

Executive summary

Consensus development methods, and their use in clinical guideline development

MK Murphy¹

NA Black¹

DL Lamping¹

CM McKee¹

CFB Sanderson¹

J Askham²

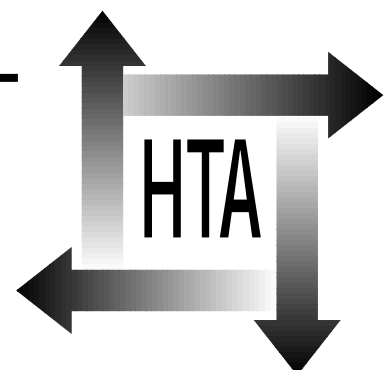
T Marteau³

¹ Health Services Research Unit, London School of Hygiene & Tropical Medicine

² King's College, London

³ United Medical & Dental Schools, London

Health Technology Assessment
NHS R&D HTA Programme





Executive summary

Background

Consensus methods are increasingly being used to develop clinical guidelines which define key aspects of the quality of health care, particularly appropriate indications for interventions. This review is restricted to formal consensus methods in which the structure, process and output are explicit from the outset. Three main approaches have been used in the health field: the Delphi method, the nominal group technique (NGT) and the consensus development conference.

Objectives

- To identify the factors that affect the decisions that emerge from consensus development methods.
- To assess the implications of the findings for the development of clinical guidelines.
- To recommend further methodological research for improving the use of consensus development methods as a basis for guideline production.

Methods

Data sources

The majority of the literature reviewed was identified through searches of Medline, PsychLIT and the Social Science Citation Index and from reference lists in retrieved articles.

Study selection

A matrix of 15 cells was developed from three types of activity (planning, individual judgement, group interaction) and five components (questions, participants, information, method of structuring the interaction, method of synthesising individual judgements) involved in consensus development methods.

Six cells were selected for detailed review on the basis of three criteria: (1) importance to consensus decision-making in the health sector; (2) the amount and quality of the literature available; (3) the potential for offering practical guidance. For each of the six cells the review drew on the results of the principal general search. For some cells, further focused searches were undertaken.

In all, 177 primary research and review articles were selected.

Data extraction and synthesis

If substantial literature was available from the health sector, we paid little or no attention to evidence from other sectors. If few or no studies had been conducted in the health sector, we sought relevant evidence from other fields. We used a narrative approach, sometimes based around tables of results. The extent to which research support exists for any conclusion is indicated, although these should not necessarily be considered as a hierarchy: A = clear research evidence; B = limited supporting research evidence; C = experienced common-sense judgement.

Results and conclusions

Setting the task or question to be addressed

- Cues included in scenarios must be selected with care. As well as reviewing the relevant literature, clinicians in the consensus group should give their opinions (most usefully in the first round) about which cues are important. Doing so may help maintain their participation and help them justify their judgements. [C]
- Contextual cues included in scenarios are as important as ones specific to the topic at issue, and they should be made explicit. [B]
- It must be decided whether to focus on ways of managing a specific condition or on indications for using an intervention. If the focus is on an intervention, care should be taken about how to deal with other relevant interventions. [C]
- Is a global judgement elicited, or is an attempt made to break the judgement down into probability and utility estimates? Although there are theoretical advantages to the latter, it is likely to be a more difficult task for participants and it may not enhance judgements. [C]
- Inclusion of all possible scenarios may increase comprehensiveness, but if many of the scenarios never occur in practice, the increased burden on the respondents may not be justified by the limited value of the information provided. Judgements of scenarios which never or rarely occur in practice may be less reliable. [B]

- Requiring participants to judge what may be seen as numerous irrelevant scenarios may alienate them from the task. [C]

Selecting the participants

- Within defined specialist or professional categories, the selection of the particular individuals is likely to have little impact on the decision of a group of sufficient size. To enhance the credibility and widespread acceptance of the guidelines, the participants should reflect the full range of key characteristics of the population that it is intended to influence. Selection should be seen to be unbiased. [C]
- To define common ground and maximise areas of agreement, groups should be homogeneous; to identify and explore areas of uncertainty, a heterogeneous group is appropriate. [B]
- In judgements of clinical appropriateness, the most influential background factor is the particular medical specialty. Specialists tend to favour the interventions with which they are most familiar. Consensus-based guidelines should therefore be interpreted in the context of the specialty composition of the group. [A]

Choosing and preparing the scientific evidence

- A review of research-based information should be provided to all participants at an early stage. Participants should be encouraged to bring the review and any personal notes to the group sessions as memory aids. [B]
- Information presented in a synthesised form (e.g. tables) is more likely to be assimilated. Participants may be more likely to use information that is presented in an accessible format. Information tabulated so as to increase the salience of the dimensions to be used for making judgements is more likely to be processed in this manner. [C]
- Methodologists should be involved in conducting any literature review. [C]
- Grading the quality of studies using a reliable method may mitigate the biases of the reviewers somewhat, but may not eliminate them. [B]

Structuring the interaction

- With NGTs and the Delphi method, two or more rating rounds are likely to result in some convergence of individual judgements, though it is unclear whether this increases the accuracy of the group decision. [A]
- With the Delphi method, it is advisable to feed back reasons or arguments as well as measures of central tendency or dispersion. [B]
- Efforts should be made to mitigate the effects of status of participants (which can affect their con-

tribution to and influence within a group). [B]

- A comfortable environment for meetings is likely to be preferred by participants and to be conducive to discussion. [C]
- A good facilitator will enhance consensus development and can ensure that the procedure is conducted properly. [C]

Methods of synthesising individual judgements

- An implicit approach to aggregating individual judgements may be adequate for establishing broad policy guidelines. More explicit methods based on quantitative analysis are needed to develop detailed, specific guidelines. [C]
- The more demanding the definition of agreement, the more anodyne the results will be. If the requirement is too demanding, either no statements will qualify or those that do will be of little interest. [C]
- Differential weighting of individual participants' views produces unreliable results unless there is a clear empirical basis for calculating the weights. [B]
- The exclusion of individuals with extreme views (outliers) can have a marked effect on the content of guidelines. [A]
- There is no agreement as to the best method of mathematical aggregation. [B]
- Reports of consensus development exercises should include an indication of the distribution or dispersal of participants' judgements, not just the measure of central tendency. In general, the median and the inter-quartile range are more robust than the mean and standard deviation. [A]

Priorities for future research

- What impact does the framing or presentation of the question have on individual judgements?
- In what form and how inclusive should scenarios be?
- How does the extent of heterogeneity of a group affect the process and outcome?
- What effect does research-based information have on individual and on group judgements? Does the effect depend on the amount of information or how it is presented?
- What effect does the method of feedback of participants' views have on group judgement?

Publication

Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al.* Consensus development methods, and their use in clinical guideline development. *Health Technol Assessment* 1998; 2(3).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Methodology Panel.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will, in England, be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

Series Editors: Andrew Stevens, Ruairidh Milne and Ken Stein
Assistant Editor: Jane Robertson

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
<http://www.soton.ac.uk/~hta>

ISSN 1366-5278