International Journal for Quality in Health Care 2006; **Volume 18, Number 5:** pp. 365–369
Advance Access Publication: 7 September 2006
10.1093/intqhc/mzl040

Considered judgement in evidence-based guideline development

KARIN VERKERK[1,2,6], HASKE VAN VEENENDAAL[1], JOHAN L. SEVERENS[2,3], ERIK J. M. HENDRIKS[4,5], AND JAKO S. BURGERS[1]

[1]Dutch Institute for Healthcare Improvement (CBO), Utrecht, [2]Department of Health, [3]Department of Clinical Epidemiology, University of Maastricht, Maastricht, [4]Department of Research, Dutch Institute of Allied Health Care, Amersfoort, [5]Department of Epidemiology, University of Maastricht, Maastricht and [6]Hogeschool Rotterdam, Department of Physiotherapy, Rotterdam, The Netherlands

Abstract

Background. Clinical practice guidelines should be based on the best available evidence. However, this evidence is often incomplete, controversial, or lacking. Other considerations beyond the evidence are therefore needed to be able to formulate specific and applicable recommendations for clinical practice.

Objective. The aim of this study is to obtain consensus among experts about a set of domains and items covering the most relevant 'other considerations' to formulate recommendations in evidence-based guideline development.

Methods. An initial list of 10 domains and 49 items for a systematic and considered judgement of scientific evidence was generated from the literature. A panel of Dutch experts in guideline development tested this list using a two-round Delphi consensus technique. Each expert was asked to independently score the relevance of the items on a 4-point Likert scale, ranging from 'very important' to 'not important'. The final list consisted of items that were included by at least 60% consensus.

Results. Twenty-eight experts participated in the first Delphi round and 21 of them in the second round. High scoring domains were 'clinical relevance', 'safety', and 'availability of resources'. There was consensus about the relevance of 37 items. The domain 'conflicts of interest by industry' was excluded because of lack of consensus.

Conclusion. This is the first formal consensus approach towards structuring the considered judgement process in formulating recommendations in clinical guidelines. The final list of items can be used to facilitate the process of guideline development. The next step is to test the practical usefulness and applicability of this list in guideline development.

Keywords: clinical practice guideline, considered judgement, Delphi technique

In the last few decades, the number of published clinical practice guidelines has grown exponentially. An important aim of guidelines is to provide clinicians with information that helps them in making clinical decisions. Most professionals also use guidelines to maintain and improve the quality of health care services. To ensure high quality, the guidelines should be based on the best available scientific evidence. However, the evidence is often incomplete and controversial, so the transformation of evidence into recommendations is not straightforward [1,2]. The health benefits, side effects, and risks of different options for managing the disease or condition should also be considered. This step can be defined as 'considered judgement' [3,4], which is an extra dimension beyond the available evidence in formulating recommendations in guidelines. A valid tool for a systematic approach to this part of the guideline-development process is not available yet, in contrast to the numerous instruments that can be used for the critical appraisal and grading of scientific literature. In this article, we are introducing a systematic approach to identify items that should be included in the process of considered judgement to formulate recommendations in clinical guidelines.

Methods

Design and validity testing of the framework

This study was conducted from June 2003 until February 2004. On the basis of a literature review [5] and an existing list used in guideline development by the Dutch Institute for Healthcare Improvement CBO, we considered 49 items grouped into 10 potentially relevant domains: (i) clinical relevance, (ii) safety, (iii) patient's perspective, (iv) availability of resources, (v) health care costs, (vi) organization of care, (vii) professional's perspective, (viii) legal consequences, (ix) possible conflicts of interest by the industry, and (x) other 'considered judgements'.

A policy, Delphi procedure, was used to test the content and concurrent validity of this framework [6–11]. A panel of experts involved in evidence-based guideline development

Address reprint requests to Karin Verkerk, Department of Physiotherapy, Locatie Museumpark, Museumpark 40, Postbus 25035, 3001 HA Rotterdam, The Netherlands. E-mail: k.verkerk@hro.nl

was recruited from the Dutch network of guideline organizations [12]. Aims of this network are to improve the methodology of guideline development and to avoid duplication of efforts. We asked the representatives of the member organizations to participate in our study at the regular meeting of the network. As all representatives were experienced in guideline development, we consider them as experts.

In the first Delphi round, each panel expert was asked to independently rate the relevance of the items on a 4-point Likert scale (ranging from 'very important' to 'not important') and to comment on the individual domains and items (in a comment box adjacent to the items). The participants were also asked to comment on the list in general and to suggest new domains and/or items.

The panel responses were aggregated, tabulated, summarized, and returned to the participants in the second Delphi round. The domains and items were the same as in the first round. The participants rated the relevance of each item again and commented on the feedback. In addition, the participants were asked to decide whether to keep or remove the item from the list. The participants were reminded by electronic mail over a period of 3–8 weeks to send back their list (and questionnaire).

Analysis

The mean score and standard deviation (SD) were calculated for each item. Standardized domain scores were calculated by adding up the scores from all the experts and standardizing them as a percentage of the possible maximum score [10,13,14].

The final list was composed of items that at least 60% of the experts wanted to have included [6,7]. In contrast to the threshold of 70–75% that is recommended in the literature, we used a lower threshold of 60% because of the explorative nature of the study [6].

The overall agreement of the first and second Delphi rounds was calculated as well as the correlation between the decision to accept an item as 'relevant to be considered' for formulating recommendations and the item score (= equivalence reliability). The Spearson's product correlation coefficient (r_s) was used, with 0.8 as the limiting value to demonstrate a significant correlation [15,16].

Results

Participants

Twenty-eight experts from 16 different professional organizations in guideline development agreed to participate in the first Delphi round, and 21 of them participated in the second round. The lower response in the second round was partly because of lack of time ($n = 3$); four participants did not respond at all.

First Delphi round

The initial list generated significant agreement on the domains 'clinical relevance', 'safety', 'availability of resources', 'patient's perspective', and 'legal consequences'. Eight of the 12 items in the domain 'organization of care' had an average score lower than 3.0 (range 2.2–2.8). A few participants ($n = 10$) noted that these items are part of the implementation phase. Some of the participants ($n = 7$) felt that the domain 'possible conflicts of interest by the industry' could be a 'threat' in formulating a recommendation, but this had the lowest mean (2.0) of all 10 domains. Some participants ($n = 10$) had difficulties with certain items in the domain 'health care costs'.

Second Delphi round

The domains and their items and how they were scored by the Delphi panel are summarized in the Appendix. In response to the results of the two Delphi sessions, the number of domains was reduced to nine, and a new item 'addressing ethical issues' was included. Following comments from the first round, some items in the domains 'organization of care' and 'clinical relevance' were reformulated, and more explanation was added to the items described in the domain 'health care costs'. A comment from the participants was that a health technology assessment expert should be represented in each working group because of the limited knowledge of economic studies ('health technology assessment studies') and cost analysis.

The final checklist for 'considered judgement' contained 37 items grouped into nine domains. The items in the domains 'clinical relevance', 'safety', and 'availability of resources' had consensus scores ranging from 80 to 95%. The domain 'safety' had the highest score.

Six of the 12 items in the domain 'organization of care' were excluded because of a lack of consensus. A possible explanation for this, as several participants stated, is that these items belong in the implementation phase after the guideline was finalized.

Four of the nine items in the domain 'professional's perspective' had a consensus score lower than 60% and were excluded.

The impact of the legislation and rules on the professional and on the intervention when formulating a recommendation had a consensus score of 70–75%. 'The legal consequences when a guideline is not followed' was excluded (consensus 55%).

The domain 'possible conflicts of interest by industry' had a low consensus score (25%) and was therefore excluded.

The criteria of 'trustworthiness' that underwrites the reliability and validity of this research are all taken into account [7]. The acceptance of an item ('yes' option) was related to a high mean score on the 4-point Likert scale for that item (second Delphi round versus 'yes' $S_p = 0.74$, $P < 0.001$). The final checklist 'considered judgement' contained 37 items grouped into nine domains.

Discussion

We designed a framework in this study to structure the process of 'considered judgement' to formulate the recommendations into guideline development in the Netherlands. This was validated in two Delphi rounds by a panel of experts.

High scoring domains were 'clinical relevance', 'safety', and 'availability of resources'. The most commonly reported problem was the interpretation of several items in the initial list. After providing a more detailed (or improved) explanation of the items, the participants made only minor comments on the second Delphi list.

Using a threshold of 60%, we achieved consensus on 9 domains and 37 items. This threshold was set because of the explorative character of the study. A threshold of 70–75% is often recommended in the literature [6]. If this higher threshold had been used, there would be 29 items in the final checklist. Further testing of the usefulness and applicability of the items in the list is needed. This may further reduce the number of items.

Guidelines should be based on the integration of the best available evidence, clinical expertise, and patients' values and needs [17]. Almost all the participants in this study acknowledged the relevance of the concept of 'considered judgement' in guideline development, including patients' values and perspectives. Patient participation during guideline development may help to remind the guideline-working group not to focus exclusively on the scientific evidence [3–5]. Guidelines should also include information about possible side effects and the risks of the procedures and interventions that are recommended.

Organizational barriers and the impact on health care costs should also be considered in the development of guidelines [18,19]. In addition, effective guidelines include recommendations that are compatible with existing norms and values in daily practice and are feasible and applicable on a day-to-day basis. In practice, the need to acquire new knowledge or skills and changes in the organization and in existing routines may hamper the application of the guideline [14,20]. Using the list of 'considered judgements', these aspects can be systematically considered at all phases of the guideline-development process [3,18].

Owing to a lack of consensus, the domain 'possible conflicts of interest by industry' was excluded. It has been reported that contacts between physicians and the pharmaceutical industry are increasing [21]. Guidelines may also be influenced by industry through industrial funding of trials and by working group members having conflicts of interest [21–23]. Although the domain was not considered to be relevant by the expert panel, we still recommend being alert for conflicts of interests by all stakeholders involved in the guideline-development process.

The formulation of recommendations, in particular when the evidence is inconsistent, scarce, or lacking, is complex and value-loaded. The composition of guideline-development group as well as group dynamics can influence this process [2,24]. As a consequence, it may be difficult to describe arguments beyond the evidence in a transparent way. Attempts to design a uniform system for grading the strength of recommendations should balance the need for simplicity with the need for full and transparent consideration of all relevant issues [25,26].

Conclusions

This is the first formal consensus approach to explore other considerations apart from the scientific evidence relevant in formulating recommendations in clinical guidelines. The final checklist can be used during the process of guideline development and offers guideline developers the opportunity of systematically considering many additional items beyond the evidence when formulating recommendations. A systematic approach may also help in grading the strength of the final recommendations. If the list is used in parallel with the AGREE Instrument, it could enhance the quality and transparency of the guideline [18]. We encourage further testing of the validity and usefulness of this list for guideline developers in practice.

Acknowledgements

The authors thank the following people for participating in the Delphi rounds: W.J.J. Assendelft, T. Bekkering, J. Benraadt, N. Boluyt, I.J.M. Boonman-de Winter, H. den Bosch, P.J. Broek, Th. Daha, D. Daemers, A. Eliens, J.J.E. van Everdingen, L. Hakkaart-van Roijen, M. Heemskerk, H.J.M. Hendriks, M. Jansen, J. Leytens, R. van Peppen, E. Poot, C.J.G.M. Rosenbrand, P. Rossier, H.M.J. Slot, T. Spies, I. van der Stelt, J. van Vuuren, E. Walma, T. van der Weijden, T. Wiersma, and R. Wolters. All views expressed and all remaining errors are the full responsibility of the authors. No grants were received, and the authors had neither commercial nor financial interests to support this study.

References

1. Burgers JS, Van Everdingen JJE. Comment. Beyond the evidence in clinical guidelines. *Lancet* 2004; **364:** 392–393.
2. Raine R, Sanderson C, Hatchings A, Carter S, Larkin K, Black N. An experimental study of determinants of group judgements in clinical guideline development. *Lancet* 2004; **364:** 429–437.
3. Scottish Intercollegiate Guidelines Network. SIGN 50: A. Guideline Developers' Handbook: http://www.sign.ac.uk.
4. Richtlijnen binnen het kwaliteitsinstituut voor Gezondheidszorg CBO. *Handleiding voor de werkgroepen*. Utrecht: Kwaliteitsinstituut voor gezondheidszorg CBO, 2000.
5. Verkerk K. *Methodiek van evidence-based richtlijnen: een klassieke review*. Maastricht: Jaarwerkstuk Universiteit Maastricht, 2003.
6. van der Bruggen H. Delphi methoden. *Verpleegkunde* 2002; **17:** 45–57.
7. van der Bruggen H, Groen M. Patient outcome. Naar definiëring en classifisering van resultaten van verpleegkundige zorg. *Verpleegkunde* 1997; **12:** 68–81.
8. Verhagen AP, de Vet HCW, de Bie RA *et al*. The Delphi list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; **51:** 1235–1241.
9. Hesseling NM, Ketelaars CAJ, Halfens RJG, Borghouts JAJ. Rangordenen en wegen van kwaliteitscriteria. *Verpleegkunde* 1996; **11:** 167–174.
10. Snyder-Halpern R, Bagley Thompson C, Schaffer J. Comparison of mailed vs. internet applications of the Delphi technique in clinical informatics research: www.bmj.com.

11. Jones J, Hunter D. Qualitative research: consensus methods for medical and health services research. *Br Med J* 1995; **311:** 376–380.

12. Burgers JS, Van Everdingen JJE. Evidence-based richtlijnontwikkeling: het EBRO-platform. *Ned Tijdschr Geneesked* 2004; **148:** 2057–2059.

13. Imbos Tj, Jansen MPE, Berger MPF. *Methodologie en Statistiek*. Maastricht: Faculteit der Gezondheidswetenschappen, Universiteit Maastricht, 1996.

14. Burgers JS. *Quality of Clinical Practical Guidelines*. Nijmegen: Proefschrift UMC St. Radboud, 2002.

15. Bouter LM, Dongen van MCJM. *Epidemiologisch Onderzoek: Opzet en Interpretatie*. Houten: Bohn Stafleu Van Loghum, 1995.

16. Altman DG. *Practical Statistics for Medical Research*. Lonen: Chapman & Hall, 1985.

17. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based Medicine. How to Practice and Teach EBM*, second edition. Edinburgh: Churchill Livingstone, 2000.

18. The AGREE Collaboration. 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.

19. Offringa M, Assendelft WJJ, Scholten RJPM. Inleiding in evidence-based medicine. *Klinisch handelen gebaseerd op bewijsmateriaal*. Houten/Diegem: Bohn Stafleu Van Loghum, 2000.

20. Burgers JS, Grol RPTM, Zaat JOM, Spies TH, van der Bij AK, Mokkink HGA. Characteristics of effective clinical guidelines for general practice. *Br J Gen Prac* 2003; **53:** 15–19.

21. Choudhry N, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* 2002; **287:** 612–617.

22. Taylor R, Giles J. Cash interests taint drug advice. *Nature* 2005; **437:** 1070–1071.

23. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337:** 867–872.

24. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. *J Eval Clin Pract* 2001; **7:** 165–173.

25. Atkins D, Best D, Briss PA *et al.* for the GRADE Working Group. Grading quality of evidence and strength of recommendations. *Br Med J* 2004; **328:** 1490.

26. Guyatt G, Gutterman D, Baumann MH *et al.* Grading strength of recommendations and quality of evidence in clinical guidelines: report from an american college of chest physicians task force. *Chest* 2006; **129:** 174–181.

Accepted for publication 5 August 2006

Appendix

Table A1 Results of the second Delphi session examining the different domains, each with several items

		n	M (SD)	Consensus
Clinical relevance				
1	The sizes of patient populations to which the intervention will apply	20	2.4 (0.9)	85%
2	The magnitude of the effectiveness of an intervention compared with no intervention	20	3.6 (0.5)	85%
3	The consistency of results across different studies	20	3.4 (1.1)	80%
4	The benefit of the intervention compared with another intervention or other interventions	20	3.5 (1.0)	90%
5	The generalizability, taking into account the demographic characteristics of the study population	19	3.8 (0.4)	95%
Domain score 77.25%				
Safety				
6	Side effects	19	3.6 (0.7)	90%
7	Short term risks or complications	19	3.7 (0.6)	85%
8	Long term risks or complications	19	3.6 (1.0)	90%
Domain score 80.30%				
Patient perspective				
9	The needs and expectations of the patients	20	3.1 (1.0)	79%
10	Ability to decide between different interventions (autonomy)	20	2.8 (1.2)	84%
11	The (expected) compliance	20	3.4 (0.5)	95%
12	The patient has influence on whether to follow the guideline or not	18	1.9 (0.8)	47% Exit
13	The expected satisfaction about the outcome of the intervention	20	2.8 (0.8)	84%
14	Specific personal entities concerning the use of the intervention	20	2.6 (1.1)	75%
15	The accessibility to health care or intervention	18	3.3 (0.8)	84%
16	Being informed about the benefit and harm of an intervention (safety)	19	3.5 (1.0)	74%
17	Legislation and regulation that applies to the patient	19	3.2 (1.0)	63%
Domain score 64.46%				

continued

Table A1 *continued*

		n	M (SD)	Consensus
Availability of resources				
18	The resources are available in the Netherlands for the organization, the professional and the patient	19	3.5 (1.0)	95%
19	The (required) experience and competence of professionals	19	3.4 (1.0)	95%
Domain score 80.70%				
Health care costs				
20	Cost consequence analysis	19	2.6 (0.9)	79%
21	Cost effectiveness (efficiency)	19	2.3 (2.1)	90%
22	Cost utility analysis	19	2.5 (1.8)	84%
23	Cost benefit analysis	19	2.7 (0.9)	79%
24	Cost minimization analysis	18	3.1 (0.8)	83%
Domain score 55.50%				
Organization of care				
25	The way in which the organization of care has to be offered	19	2.9 (0.8)	80%
26	Differences in the fee between health care insurers	19	1.6 (0.8)	20% Exit
27	The role of the organization and management	20	2.0 (0.9)	35% Exit
28	Multidisciplinary approach to implementing the intervention	20	2.7 (1.1)	75%
29	Required education of the staff or personnel	20	3.0 (1.2)	75%
30	The attitude of the group for which the guidelines are intended	20	2.9 (1.0)	65%
31	Aims, mission and priorities of an organization implementing an intervention	20	2.4 (1.0	40% Exit
32	The magnitude of change in the organization or health care process	20	3.0 (1.1)	75%
33	The organization culture and the willingness to change	20	2.4 (1.0)	55% Exit
34	Supply and demand-driven care	18	2.1 (1.0)	29% Exit
35	The infrastructure for implementation	20	3.0 (1.0)	80%
36	Consequences of politics and strategy	20	2.1 (1.0)	47% Exit
Domain score 49.42%				
Professionals' perspective				
37	Clinical autonomy	20	2.4 (0.9)	53% Exit
38	Access to financial and information sources	16	2.4 (0.8)	50% Exit
39	Positive or negative financial consequences for the professional	20	2.3 (0.9)	65%
40	The willingness to acquire new knowledge and skills	20	2.3 (1.0)	55% Exit
41	The standards and values of a professional	20	2.4 (1.2)	60%
42	The risk imposed on the professional when applying the intervention	19	3.5 (0.8)	90%
43	Loss or gain of time by applying the intervention	18	3.1 (0.8)	83%
44	Co-management of the availability of resources	19	2.2 (0.9)	58% Exit
45	The organization culture of the (different) professional(s)	20	2.5 (0.9)	65%
Domain score 52.00%				
Legal consequences				
46	Legal consequences when complying with a guideline or not	19	2.6 (1.0)	55% Exit
47	Legislation and regulations that apply to the professional	19	3.1 (0.7)	75%
48	Specific legislation and regulations	19	3.2 (0.7)	70%
Domain score 65.00%				
Possible conflicts of interest by the industry				
49	Commercial value by applying an intervention	19	1.6 (0.9)	25% Exit
Domain score 19.30%				
Other considered judgments				
50	Ethics considerations	17	3.2 (0.8)	Yes
Domain score 72.50%				

The mean (M), standard deviation (SD), and domain score are given for each item.