probabilities were calculated” and 3.12 “describe how probabilities were derived for patient subgroups”). Three of the five criteria in the domain “patient stories” (5.1, 5.3, and 5.4) all had equimedian ratings of 6. Two of two criteria that focused on offering a “trained coach to prepare patients to discuss decisions with their



**Fig 2** Criteria for which stakeholder effects were present at second round (see table for data and P values)

practitioners” had equimedians ratings of 5 (6.4 and 6.5). None of the criteria had evidence of disagreement, although two criteria (10.1 and 10.2 in the “plain language” domain) came close to this threshold.

For 16 criteria, evidence existed of significant differences between stakeholder groups’ ratings (table, fig 2). Compared with other stakeholder groups, researchers generally gave lower ratings to criteria. Although these group differences achieved statistical significance, only five criteria had medians that straddled the threshold for inclusion. Exclusion of any one set of stakeholder results did not change the overall inclusion or exclusion of criteria.

## Discussion

### Principal findings

This Delphi process, supported by summarised evidence reports, has provided substantial consensus about a framework of quality criteria for patient decision aids. The decision aid criteria that were most strongly endorsed also had the greatest empirical support. In addition, the process revealed areas of disagreement and opportunities for further research. Where stakeholder groups’ ratings differed, the researcher group tended to give lower ratings, presumably because these participants were more conservative about the feasibility of simultaneously achieving a large number of quality criteria and perhaps more aware of the difficulty in obtaining supportive empirical evidence. For example, the use of “patient stories” in decision aids caused considerable discussion. Some researchers believe that patient decision aids should avoid using patient stories until their impact is better understood. Concerns exist because patient stories have the potential to introduce significant bias and depend on how the stories are selected and presented.<sup>30</sup> Given that decision making is strongly influenced by self identification with “similar others,” this area needs further investigation. The values clarification technique of describing the physical, emotional, and social effects of options to help patients to explore “experienced utility” was strongly endorsed.<sup>31</sup> Another area of disagreement was about the use of trained coaches to prepare patients to discuss options; although the addition of coaching has shown positive effects,<sup>10</sup> worries exist about its feasibility, and further empirical work would be of value.

The endorsed criteria are available as a checklist (see [bmj.com](http://bmj.com)). In this checklist, the first subset of criteria (content) refer to the information, probabilities, values clarification, and guidance in deliberation that are context specific—that is, specific to the health condition and therapeutic/screening options covered by a particular patient decision aid. The second subset of criteria (development process) are generic, in that they refer to design and developmental criteria that are relevant to all patient decision aids, regardless of the health context. The third subset of criteria (effectiveness) are also generic, in that they refer to the general principles of fostering a high quality decision process and a high quality choice. This checklist enables the users of existing decision aids, such as patients and health professionals, to assess the content, development process, and effectiveness of patient decision aids they encounter. The framework for the quality criteria that appear in the checklist, in conjunction with the IPDAS Collaboration’s supporting background evidence documents,<sup>28</sup> form an important resource for the developers of new decision aids who need empirical evidence about the different components and processes required to produce a high quality decision aid. In this sense, the checklist is comparable to the GRADE working group outputs.<sup>32</sup>

Furthermore, the quality criteria that appear in the checklist could guide researchers in the decision sciences to create a validated quality assessment scale that could generate quantitative scores. These quantitative data could, in turn, be used in comparative studies of patient decision aids and in systematic reviews of these technologies.<sup>3</sup> In this sense, the checklist is comparable to the AGREE tool, which is used for the assessment of clinical guidelines.<sup>16</sup>

### Weaknesses and strengths

A potential weakness of the study is the extent to which the participants were not independent of the research agenda. Although summaries of empirical evidence were presented, some raters might have considered patient decision aids to offer more advantages than disadvantages and could have introduced bias. We note, however, that consistency was high across groups and that, where differences appeared, researchers generally gave lower ratings than the other stakeholders. A second possible weakness is that we asked the participants to rate the criteria against only the “importance” of the criterion for the quality of a decision aid. Ideally, factors such as measurability and feasibility would have been included.<sup>27</sup> However, we opted for high response rates and attempted not to overburden participants.

A strength of the study is the appropriate use of a Delphi consensus process.<sup>21</sup> The method generated a wide “ownership” of the exercise and involvement of many recognised research groups in this field. Moreover, we took care to ensure the availability of existing empirical evidence,<sup>28</sup> the use of plain language, and that equal weighting was given to stakeholder groups’ ratings.

### Results in context

This study represents the first international effort to build on the work of the Cochrane Collaboration’s systematic review group and establish a normative consensus on quality criteria for patient decision aids.<sup>5</sup> We recognise that the checklist contains a substantial number of criteria and might be considered to represent an “ideal” construction that may be difficult to attain. However, this quality framework emphasises the need to strive for designs that have favourable effects on decision quality. The criteria are not meant to be prescriptive; many different ways of achieving the same ends exist, and an unresolved debate remains about what constitutes a minimum set of domains and criteria that should be met by a patient decision aid. Nevertheless, we believe that the IPDAS Collaboration has set an agenda for both developers and researchers. The development of future decision aids should also be based on theoretical underpinning<sup>12</sup> and on the measurement of appropriate outcomes, in order to determine whether patient decision aids accomplish their primary objective—to improve the quality of decisions (the extent to which patients’ decisions are consistent with their informed values).<sup>13</sup>

### Implications

Quality criteria for patient decision aids are relevant to patients, healthcare professionals, healthcare service purchasers, and policy makers, all of whom need to be confident about the development and testing that these tools have undergone before their release. The IPDAS Collaboration checklist is designed for existing users and for developers of new decision aids. To take the field forward, two things are now needed: firstly, to use the IPDAS framework as a basis for developing a validated instrument for assessing what could be termed the internal quality (content and technology) of the decision tool; secondly, to develop an agreed way to measure with more precision the impact of such tools

## What is already known on this topic

Decision support technologies for patients (also known as decision aids) have received increasing interest over the past decade

A systematic review of randomised controlled trials confirmed many positive outcomes when these tools are used by patients and healthcare providers

No agreement exists about the content of the active components of decision aids, and no guidance exists about quality standards for their development and evaluation

## What this study adds

A Delphi process, supported by summarised evidence reports, has provided substantial consensus about a framework of quality criteria for patient decision aids

The criteria are available as a users' checklist and are being used as a guide to developers of decision support

(their efficacy and effectiveness) on a range of outcomes. At that point that we may be able to examine in more depth the goal of helping patients and their advisers to make the best healthcare decisions.

The following people contributed to background evidence documents: J Austoker (UK), H Bekker (UK), J Belkora (USA), C Braddock (USA), P Butow (AU), E Chan (USA), A Charvet (Switzerland), J Davison (Canada), J Dolan (USA), A Fagerlin (USA), J Fowler (USA), D Frosch (USA), P Hewitson (UK), T Hope (UK), M Jacobsen (Canada), M O'Kane (USA), A Kennedy (Switzerland), S Knight (USA), M Kupperman (USA), B Ling (USA), T Marteau (UK), K McCaffery (Australia), M O'Connor (USA), E Ozanne (USA), M Pignone (USA), A Raffle (UK), L Schwartz (USA), S Sheridan (USA), S Stableford (USA), D Stilwell (USA), V Tait (Canada), D Timmermans (Netherlands), L Trevena (Australia), C Wills (USA), S Woloshin (USA), S Ziebland (UK). We also thank the following for their assistance: P Shekelle (USA), J Muir Gray (UK), P Tugwell (Canada), J Wenberg (USA), and the other participants in the consensus process. We acknowledge Raymond Ramirez and J Scott Smith for development and management of the Delphi process website.

Contributors: GE, AO'C, DS, Michael Barry, Steven Bernstein, Richard Thomson, AC, and AE were responsible for the study concept, design, and management. AO'C, Hilary Llewellyn-Thomas, DS, and the chairs of the evidence groups were responsible for producing background documents. DS and RV were responsible for data acquisition, and GE, AO'C, and DS were responsible for the analysis and interpretation. GE and AO'C drafted the manuscript, and all the authors revised it for important intellectual content. Alan Sykes did the statistical analysis. AO'C and GE obtained funding for the work. GE is the guarantor.

Funding: Canadian Institutes of Health Research Group grant; Cardiff University internal funding.

Competing interests: AO'C receives financial support from the not for profit Foundation for Informed Medical Decision Making, Boston. This foundation receives royalties from Health Dialog, a commercial producer and promoter of patient decision aids. Other authors: none declared.

- O'Connor AM, Llewellyn-Thomas HA, Flood AB. Modifying unwarranted variations in health care: shared decision making using patient decision aids. *Health Aff (Millwood)* 2004;suppl web exclusive:VAR63-72.
- O'Connor AM, Fiset V, Degrasse C, Graham I, Evans W, Stacey D, et al. Decision aids for patients considering health care options: evidence of efficacy and policy implications. *J Natl Cancer Inst* 1999;25(monograph):67-80.
- Barratt A, Trevena L, Davey HM, McCaffery K. Use of decision aids to support informed choices about screening. *BMJ* 2004;329:507-10.
- Woolf SH, Chan EC, Harris R, Sheridan SL, Braddock CH, Kaplan RM, et al. Promoting informed choice: transforming health care to dispense knowledge for decision making. *Ann Intern Med* 2005;143:293-300.
- O'Connor AM, Stacey D, Entwistle V, Llewellyn-Thomas H, Rovner D, Holmes-Rovner M, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2003;(1):CD001431.

- Molenaar S, Sprangers MA, Postma-Schuit FC, Rutgers EJ, Noorlander J, Hendricks J, et al. Feasibility and effects of decision aids. *Med Decis Making* 2000;20:112-27.
- Estabrooks C, Goel V, Thiel E, Pinfold P, Sawka C, Williams I. Decision aids: are they worth it? A systematic review. *J Health Serv Res Policy* 2001;6:170-82.
- Entwistle V. The potential contribution of decision aids to screening programmes. *Health Expect* 2001;4:109-15.
- Whelan T, Levine M, Willan A, Gafni A, Sanders K, Mirsky D, et al. Effect of a decision aid on knowledge and treatment decision making for breast cancer surgery: a randomized trial. *JAMA* 2004;292:435-41.
- Kennedy AD, Sculpher MJ, Coulter A, Dwyer N, Rees M, Abrams KR, et al. Effects of decision aids for menorrhagia on treatment choices, health outcomes, and costs. *JAMA* 2002;288:2701-8.
- Evans R, Elwyn G, Edwards A. Making interactive decision support for patients a reality. *Inform Prim Care* 2004;12:109-13.
- Charles C, Gafni A, Whelan T, O'Brien MA. Treatment decision aids: conceptual issues and future directions. *Health Expect* 2005;8:114-25.
- Sepucha KR, Fowler FJ, Mulley AG. Policy support for patient-centered care: the need for measurable improvements in decision quality. *Health Aff (Millwood)* 2004;suppl web exclusive:VAR54-62.
- Altman DG. Better reporting of randomized controlled trials: the CONSORT statement. *BMJ* 1996;313:570-1.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. *Lancet* 1999;354:1896-900.
- Cluzeau FA, Burgers JS, Brouwers M, Grol R, Mäkelä M, Littlejohns P, et al. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003;12:18-23.
- Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999;53:105-11.
- Coulter A, Entwistle V, Gilbert D. *Informing patients: an assessment of the quality of patient information materials*. London: King's Fund, 1998.
- Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ* 2002;324:569-72.
- Heaney DJ, Walker JJ, Howie JG, Maxwell M, Freeman GK, Berrey PN, et al. The development of a routine NHS data-based index of performance in general practice (NHSPPD). *Fam Pract* 2002;19:77-84.
- Campbell SM, Cantrill JA, Roberts D. Prescribing indicators for UK general practice: Delphi consultation study. *BMJ* 2000;321:425-8.
- Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;2(3):1-88.
- Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376-80.
- Jones J, Hunter D. Using the Delphi and nominal group technique in health services research. In: Mays N, Pope C, eds. *Qualitative research in health care*. London: BMJ Books, 1999.
- Engels Y, Campbell S, Dautzenberg M, van den Hombergh P, Brinkmann H, Szécsényi J, et al. Developing a framework of, and quality indicators for, general practice management in Europe. *Fam Pract* 2005;22:215-22.
- Fitch K, Lázaro P, Aguilar MD, Kahan J, van het Loo M, Bernstein SJ. European criteria for the appropriateness and necessity of coronary revascularization procedures. *Eur J Cardiothorac Surg* 2000;18:380-7.
- Brook RH. Appropriateness: the next frontier. *BMJ* 1994;308:217-8.
- International Patient Decision Aid Standards (IPDAS) Collaboration. Background evidence reports from expert panels for 12 quality domains. <http://ipdas.ohri.ca/> (accessed Jan 2006).
- Conover WJ. *Practical nonparametric statistics*. 2nd ed. New York: John Wiley & Sons, 1980.
- Ubel PA, Loewenstein G. The role of decision analysis in informed consent: choosing between intuition and systematicity. *Soc Sci Med* 1997;44:647-56.
- Kahnemann D. Experienced utility and objective happiness: a moment-based approach. In: Kahnemann D, Tversky A, eds. *Choices, values and frames*. New York: Cambridge University Press and Russell Sage Foundation, 2000:673-92.
- Schunemann HJ, Best D, Vist G, Oxman AD, Grade Working Group. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003;169:677-80.

(Accepted 13 July 2006)

Department of General Practice, Centre for Health Sciences Research, Cardiff University, Cardiff CF14 4YS

Glyn Elwyn *research professor, primary medical care*

Adrian Edwards *research professor, primary medical care*

University of Ottawa and Ottawa Health Research Institute, Clinical Epidemiology Program, Ottawa, ON, Canada K1Y 4E9

Annette O'Connor *professor*

School of Nursing, University of Ottawa, Ottawa, ON, Canada K1H 8M5

Dawn Stacey *assistant professor*

Department of Family and Community Medicine, Baylor College of Medicine, Houston, TX 77098-3915, USA

Robert Volk *associate professor*

Picker Institute Europe, King's Mead House, Oxford OX1 1RX

Angela Coulter *chief executive*

Newcastle upon Tyne Medical School, School of Population and Health Sciences, Newcastle upon Tyne NE2 4HH

## Research

---

**Richard Thomson** *professor of epidemiology and public health*

Screening and Test Evaluation Program, School of Public Health, University of Sydney, NSW 2006, Australia

**Alexandra Barratt** *associate professor, epidemiology*

Medical Psychology Research Unit, School of Psychology, University of Sydney

**Phyllis Butow** *professor*

Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA

**Michael Barry** *chief, general medicine unit*

**Albert G Mulley** *chief, general medicine division*

**Karen Sepucha** *senior scientist*

HSR&D Center of Excellence, VA Ann Arbor Healthcare System, USA

**Steven Bernstein** *research scientist*

Public Health Resource Unit, Oxford OX4 2GX

**Aileen Clarke** *consultant in public health*

Social Dimensions of Health Institute, University of Dundee, Dundee DD1 4HJ

**Vikki Entwistle** *reader*

Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's University, Kingston, ON, Canada

**Deb Feldman-Stewart** *associate professor*

Center for Ethics, Michigan State University, East Lansing, MI 48824, USA

**Margaret Holmes-Rovner** *professor*

Center for the Evaluative Clinical Sciences, Department of Community and Family Medicine, Dartmouth Medical School, Hanover, NH, USA

**Hilary Llewellyn-Thomas** *professor*

GRESAC, Centre Léon Bérard, 69008 Lyon, France

**Nora Moumjid** *health economist*

Rikshospitalet Radiumhospitalet, Oslo, Norway

**Cornelia Ruland** *professor*

Acadvent Ltd, 171 Gower Road, Swansea

**Alan Sykes** *statistician*

Department of Medicine, McMaster University and Juravinski Cancer Centre,

Hamilton, ON, Canada

**Tim Whelan** *professor*

Correspondence to: Glyn Elwyn [elwyn@cardiff.ac.uk](mailto:elwyn@cardiff.ac.uk)

doi 10.1136/bmj.38926.629329.AE

This is Version 2 of the article, which includes the full list of authors.