

Expert consensus on the desirable characteristics of review criteria for improvement of health care quality

H M Hearnshaw, R M Harker, F M Cheater, R H Baker, G M Grimshaw

Abstract

Objectives—To identify the desirable characteristics of review criteria for quality improvement and to determine how they should be selected.

Background—Review criteria are the elements against which quality of care is assessed in quality improvement. Use of inappropriate criteria may impair the effectiveness of quality improvement activities and resources may be wasted in activities that fail to facilitate improved care.

Methods—A two round modified Delphi process was used to generate consensus amongst an international panel of 38 experts. A list of 40 characteristics of review criteria, identified from literature searches, was distributed to the experts who were asked to rate the importance and feasibility of each characteristic. Comments and suggestions for characteristics not included in the list were also invited.

Results—The Delphi process refined a comprehensive literature based list of 40 desirable characteristics of review criteria into a more precise list of 26 items. The expert consensus view is that review criteria should be developed through a well documented process involving consideration of valid research evidence, possibly combined with expert opinion, prioritisation according to health outcomes and strength of evidence, and pilot testing. Review criteria should also be accompanied by full clear information on how they might be used and how data might be collected and interpreted.

Conclusion—The desirable characteristics for review criteria have been identified and will be of use in the development, evaluation, and selection of review criteria, thus improving the cost effectiveness of quality improvement activities in healthcare settings.

(*Quality in Health Care* 2001;10:173–178)

Keywords: review criteria; Delphi process; audit criteria; quality improvement

The issue of defining and assessing the quality of health care is central to improving clinical practice. In recent years quality improvement methods such as clinical audit and clinical utilisation review have been actively promoted by many healthcare providers and policy makers.^{1–3}

Key messages

- Review criteria are the elements against which quality of health care is assessed in quality improvement.
- An expert consensus view was generated of 26 desirable characteristics of review criteria.
- Review criteria should be developed through a well documented process involving consideration of valid research evidence, possibly combined with expert opinion, prioritisation according to health outcomes and strength of evidence, and pilot testing.
- Review criteria should be accompanied by full and clear information on how they might be used and how data might be collected and interpreted.
- Information on the characteristics of review criteria will enable rational selection of review criteria for quality improvement activities.

What is already known on the subject

Review criteria are systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services and outcome. There is no generally accepted method of defining appropriate review criteria. If quality of care is assessed against inappropriate criteria, resulting improvements in performance against these criteria may not effect any improvement in care, and resources may be wasted in ineffective quality improvement activities.

What this paper adds

An expert consensus on the desirable characteristics of review criteria has been generated and the criteria of quality for quality review criteria have been identified. Review criteria can now be effectively identified and presented so that data can be collected on justifiable, appropriate, and valid aspects of care.

The first stage of quality improvement for a given topic is to establish the review criteria to use. Review criteria are “systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services and outcomes”.⁴ If appropriate review criteria are used, improvements in performance measured against these criteria should result in improved care. In contrast, if

Centre for Primary Health Care Studies, University of Warwick, Coventry CV4 7AL, UK

H M Hearnshaw, senior lecturer in primary care

National Children's Bureau, London, UK
R M Harker, senior research officer

School of Healthcare Studies, University of Leeds, Leeds, UK
F M Cheater, professor of public health nursing

Clinical Governance Research and Development Unit, Department of General Practice and Primary Health Care, University of Leicester, Leicester, UK
R H Baker, professor of quality in health care

Centre for Health Services Studies, University of Warwick, Coventry, UK
G M Grimshaw, senior research fellow

Correspondence to:
Dr H Hearnshaw
hilary.hearnshaw@warwick.ac.uk

Accepted 21 June 2001

quality of care is assessed against inappropriate or irrelevant criteria, then resulting improvements in performance against these criteria may not effect any improvement in care, and resources may be wasted in ineffective quality improvement activities.⁵

There is no generally accepted method of defining appropriate review criteria. A few authors have proposed what the desirable characteristics of review criteria are,^{4 6} but there is no clear indication of how appropriate criteria might be developed. Often review criteria have been generated from guidelines.⁷⁻¹⁰ Alternatively, instead of directly translating guidelines into review criteria, it has been argued that criteria should be based directly on high quality research evidence and prioritised according to strength of evidence and impact on health outcomes.^{11 12}

Unfortunately, high quality research evidence is not readily available for all clinical topics,¹³ and expert opinion is often relied upon to develop criteria.^{14 15} Although some authors recommend that criteria should not be developed at all if research evidence is lacking,¹¹ an alternative approach is to synthesise the two methods, drawing on expert opinion when there is no research evidence.^{16 17}

Consensus methods are increasingly used to develop clinical guidelines^{18 19} and can provide a way of determining how review criteria should be selected and defining their desirable characteristics. The Delphi technique²⁰ is a consensus method that gathers expert opinion through an iterative questionnaire process. The researchers communicate in writing with a panel of experts comprising between 10 and 50 members. Experts are anonymous to the extent that other panel members do not know their identity at the time of data collection.

It is recommended that the panel should include both "advocates" and "referees".²¹ The expertise of advocates stems from participant involvement in the area under study—for example, clinicians or quality managers. Referees have less direct involvement and their expertise is derived from study of the topic—for example, academic researchers. Hence, we shall refer to advocates as practitioner experts and referees as academic experts.

Modifications to a "pure" Delphi process are common.²²⁻²⁴ The preparatory stage of formulating issues can be supplanted by reference to existing research²⁵ and subsequent rounds can be used to develop, rather than directly reiterate, the concerns of previous rounds.²⁰ This study aimed to determine both how review criteria should be selected and their desirable characteristics, using a modified Delphi technique. This information will inform both those who develop review criteria and those who select review criteria to make quality improvement in health care more effective.

Methods

A two round modified Delphi process was used to generate consensus amongst an international panel of experts. A decision was made to restrict the Delphi process to two rounds before inviting experts to participate, since the

initial questionnaire was based upon a careful review of available literature. It was considered that two rounds would be enough to reach adequate consensus and would minimise the workload for participants.

THE EXPERT PANEL

We identified an international group of experts in quality improvement in health care from a variety of professional disciplines. Three sources of information were used to identify experts: (1) publication record, (2) membership of quality improvement groups in relevant organisations such as the Royal Colleges in the UK, and (3) recommendations from researchers in the field. Forty nine experts were contacted, mostly by email, and asked to contribute to the study. The expert group was categorised by the researchers into 26 academic experts ("referees") and 23 practitioner experts ("advocates"). Individuals for whom contact details were rapidly obtained were contacted first. Having received a sufficient number of positive responses from these experts, we ceased to seek contact details for further individuals. Although we acknowledge that our list of experts is not exhaustive and other individuals could have contributed to the study, the expert group provided a wide range of views and adequate representativeness.²⁰

ROUND 1

Medline and Embase databases were searched from 1990 to March 1999 using the topic headings "clinical audit", "medical audit", "clinical utilization", "quality assurance", and "guidelines" and text words "review criteria", "appropriateness criteria", "clinical indicators", and "performance indicators". The abstract of each citation was reviewed and all studies concerned with the development of review criteria were retrieved. In addition, publications of expert panel members on the development of clinical guidelines were reviewed. The literature review was used to compile a list of identified desirable characteristics of review criteria from which the questionnaire for round 1 of the Delphi process was constructed. The questionnaire contained 40 items in three sections:

- (1) The process of developing review criteria (18 items).
- (2) Attributes of review criteria (11 items).
- (3) The usability of review criteria (11 items).

The experts were asked to rate importance and feasibility for each item using 7 point scales, anchored by "not at all important" and "very important", and "not at all feasible" and "very feasible". Free comments on each item and suggestions about items overlooked in the questionnaire were also invited. Questionnaires were distributed by email to 31 experts and by post to seven experts who did not have access to email. Experts were asked to complete the questionnaire within 2 weeks. Non-responders were sent reminders after 2 weeks and, where necessary, after a further 10 days. Round 1 was concluded 5 weeks after the distribution of the questionnaire.

Table 1 Experts involved in each stage of the Delphi process

	Academics	Practitioners	Total
Invited to participate	26	23	49
Agreed to participate	21	17	38
Round 1 returns			
Without reminder	10	7	17
After 1 reminder	5	3	8
After 2 reminders	3	2	5
Completed round 1	18	12	30
Round 2 returns			
Without reminder	7	6	13
After 1 reminder	2	2	4
After 2 reminders	6	3	7
Completed round 2	15	10	25

Table 2 Items included/excluded after each round of the Delphi process

	No of items	Inclusion	Exclusion
Round 1			
Section 1: Development	18	17	1
Section 2: Attributes	11	7	4
Section 3: Usability	11	5	6
Total	40	29	11
Round 2			
Section 1: Development	19 (2 new)	14	5
Section 2: Attributes	7	7	0
Section 3: Usability	5	5	0
Total	31	26	5

Round 1 responses were aggregated and used to identify aspects to retain for round 2. A definition of disagreement based upon the RAND/UCLA appropriateness method¹⁶ was generated by the project team and used to exclude items from round 2 if their ratings were polarised to the extreme points of the scale—that is, if three or more experts gave a high rating of 6 or 7 whilst, in addition, three or more gave a low rating of 1 or 2. Cumulative percentage scores were then used to determine which of the remaining items met the inclusion criteria, firstly, of at least 80% of the expert panel providing an importance rating of 5 or more and, secondly, a feasibility rating of 4 or more. Items thus excluded showed lack of consensus on being given a rating for importance on the top two points on the scale. Nevertheless, there was opportunity later for experts to request the exclusions to be revoked. Where experts provided comments, these were carefully considered in project team discussions. Some comments resulted in a refinement of item wording for round 2 of the Delphi process, others led to the inclusion of additional items where experts felt significant omissions arose. The aggregation of round 1 results and subsequent development of round 2 occurred over a 2 week period. The round 2 questionnaire was ready 7 weeks after the round 1 questionnaire was sent out.

ROUND 2

The round 2 questionnaire informed the experts of the method used to identify items to be included or excluded for round 2. Experts were asked to re-rate each item for round 2 and to provide additional comments if they wished. The questionnaire reminded experts of their own round 1 rating for each item and presented the expert group’s mean rating for that item. If the wording of items had been altered, ratings for the original item were provided and the initial wording was shown below the altered item.

Some new items were added to section 1 in response to expert comment. These were clearly labelled as being new. All excluded items were shown separately at the end of each section. Experts could alter their ratings for these items and comment on their exclusion, if they wished.

The same processes for distribution, reminding, and analysis were used in round 2 as in round 1. Items retained after round 2 identified the desirable characteristics of review criteria and the method of selecting them.

Members of the panel of experts were sent details of the outcomes of the Delphi process.

Results

Thirty eight of the 49 experts invited to take part agreed to do so. The number of experts responding to each round of the Delphi is shown in table 1. The table also gives details of the number of practitioner and academic experts included in each round. There were no significant differences in the proportion of practitioners and academics responding to the initial participation request ($\chi^2 = 0.3, p>0.05$), nor did experts’ status as a practitioner or academic relate to their likelihood of completing round 1 ($\chi^2 = 1.5, df = 1, p>0.05$) or round 2 ($\chi^2 = 0.5, df = 1, p>0.05$). Participating experts are listed in appendix 1.

From a starting point of 40 items in round 1, 26 items qualified for inclusion after two rounds of the Delphi process—that is, 80% of the experts gave importance ratings of 5 or more and feasibility ratings of 4 or more for these 26 items and there was no polarisation of expert ratings. Table 2 shows the number of items in each round resulting in exclusion and inclusion.

In the final list of desirable characteristics, 13 items retained the original round 1 wording, 12 items were reworded for round 2, and one item was introduced in round 2. Table 3 shows the final list and the mean importance and feasibility rating associated with each characteristic.

The round 2 questionnaire also allowed experts to reconsider the 11 items excluded at round 1. The round 2 responses confirmed that all these items should be excluded. In addition, a further five items were excluded after round 2 (including one of the items newly introduced at round 2). All the 16 excluded items are shown in table 4 with the reasons for exclusion.

The final lists of included and excluded items (tables 3 and 4) were given to all expert participants for comment. There was no dissent to the list.

Discussion

It has been possible to define the desirable characteristics of review criteria by use of a modified Delphi process. This method has refined and validated the list of characteristics initially based on literature alone. The use of expert judgement has identified which of the literature based characteristics of review criteria are highly important and feasible. The final list of desirable characteristics can inform developers and users of review criteria and lead

Table 3 List of the desirable characteristics of review criteria and mean importance and feasibility ratings, ordered by importance rating

Characteristic	Mean rating	
	Importance	Feasibility
Criteria are described in unambiguous terms	6.7	5.6
Criteria are based on a systematic review of research evidence	6.6	5.0
The validity of identified research is rigorously appraised	6.5	5.5
Criteria include clear definitions of the variables to be measured	6.5	5.9
Criteria explicitly state the patient populations to which they apply	6.4	5.9
Criteria are capable of differentiating between appropriate and inappropriate care	6.3	5.0
Criteria are linked to improving health outcomes for the care being reviewed	6.3	4.7
Criteria explicitly state the clinical settings to which they apply	6.2	5.8
The collection of information required for criteria based review minimises demands on staff	6.2	6.4
The method of selecting criteria is described in enough detail to be repeated	6.1	5.8
Criteria are accompanied by clear instructions for their use in reviewing care	6.1	6.0
The systematic review used to guide the selection of criteria is up to date	6.0	5.4
Criteria are pilot tested for practical feasibility	6.0	5.1
Criteria include aspects of care that are relevant to patients	6.0	4.9
The collection of information for criteria based review is acceptable to those patients whose care is being reviewed	6.0	5.2
The bibliographic sources used to identify research evidence are specified	5.9	6.3
In selecting criteria, decisions on trade-offs between outcomes from different treatment options are stated	5.9	4.5
The collection of information required for criteria based review minimises demands on patients	5.9	5.3
The method of synthesising evidence and expert opinion is made explicit	5.8	5.3
Criteria are prioritised according to the quality of supporting evidence	5.8	4.9
Criteria are prioritised according to their impact on health outcomes	5.8	4.5
The criteria used to assess the validity of research are stated	5.6	5.8
Similar criteria should emerge if other groups review the same evidence	5.5	4.6
The collection of information for criteria based review is acceptable to those staff whose care is being reviewed	5.5	4.8
Expert opinion is included in the process of developing review criteria	5.3	5.5
Criteria used in previous quality reviews of the same clinical topic are considered for inclusion	5.3	5.8

Table 4 Excluded items not included in the final list of desirable characteristics

Characteristic	Reason for exclusion*		
	Low importance	Low feasibility	Polarisation
<i>Development of review criteria</i>			
The views of patients are included in the process of developing review criteria	Y	Y	
Criteria are prioritised according to the cost implications (of complying with the criteria for equal health outcomes)	Y	Y	
The search strategy or keywords used to search the literature are stated	Y		
The experts involved in the process of developing review criteria are stated	Y		
Criteria are prioritised	Y		
Criteria are prioritised according to their importance to patients		Y	
<i>Attributes of review criteria</i>			
Criteria are linked to lowering resources used in the care being reviewed	Y	Y	
Criteria are presented in computer compatible format, where possible	Y	Y	
All criteria that are necessary for a review of the topic of care are included	Y	Y	
The same results should emerge if other people collect data for the criteria		Y	
<i>Usability of review criteria</i>			
Criteria facilitate data collection over a reasonable time period which may include training in data collection	Y	Y	Y
Criteria facilitate data collection from acceptable sample sizes	Y		Y
The average time range required to collect information for criteria based review is indicated	Y		
The individuals conducting the review are confident that the criteria are valid		Y	
The criteria produce ratings that are easy to interpret into appropriateness of care		Y	
The degree of professional expertise required to collect information for criteria based review is indicated	Y		

*Low importance means that less than 80% of the experts gave importance ratings of 5 or more on the 7 point scale. Low feasibility means that less than 80% of the experts gave feasibility ratings of 4 or more. Polarisation means that more than two experts rated this item at 6 or 7 and more than two rated it at 2 or less on one of the scales.¹⁶

to more appropriate, less wasteful, use of review criteria in quality improvement activities in health care.

Our original list of important aspects of review criteria consisted mainly of items mentioned in publications by expert panel members. While the Delphi process confirmed the importance of most of these items, it also excluded some. For instance, although the process retained items such as "Criteria are based on a systematic review of research evidence" and "Expert opinion is included in the process of developing review criteria", it excluded specifying the search strategy used or the names of the experts involved. We feel this has demonstrated the effectiveness of the Delphi process in excluding unimportant or extreme views to arrive at a more centralised accepted definition.

The literature based initial list of criteria included four items specifically related to patient issues. The items "Criteria include aspects of care that are relevant to patients"

and "The collection of information for criteria based review is acceptable to those patients whose care is being reviewed" were retained in the final list. However, the item "The views of patients are included in the process of developing review criteria" was excluded as having low feasibility. Measuring the importance to patients of criteria could certainly be very costly in time and resources, making it infeasible, even though it was rated as important. The item "Criteria are prioritised according to their importance to patients" was rated as of low feasibility and low importance. The prioritisation of criteria was generally rated as of low importance, perhaps indicating that all criteria which met the other characteristics should be used, and prioritisation added nothing of importance. Thus, it was not specifically the patients' priorities that were being excluded, but prioritisation itself.

The external validity of the collective opinion produced by a Delphi method is dependent

on the composition of the expert panel. This study aimed to include both referees (academic researchers) and advocates (quality improvement practitioners) to balance insight from theoretical understanding with practical experience. However, the eventual composition of our expert panel was marginally biased towards an academic research perspective as slightly more “referees” than “advocates” agreed to participate. This may have caused practical aspects of review criteria to be under-represented in the final definition, and could explain the exclusion of items reflecting resource allocation and patient considerations. Given the increase in the political importance of resource allocation and patient considerations, these exclusions could be deemed inappropriate.

The identification of desirable characteristics of review criteria has emerged from the expert consensus process used in this study. Although the problems associated with the nature of consensus and the structure of our expert panel should be acknowledged, the definition created here represents a considerable advance in our understanding of what appropriate review criteria are and how they might be developed. Previous published literature alone did not directly translate into a definitive list of the desirable characteristics of review criteria. Inconsistency was found in the literature on the relative importance which individual researchers assigned to different aspects of review criteria. The use of this international panel of experts from a variety of professional disciplines lends a universal relevance to this definitive list. It is not dependent on the views of professionals from one specific context or working within one particular healthcare system.

The modified Delphi process used here was very efficient in terms of time and resources spent in consultation with the experts involved. The process allowed experts from geographically disparate areas to be included at relatively low cost. Distributing questionnaires by email further minimised the resources needed and the time taken to complete each round of the Delphi process. The method proved a fast and low cost way of gathering the opinions of an international panel.

The knowledge gained from this study should be of relevance to all those involved in the development of review criteria for quality improvement. We have provided information on how review criteria can be effectively identified and presented so that data can be collected on justifiable, appropriate, and valid aspects of care. This can also be used to guide the assessment of review criteria for those who need to decide which criteria are appropriate for their particular quality assurance or quality improvement project. In short, the criteria of quality for quality review criteria have been identified. Future studies should be directed at establishing the value of this set of meta-criteria as a developmental tool to aid the selection of review criteria for quality improvement activities. The set is currently being used to develop an instrument to appraise the quality of review

criteria. This instrument will have the potential to raise the standard of all quality reviews and thus improve the quality of health care.

In summary, the expert consensus view is that review criteria should be developed through a well documented process involving consideration of valid research evidence, possibly combined with expert opinion, prioritisation according to health outcomes and strength of evidence, and pilot testing. Review criteria should also be accompanied by full clear information on how they might be used and how data might be collected and interpreted.

Appendix 1: List of expert participants

Andrew Booth, University of Sheffield, UK
 John Cape, British Psychological Society, UK
 Alison Cooper, Fosse Healthcare NHS Trust, UK
 Gregor Coster, Auckland University, New Zealand
 Susan Dovey, University of Otago, New Zealand
 Jeremy Grimshaw, Aberdeen University, UK
 Gordon Guyatt, McMaster University, Canada
 Gill Harvey, Royal College of Nursing, UK
 Nick Hicks, Oxford Public Health, UK
 Jaki Hunt, Kettering General Hospital, UK
 Lyn Juby, Clinical Governance Research and Development Unit, UK
 Kamlesh Khunti, Clinical Governance Research and Development Unit, UK
 Beat Kuenzi, SGAM Research Group, Switzerland
 Mayur Lakhani, Clinical Governance Research and Development Unit, UK
 Philip Leech, National Health Service Executive, UK
 Katherine Lohr, Research Triangle Institute, USA
 Adrian Manhire, Royal College of Radiologists, UK
 Karen Mills, Warwickshire Multi-disciplinary Audit Advisory Group, UK
 Andrew Moore, Bandolier, UK
 Mary Ann O'Brien, McMaster University, Canada
 Frede Oleson, Aarhus University, Denmark
 Barnaby Reeves, Royal College of Surgeons, UK
 James Rimmer, Avon Primary Care Audit Group, UK
 Tom Robinson, Leicester General Hospital, UK
 Martin Roland, National Primary Care Research & Development Centre, UK
 Charles Shaw, CASPE Research, UK
 Paul Shekelle, RAND, USA
 Chris Silagy, Flinders University, Australia
 Tim van Zwanenburg, University of Newcastle, UK
 Kieran Walshe, Birmingham University, UK
 Geoff Woodward, Royal College of Optometrists, UK

The researchers are indebted to all the members of the expert panel who contributed their time and expertise so willingly. Acknowledgement is also given to Dr Mayur Lakhani, University of Leicester, UK, for his helpful comments on an earlier version of this paper, and to the anonymous referees of an earlier version.

This research was funded by the UK National Health Service Research and Development Health Technology Assessment programme.

The views and opinions expressed here are those of the authors and do not necessarily reflect those of the NHS Executive.

- 1 Department of Health. *Working for patients*. London: HMSO, 1989.
- 2 Department of Health. *Working paper 6: medical audit*. London: HMSO, 1989.
- 3 Department of Health and Human Services. *The challenge and potential for assuring quality health care for the 21st century*. Publication No. OM 98-0009. Prepared for the Domestic Policy Council, Washington DC, 17 June 1998. <http://www.ahcpr.gov/qual/21stscena.htm>.
- 4 Field MJ, Lohr KN, eds. *Guidelines for clinical practice: from development to use*. Washington: Institute of Medicine, 1992.
- 5 Naylor CD, Guyatt GH. Users' guides to the medical literature XI. How to use an article about a clinical utilization review. *JAMA* 1996; 275: 1435-9.
- 6 Lohr KN, ed. *Medicare: a strategy for quality assurance. Volume I*. Washington: National Academy, 1990.
- 7 Bozzett SA, Asch S. Developing quality review criteria from standards of care for HIV disease: a framework. *J Acquir Immune Defic Syndr Hum Retroviral* 1995; 8: S45-52.

- 8 Lanska D. A public/private partnership in the quest for quality: development of cerebrovascular disease practice guidelines and review criteria. *Am J Med Qual* 1995; 10: 100–6.
- 9 James PA, Cowan TM, Graham RP, et al. Using a clinical practice guideline to measure physician practice: translating a guideline for the management of heart failure. *J Am Board Fam Pract* 1997; 10: 206–12.
- 10 Hadorn DC, Baker RW, Kamberg CJ, et al. Phase II of the AHCPR-sponsored heart failure guideline: translating practice recommendations into review criteria. *Jt Comm J Qual Improv* 1996; 22: 265–76
- 11 Fraser RC, Khunti K, Baker R, et al. Effective audit in general practice: a method for systematically developing audit protocols containing evidence-based review criteria. *Br J Gen Pract* 1997; 47: 743–6
- 12 Ashton CM, Kuykendall DH, Johnson MS, et al. A method of developing and weighting explicit process of care criteria for quality assessment. *Med Care* 1994; 32: 755–70.
- 13 Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998; 338: 1888–95.
- 14 Johnson N. Feasibility of developing and selecting criteria for the assessment of clinical performance. *Br J Gen Pract* 1993; 43: 499–502.
- 15 Bateman DN, Eccles M, Campbell M, et al. Setting standards of prescribing performance in primary care: use of a consensus group of general practitioners and application of standards to practices in the North of England. *Br J Gen Pract* 1996; 46: 20–5.
- 16 Brook RH, Chassin MR, Fink A, et al. A method for the detailed assessment of the appropriateness of medical strategies. *Int J Technol Assess Health Care* 1986; 2: 53–63.
- 17 Campbell SM, Roland MO, Shekelle PG, et al. Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice. *Quality in Health Care* 1999; 8: 6–15.
- 18 Eccles M, Clapp Z, Grimshaw J, et al. North of England evidence based guidelines development project: methods of guideline development. *BMJ* 1996; 312: 760–2.
- 19 Murphy MK, Black NA, Lamping DL, et al. Consensus development methods and their use in clinical guideline development. *Health Technol Assess* 1998;2:i-iv.
- 20 Turoff M. The policy Delphi. In: Linstone HA, Turoff M, eds. *The Delphi method: techniques and applications*. London: Addison-Wesley, 1975.
- 21 Critcher C, Gladstone B. Utilizing the Delphi technique in policy discussion: a case study of a privatized utility in Britain. *Public Administration* 1998;76:431–49.
- 22 McKnight JN, Edwards N, Pickard L, et al. The Delphi approach to strategic planning. *Nurs Manage* 1991;22:55–7.
- 23 Linstone HA, Turoff M, eds. *The Delphi method: techniques and applications*. London: Addison-Wesley, 1975.
- 24 Mitchell VW. Using Delphi to forecast in new technology industries. *Market Intelligence and Planning* 1992;10:4–9.
- 25 Guba ES, Lincoln YS. *Fourth generation evaluation*. London: Sage, 1989.