Papers

Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice

S M Campbell, M O Roland, P G Shekelle, J A Cantrill, S A Buetow, D K Cragg

National Primary Care Research and Development Centre, University of Manchester, UK
S M Campbell, research fellow
M O Roland, professor
J A Cantrill, senior lecturer
S A Buetow, research fellow

RAND Corporation, Santa Monica, California, USA
P G Shekelle, atlantic fellow

Department of General Practice, University of Manchester, UK
D K Cragg, lecturer

Correspondence to:
Mr Stephen Campbell, National Primary Care Research and Development Centre, Williamson Building, University of Manchester, Oxford Road, Manchester M15 6PL. Telephone 0044 161 275 7655; fax 0044 161 275 7600; email stephen.campbell@man.ac.uk

Abstract
Objective—To develop review criteria to assess the quality of care for three major chronic diseases: adult asthma, stable angina, and non-insulin dependent diabetes mellitus.
Subjects and methods—Modified panel process based upon the RAND/UCLA (University College of Los Angeles) appropriateness method. Three multiprofessional panels made up of general practitioners, hospital specialists, and practice nurses.
Results—The RAND/UCLA appropriateness method of augmenting evidence with expert opinion was used to develop criteria for the care of the three conditions. Of those aspects of care which were rated as necessary by the panels, only 26% (16% asthma, 10% non-insulin dependent diabetes, 40% angina) were subsequently rated by the panels as being based on strong scientific evidence.
Conclusion—The results show the importance of a systematic approach to combining evidence with expert opinion to develop review criteria for assessing the quality of three chronic diseases in general practice. The evidence base for the criteria was often incomplete, and explicit methods need to be used to combine evidence with expert opinion where evidence is not available.
(*Quality in Health Care* 1999;**8**:6–15)

Keywords: review criteria; quality; general practice

Introduction
Defining and measuring the quality of medical care has become increasingly important in recent years. Approaches to the measurement of clinical quality include the development of clinical guidelines,[1–6] performance indicators,[7–10] and review criteria.[11 12] Review criteria are "systematically developed statements that can be used to assess the appropriateness of specific healthcare decisions, services and outcomes".[6] Review criteria, although often derived from guidelines, can be confused with them. In essence, guidelines guide care prospectively whereas review criteria assess care retrospectively. For example, if a guideline recommends that the blood pressure of diabetics should be taken annually, a corresponding review criterion might be that "the records of diabetics show a blood pressure reading in the previous 12 months".

The characteristics of ideal review criteria have been described by Baker and Fraser[11] and are outlined in box 1. This paper describes the application of a method of developing review criteria; in which we have attempted to get as close as possible to these characteristics.

The focus on research evidence is important. The extent to which research based criteria can be developed is an important issue for quality improvement, given the relative lack of evidence for many healthcare procedures.

Review criteria are most often used to assess aspects of care which can be verified retrospectively from patient records. They therefore measure both whether something was done, and whether it was recorded. Although the link is not necessarily direct, there is evidence both that information in medical records does relate to care actually given,[13] and that there is an association between the quality of record keeping and the outcome of care.[14] In the study by Kahn *et al*, the process of care was related to outcome both for measures which were heavily dependent on the quality of record keeping, and for those which were not.

In this study, we aimed to develop evidence-based review criteria for the assessment of necessary and appropriate care for three major chronic diseases: adult asthma, stable angina, and non-insulin dependent diabetes mellitus

- Based on research evidence
- Prioritised according to strength of research evidence and influence on outcome
- Measurable: clear and precise
- Appropriate to the clinical setting

Box 1   Ideal characteristics of review criteria.[11]

- Combines systematic review of the scientific literature with expert opinion
- Yields specific criteria that can be used for review criteria or practice guidelines, or both
- Provides a quantitative description of the expert judgement of a multidisciplinary group of clinicians
- Iterative, anonymous rating and a face to face panel discussion
- Each panellist has equal weight in determining the final result

Box 2   Key characteristics of the RAND appropriateness method.[15]

(NIDDM). These three conditions were selected because they are associated with high levels of mortality and morbidity in the community and represent a substantial amount of chronic disease workload in general practice. We are currently applying these criteria to records in a representative sample of practices, with the intention of reducing them to a smaller set that can act as a proxy for the full sets of review criteria.

Method
The method we used was based on the RAND/UCLA (University College of Los Angeles) consensus panel method for assessing the appropriateness of medical procedures.[15] The RAND method has been used to assess the appropriateness of procedures or the necessity of procedures or both[16][17] and has been used previously in the United Kingdom.[18][19] It combines expert opinion with evidence in a more measurable way than other approaches—such as that developed by the Agency for Health Care Policy and Research (AHCPR).[12] The key characteristics of the RAND method are outlined in box 2.

There were three stages used in the study once the conditions had been selected; the generation of review criteria, panel selection, and the rating process.

GENERATION OF REVIEW CRITERIA
Review criteria are often based on clinical guidelines,[11] and accordingly we started with previously published evidence-based guidelines for the management of asthma,[20–22] angina,[23][24] and non-insulin dependent diabetes mellitus.[25][26] For asthma and angina, these were based on systematic reviews of the scientific literature. We updated the systematic literature reviews for all three conditions by repeating the search strategy described in the North of England studies which included their published methodological screen[21][24] to identify new evidence, particularly randomised controlled trials. This included searches for research published between September 1994 and May 1997 for asthma, between September 1994 and August 1997 for angina, and between January 1990 and August 1996 for diabetes. Computerised literature searches were supplemented with hand searches from August 1996 up to the date of each panel meeting (April, June, and September 1997 for diabetes, asthma, and angina respectively). Journals hand searched were *The Lancet, New England Journal of Medicine, Journal of the American Medical Association, British Medical Journal, and Archives of Internal Medicine*. All three updated literature reviews of new evidence were scrutinised by two external experts to check for completeness and accuracy. The guideline statements and new evidence from literature reviews were then used to construct lists of initial potential review criteria by members of the research team for angina (SAB/PGS), asthma (SMC/DKC) and NIDDM (JAC/MOR); in accordance with previous application of the method. The lists drawn up by the research team formed the questionnaires rated in round 1. These are available from the authors and consisted of 136 asthma, 308 angina, and 196 diabetes criteria, respectively.

PANEL SELECTION
We assembled three multidisciplinary panels of people with nationally recognised expertise in angina, asthma, and diabetes. Panellists were selected on the basis of their publication record, involvement in specialist societies—such as the General Practitioners in Asthma Group (GPIAG), or the British Diabetic Association—or after advice from researchers in the field. Panellists, who were selected to cover a range of clinical expertise and a geographical spread across the United Kingdom, were sent a letter of invitation to participate. Panellists who declined were asked to nominate a potential replacement. About three quarters of people asked to take part agreed.

Panels included general practitioners, specialists, and practice nurses. The asthma and diabetes panels had 11 members: three asthma hospital specialists, six general practitioners, and two practice nurses. The angina panel had 10 members: four hospital specialists, and six general practitioners, but no nurses as there are few specialised primary care angina nurses in the United Kingdom. All the general practitioners were principals, many working full time in general practice. The members of each of the panels are listed in appendix 1.

RATING PROCESS
Panel rating for necessary care and inappropriate care
Members of each panel were sent copies of the relevant evidence-based guidelines from which we had worked, a summary of more recent published evidence with copies of the articles identified, and a set of initial review criteria drawn up by the research team.

Panellists were asked to rate each criterion in the relevant list, and state whether each action was necessary, and also, for certain criteria, whether the action was appropriate. Sample criteria for the asthma panel are shown in box 3.

Necessary care was defined in terms of three criteria: reasonable likelihood that the action will provide net benefit to the patient; the benefit to the patient is not likely to be small; and it would be improper care not to have performed or recommended the action.[15] The

distinction between necessary and appropriate is important. Kahan *et al*[16] refer to necessary care as a "stronger" evaluation tool than appropriateness as necessary care is "crucial" care that a physician or health professional is "obliged" to recommend or undertake because not to do so would be "deleterious to the patient's health". Inappropriate care was thus defined as occurring where the expected health benefits of performing the action were exceeded by the expected healthcare risks by a sufficient amount, that the procedure was not worth doing.[16]

We asked about necessary care where the quality problem was likely to be underprovision (necessary procedures or treatments not performed). We asked about appropriateness to specifically identify inappropriate care for the much smaller number of review criteria where the quality problems were likely to be the over-provision of care, or provision of care which was potentially harmful (inappropriately performed procedures or treatments). Ratings used an ordinal scale of 1–9, where 1 meant an action was clearly not necessary or extremely inappropriate and 9 meant that an action was clearly necessary. No inappropriateness criteria were rated by the diabetes panel.

For example, from box 3, a rating of 9 on the necessity scale for criterion 10.1g would be a signal that the respondent considered use of oral steroids highly necessary if peak flow has fallen to 50% of predicted or best value; whereas a rating of 1 for the criterion 10.1a on the appropriateness scale would have indicated the respondent considered use of oral steroids extremely inappropriate on the basis of a peak flow of 80% of predicted or best.

The process was designed to set minimum, not optimal standards of care. With the scientific evidence provided and their judgement, panellists were asked to rate each criterion with reference to an average patient consulting an average general practitioner in 1997. There was explicitly no consideration of cost during the panel discussions. Throughout the sessions, panellists were reminded that their decisions should be about whether an aspect of care should be recorded on the medical records.

Ratings were confidential and carried out in two rounds. The first round of rating was done by post, with panel members scoring each of the review criteria supplied by the research team. Panels then met for a full day, when the results of all their earlier ratings were fed back in a way which allowed each panellist to know his or her score for an individual item and also the overall distribution of the whole panel's scores. Panellists had the opportunity to discuss all the first round criteria during the panel meeting under the chairmanship of a moderator (diabetes and angina, PGS, asthma, MOR). After discussion each item was rerated. In some cases, criteria were reworded, or new criteria were added during the exercise. This is an integral part of the panel process and allows the panellists to reword or restructure criteria better to fit their clinical judgement. For example, members of the asthma panel decided that the first round criteria failed adequately to reflect guidance of management of symptoms: an additional 12 criteria were added that related to control of symptoms. Only the ratings of the second round were used to develop the final lists of review criteria (appendices 2, 3, and 4).

3.1 There should be at least one peak flow recording for all patients who can use a peak flow meter

	Necessity scale
	1 2 3 4 5 6 7 8 9

10.1 The records show that in exacerbations of asthma, oral steroids are used when the peak flow has fallen to less than:

	% Predicted or best	Appropriateness scale	Necessity scale
a	80	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
b	75	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
c	70	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
d	65	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
e	60	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
f	55	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
g	50	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

Box 3   Examples of criteria used by panellists to rate for necessity and appropriateness.

Angina
Example of criterion rated as a necessary aspect of care for angina (median panel score of 9)
- The records show that, in the absence of contraindications, the patient has been offered treatment with aspirin

Example of criterion for angina regarded as inappropriate care (median panel score of 1)
- Verapamil was used as second line treatment for patients on β blockers

Asthma
Example of criterion rated as a necessary aspect of care for asthma (median panel score of 9)
- In consultations associated with an exacerbation (defined as an increase in symptoms of dyspnoea, or wheeze or limitation of activity), the peak flow is recorded

Example of criterion for asthma regarded as inappropriate care (median panel score of 1)
- Treatment with intravenous β agonists or intravenous aminophylline is inappropriate unless there is cotreatment with an inhaled β agonist

Non-insulin dependent diabetes mellitus
Examples of criteria rated as necessary aspects of care for non-insulin dependent diabetes (median panel score of 9)
- The records show a recording of HbA1c level in the past 14 months
- Visual foot examination for ulceration or deformity recorded in the past 14 months

Box 4   Examples of necessary and inappropriate criteria.

Table 1   Summary of necessity and appropriateness ratings

		Necessity		Inappropriateness	
		Round 1	Round 2	Round 1	Round 2
Asthma:					
	Criteria (n)	111	115	24	22
	Median 1 or 2	4	6	**3**	4
	Median 3–7	55	61	11	10
	Median 8 or 9	**52**	**48**	10	9
	Necessary (%)	39	38	—	—
	Inappropriate (%)	—	—	8	18
	Disagreement (%)	4	7	4	0
Diabetes:					
	Criteria (n)	196	183	—	—
	Median 1 or 2	6	28	—	—
	Median 3–7	108	73	—	—
	Median 8 or 9	**82**	**82**	—	—
	Necessary (%)	42	45	—	—
	Inappropriate (%)	—	—	—	—
	Disagreement (%)	4	7	—	—
Angina:					
	Criteria (n)	260	272	48	51
	Median 1 or 2	28	56	**3**	**10**
	Median 3–7	93	96	26	23
	Median 8 or 9	**139**	**120**	19	20
	Necessary (%)	54	44	—	—
	Inappropriate (%)	—	—	6	18
	Disagreement (%)	4	1	2	2

Figures in bold represent the criteria identified as necessary and inappropriate.

In the second round of ratings, we used Brook's definition of disagreement as three or more scores in both the top (6-9) and bottom (1-3) tertile[15] and selected necessary criteria as those which had a median score of 8 or 9 (out of a maximum possible of 9) for necessity without disagreement by the panel. We identified inappropriate care as those which had a median rating of 1 or 2 for appropriateness, again without disagreement among the panel.

Rating of evidence base of final review criteria
The RAND appropriateness method was then supplemented by the addition of a third postal round of ratings designed to assess the evidence base of aspects of care rated necessary, based on a modification of the method used previously by AHCPR.[6 12] The panellists were asked to rate the strength of evidence for aspects of care identified as either necessary or inappropriate (8 or 9 for necessity or 1 or 2 for appropriateness) on the following scale: (A) mainly convincing scientific evidence usually consisting of one or more randomised controlled trials, limited added expert opinion; (B) about equal parts scientific evidence and expert opinion; (C) mainly expert opinion, limited scientific evidence. The final review criteria listed in the appendices show median panel scores and evidence-based scores for the second (panel meeting) round of ratings.

Table 2   Summary of panels' rating of the level of evidence for final criteria

		Level of evidence		
	Total	A	B	C
Angina criteria rated 8 or 9 for necessity, or 1 or 2 appropriateness	130	52	68	10
Asthma criteria rated 8 or 9 for necessity, or 1 or 2 for appropriateness	51	8	30	13
Diabetes criteria rated 8 or 9 for necessity	82	8	54	20
	263	26%	58%	16%

A=mainly convincing scientific evidence, usually consisting of one or more RCTs, limited added expert opinion; B=about equal parts scientific evidence and expert opinion; C=mainly expert opinion, limited scientific evidence.

The data presented in the next section will describe examples of criteria rated necessary and inappropriate for each condition; which are listed in full in appendices 2, 3, and 4. Summary data for all criteria rated by each panel will be presented, as will data relating to the level of evidence for all criteria rated as either necessary or inappropriate.

Results
RATING PROCESS
Median panel score ratings for each criterion were calculated to identify criteria rated as necessary (median score 8 or 9) for all conditions and also inappropriate (median score 1 or 2) for asthma and angina. Examples of necessary and inappropriate criteria for each of the three conditions are listed in box 4.

The full list of review criteria defined as necessary and inappropriate are listed in appendices 2 (asthma), 3 (NIDDM), and 4 (angina). We only included criteria rated 8 or 9 for necessary (highly necessary) or 1 and 2 for appropriateness (highly inappropriate) because the reliability of the group process increases at the extreme ends of the scoring range.

A summary of the criteria rated as either necessary or inappropriate in the two rounds for each of the three studies is shown in table 1. All three panel meetings resulted in changes in the wording or structure of some criteria. The number of criteria listed in the appendices is smaller than the number in table 1 as panels made multiple ratings for some aspects of care. For example, for cholesterol lowering treatment in patients with angina, the angina panel made separate ratings, depending on age, sex, three levels of serum cholesterol, presence of diabetes, and previous myocardial infarction (48 separate ratings). For some of these, it was possible to condense the results in the appendices—for example, for patients with angina who had diabetes and a previous myocardial infarction, an offer of lipid lowering treatment was regarded as necessary for all patients with a total cholesterol >5.5 mmol/l (making ratings of concentrations >5.5 mmol/l superfluous).

RATING OF EVIDENCE BASE
The level of evidence which the panels scored for the necessary criteria (median 8 or 9) and inappropriate criteria (median 1 or 2) is shown in table 2. Only 26% of criteria (16% asthma, 10% NIDDM, 40% angina) were rated as being based on strong scientific evidence (level A mainly convincing scientific evidence, usually consisting of one or more randomised controlled trials, limited added expert opinion) and 76% of these related to angina. A list of criteria only containing those backed by strong scientific evidence would have excluded most criteria judged by the panel to be necessary aspects of care.

Examples of aspects of care for diabetes regarded as necessary by the panel (median score 9), but with varying levels of evidence are listed in box 5.

- The practice maintains a register of patients with a diagnosis of diabetes (level of evidence C)
- The records contain details, recorded within the past 14 months, of peripheral pulses (level of evidence B)
- If total cholesterol is >7.8 mmol/l, and the patient has ischaemic heart disease, lipid lowering treatment has been offered (level of evidence A)

Box 5   Examples of the evidence base for the review criteria for diabetes rated to be necessary.

Discussion

This paper reports the development of review criteria for assessing the quality of medical care by combining the best available evidence with expert opinion; thereby adhering to the key characteristics of the RAND method (box 2). In accordance with recommendations identified as attributes of good review criteria (box 1)[11] each criterion was assessed for magnitude of benefit, judged by high scores for necessity or inappropriateness, and rated for strength of evidence. Rating the strength of evidence for each criterion identified as necessary or inappropriate, built upon previously reported accounts of the RAND methods.

EVIDENCE BASE

In the judgement of our panels, although most of the final criteria were judged to have about equal parts scientific evidence and expert opinion, there was strong scientific evidence for only 26% of the criteria which panels none the less regarded as clearly necessary aspects of care. The higher percentage of good evidence for the angina criteria reflects the extent of previous research, particularly randomised controlled trials, in this area compared with the other two conditions. Importantly, it also emphasises that the evidence-base varies across different clinical conditions.

ADVANTAGES OF THE MODIFIED RAND METHOD

Where evidence for aspects of care is lacking, a systematic approach to gathering expert opinion is required. The prime use of the modified RAND method is its capacity to generate specific ratings for clearly necessary or inappropriate aspects of care (an essential attribute of review criteria); which can be applied equally to areas where there is either extensive or limited evidence. We have added to the previously used method an additional round in which the panel rated the strength of evidence for each necessary review criterion. We think this is a significant improvement on the published method.

We think that our results show the importance of combining evidence and expert opinion in the development of review criteria for general practice in the United Kingdom, due to lack of evidence for many aspects of care. There are significant limitations to approaches which rely solely upon a scientific evidence base to identify care which is judged as necessary or inappropriate.

VALIDITY OF THE RAND METHOD USED

Our results depend on the validity and reliability of the RAND method. The method shows acceptable reliability, although the consistency between panels is higher for judgements about what doctors should do (necessary treatments or procedures) than what they should not do (appropriate or inappropriate treatments or procedures).[27] The method is certainly sufficiently reliable for group comparisons. The recommendations of the panels are in general in line with the management of larger groups of doctors,[28] and Naylor[29] concludes that the RAND use of Delphi panels "continues to offer one of the only meticulously tested and systematic methods for leavening limited evidence with expert opinion and inference". The few studies that have been done of the predictive validity of the method support the predictive validity of the criteria produced.[30–32] However, the previously published RAND method has been criticised for not rating the strength of evidence for the necessary aspects of care, and therefore failing to differentiate between them. We therefore modified the RAND method to incorporate a rating of strength of evidence as recommended by Baker and Fraser.[11]

Concern has also been expressed about the reproducibility of panel findings and the influence of dominant personalities or power struggles.[33–35] Ultimately, as Grol[1] has emphasised it is essential that consensus methods are explicit. A fundamental feature of the RAND method is that the group judgement process is explicit, with predefined rules for what constitutes disagreement, and what ratings will pass as criteria. Furthermore, it is based upon a systematic literature review, can be multidisciplinary, and is democratic in that the views of each panellist have equal weight (box 2). The quantitative risk benefit scale facilitates relative as well as absolute judgements.

IMPLICATION FOR QUALITY IMPROVEMENT

Our findings have important implications for quality assessment and quality improvement as they highlight a potential flaw in the drive towards evidence-based medicine. Some interventions are unlikely ever to be subjected to evaluation with randomised controlled trials—for example, the use of practice diabetic registers or the recording of signs and symptoms. It would also be unethical to conduct a randomised controlled trial on an aspect of care considered inappropriate by experts. The need to include expert opinion is equally important for clinical guidelines. For example, of the 57 recommendations in the North of England evidence-based clinical practice guideline for angina[23] (which were used as the starting point for generating the angina review criteria), 37 were based on uncontrolled studies, consensus meetings, or extrapolations from other evidence.

We recognise that there are potential limitations to methods which rely on expert opinion for many of the final criteria. Concern has been expressed that expert opinion simply rates conventional wisdom without scientific evidence. However, criteria based exclusively upon evidence are arguably non-existent as it is

almost inevitable that they will contain covert expert opinion—for example, the generalisation of trial evidence to wider populations. Moreover, without the use of expert opinion a limited number of criteria would have been generated. Most guidelines and review criteria are developed with methods that incorporate expert opinion in far less formal ways, and with less empirical evidence of reliability and validity of the process, than the method presented. The RAND approach incorporates expert opinion in a measurable way. This means that it is particularly relevant for many conditions presenting in general practice and acute conditions; for which there is so little evidence that the approach described here would rely largely on opinion rather than on evidence.

POTENTIAL LIMITATIONS OF REVIEW CRITERIA

The use of review criteria—such as those presented in this paper—rely on data available in medical records. This also has important implications for quality improvement as it means that two elements of care are being assessed: the care given and the care actually recorded. Although all three panels were clear that all items in final selections should be aspects of care recorded in the medical records, the use of records to collect data may introduce bias. For example, a doctor who does not record that a patient has consulted with an exacerbation of asthma cannot be judged against the criteria that relate to management of exacerbations.

To be valid, review criteria must be related to and thus measure clearly defined aspects of care. It is important to recognise that there are additional areas of quality—for example, interpersonal care—that are also important in the management of these conditions but which are far less easy to define or measure. Furthermore, measurement quality focused on a particular condition does not take into account the complexity of real patients who may have greater medical or social problems than diabetes, asthma, or angina. There is currently little evidence on whether the results for individual conditions can be generalised to reflect the overall quality of care in a patient or a practice. Furthermore, in the development of these criteria, patients' views of management for the three conditions are lacking.

ONGOING DEVELOPMENT

The sets of review criteria in the appendices have been converted into simple data abstraction forms suitable for application to medical records. These are available from the authors, who have so far applied them to about 400 sets of general practice records for patients with stable angina, adult asthma, and non-insulin dependent diabetes mellitus. Collection of data for each set of records takes about 20 minutes. An SPSS programme to convert the condensed data from the abstraction sheets into the data required to assess the review criteria in the three appendices is also available from the authors. These review criteria form the basis of our ongoing research. We plan to identify for each condition a subset of criteria which can be used to summarise the larger sets of data and which will be usable for routine quality assessment.

Conclusion

Our experience has been that, despite certain caveats, the RAND appropriateness method of augmenting evidence with expert opinion, although time consuming, can be used to generate review criteria for general practice, which conform to essential attributes for assessing quality. The lack of research evidence for many aspects of care highlights the importance of the use of methods that allow explicit combination of both evidence and expert opinion.

1 Grol R. Development of guidelines for general practice care. *Br J Gen Pract* 1993;**43**:146–51.
2 Forrest D, Hoskins A, Hussey H. Clinical guidelines and their implementation. *Postgrad Med J* 1996;**72**:19–22.
3 Thomson R, Lavender M, Madhok R. How to ensure that guidelines are effective. *BMJ* 1995;**311**:237–42.
4 Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;**342**:1317–20.
5 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.
6 Field MJ, Lohr KN, eds. *Committee on Clinical Practice Guidelines: guidelines for clinical practice: from development to use*. Washington, DC: Institute of Medicine, 1992.
7 Majeed FA, Voss S. Performance indicators for general practice: will lead to league tables of performance. *BMJ* 1995;**311**:209–10.
8 Radical Statistics Health Group. NHS indicators of success: what do they tell us? *BMJ* 1995;**310**:1045–50.
9 Anderson BG, Noyce JA. Clinical indicators and their role in quality management. *Aust Clin Rev* 1992;**12**:15–21.
10 Angaran DM. Selecting, developing and evaluating indicators. *Am J Hosp Pharmacol* 1991;**48**:1931–7.
11 Baker R, Fraser RC. Development of review criteria: linking guidelines and assessment of quality. *BMJ* 1995;**311**:370–3.
12 Using clinical practice guidelines to evaluate quality of care. Vol 2. Methods. Rockville, MD: Agency for Health Care Policy and Research, 1995.
13 Kosecoff J, Fink A, Brook RH, *et al*. The appropriateness of using a medical procedure: is information in the medical record valid? *Med Care* 1987;**25**:196–201.
14 Kahn KL, Rogers WH, Rubenstein LV, *et al*. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *JAMA* 1990;**264**:1969–73.
15 Brook RH, Chassin MR, Fink A, *et al*. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;**2**:53–63.
16 Kahan JP, Bernstein SJ, Leape LL,et. Measuring the necessity of medical procedures. *Med Care* 1994;**32**:357–65.
17 Kahan JP, Park RE, Leape LL, *et al*. Variations by speciality in physician ratings of the appropriateness and necessity of indications for procedures. *Med Care* 1996;**34**:512–23.
18 Gray D, Hampton JR, Bernstein SJ, *et al*. Audit of coronary angiography and bypass surgery. *Lancet* 1990;**335**:1317–20.
19 Scott EA, Black N. Appropriateness of cholecystectomy in the United Kingdom: a consensus panel approach. *Gut* 1991;**32**:1066–70.
20 Centre for Health Services Research. *The primary care management of asthma in adults. North of England evidence based guideline development project*. Newcastle-upon-Tyne: Centre for Health Services Research, 1996.
21 BMJ. *Evidence based clinical guidelines. The primary care management of asthma in adults. North of England evidence based guideline development project*. http://www.bmj.com/bmj/asthma/ast01.htm
22 The British guidelines on asthma management 1995: review and position statement. *Thorax* 1997;**52**:S1–21
23 The primary care management of stable angina. North of England evidence based guideline development project. Newcastle-upon-Tyne: Centre for Health Services Research, 1996.
24 BMJ. *Evidence based clinical guidelines. The primary care management of stable angina. North of England evidence based guideline development project*. http://www.bmj.com/bmj/angina/ang01.htm
25 American Diabetic Association. Clinical practice recommendations 1997. *Diabetes Care* 1997;**20**:S1–70.
26 Saint Vincent and improving diabetes care. Specialist UK Workgroup Reports. *Diabetic Medicine* 1996;**13**:S1–128.
27 Shekelle PG, Kahan JP, Bernstein SJ, *et al*. The reproducibility of a method to identify the overuse and underuse of procedures. *N Engl J Med* 1998;**338**:1888–95.
28 Ayanian JZ, Landrum MB, Normand ST, *et al*. Rating the appropriateness of coronary arteriography: do practising physicians agree with an expert panel and with each other? *N Engl J Med* 1998;**338**:1896–904.
29 Naylor D. What is appropriate care [editorial]? *N Engl J Med* 1988;**338**:1918–20.

30 Kravitz RL, Laouri M, Kahan JP, *et al*. Validity of criteria used for detecting underuse of coronary revascularization. *JAMA* 1995;**274**:632–8.

31 Selby JV, Pireman BH, Lundstrom RJ, *et al*. Variation among hospitals in coronary angiography practices and outcomes after myocardial infarction in a large health maintenance organisation. *N Engl J Med* 1996;**328**:1888–96.

32 Shekelle PG, Chassin MR, Park RE. Evidence for the predictive validity of the RAND/UCLA appropriateness method: criteria for performing carotid endarterectomy. *Int J Technol Assess Health Care* (in press).

33 Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ* 1994;**309**:730–3.

34 Charlton BG. Practice guidelines and practical judgement: the role of mega-trials, meta-analyses, and consensus. *Br J Gen Pract* 1994;**44**:290–1.

35 Mulholland AV, Wheeler SG, Heieck JJ. Medical assessment by a Delphi group opinion technique. *N Engl J Med* 1973;**288**:1272–5.

Appendix 1: List of panellists

Angina

Professor D De Bono, Department of Cardiology, Glenfield General Hospital, Leicester LE3 9QP.

Dr M Gandhi, Department of Cardiology, Southampton General Hospital, Trenmora Road, Southampton SO16 6YD.

Dr H Dargie, Department of Cardiology, West Medical Building, University of Glasgow, Glasgow G12 8QQ.

Professor D Wood, National Heart and Lung Institute, Dovehouse, London SW3 6LY.

Professor M Eccles, Centre for Health Services Research, University of Newcastle upon Tyne, 21 Claremont Place, Newcastle NE2 4AA.

Dr K Khunti, Eli Lilly National Clinical Audit Centre, Department of General Practice, University of Leicester, Leicester General Hospital, Leicester LE5 4PW.

Dr A Farmer, The Health Centre, East Street, Thame, Oxon OX9 3JZ.

Dr I Hughes, GP, Cumber land House, 58 Scarisbrick New Road, Southport, Merseyside PR8 6PQ.

Dr M Cupples, Department of General Practice, Dunluce Health Centre, 1 Dunluce Avenue, Belfast B79 7HR.

Dr N Hicks, Department of Public Health and Health Policy, Oxfordshire Health Authority, Old Road, Headington, Oxford OX3 7LG.

Asthma

Dr KP Jones, Department of Primary Health Care, University of Newcastle-upon-Tyne, The Medical School, Framlington Place, Newcastle-upon-Tyne NE2 4HH.

Dr C Griffiths, Department of General Practice and Primary Care, St Barts and the Royal London School of Medicine, New Science Block, Charterhouse Square, London EC1M 6BQ.

Dr R Neville, Tayside Centre for General Practice, University of Dundee, Charleston Drive, Dundee DD2 4AD.

Dr D Ryan, Woodbrook Medical Centre, 28 Bridge Street, Loughborough, Leics LE11 1NH.

Dr D Bellamy, The Surgery, 4 Tolpuddle Gardens, Bournemouth BH9 3LQ.

Dr R Baker, Eli Lilly National Clinical Audit Centre, Department of General Practice, University of Leicester, Gwendolen Road, Leicester LE5 4PW.

Dr JAR Friend, Department of Thoracic Medicine, Aberdeen Royal Infirmary, Fosterhill, Aberdeen AB25 2ZN.

Professor Geddes, National Heart and Lung Institute, Royal Brompton Hospital, Sydney Street, London SW3 6NP.

Dr C Hardy, Department of Medicine, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL.

Ms S Buck, Asthma specialist nurse, 8 Mont le Grand, Exeter EX1 2PD.

Ms A Pearson, National Asthma Campaign, Providence House, London N1 0NT.

Diabetes

Dr R Young, Department of Diabetes and Endocrinology, Hope Hospital, Salford M6 8HD.

Dr S Bootle, Goyt Valley General Practice, Chapel Street, Whalley Bridge, Stockport.

Ms G Hall, Stable Cottage, Headley Road, Greyshott, Hindhead, Surrey GU26 6DL.

Ms P Wilson, 113 Station Avenue, Sandown, Isle of Wight , PO36 8HD.

Dr K Paterson, Strathcashel, Lochlibo Road, Uplawmoor, Glasgow G78 4AA.

Professor A Sinclair, Department of Medicine, University of Birmingham, Heywood Building, Selly Oak Hospital, Raddleton Road, Birmingham B29 6JD.

Dr P Tasker, St James House surgery, County Court Road, Kings Lynn, Norfolk PE30 5EJ.

Dr R Gadsby, Redroofs, 31 Coton Road, Nuneaton CV11 5TW.

Dr M Pierce, Academic General Practice Unit, Chelsea and Westminster Hospital, 360 Fulham Road, London SW10 9NH.

Dr P M Greenhalgh, Department of Primary Care and Population Sciences, Whittington Hospital, Highgate Hill, London N19 5NF.

Dr C St J Kear, 35 Church Street, Wath upon Dearne, South Yorkshire S63 7RF.

Appendix 2: Asthma in adults

Key

Numbers: indicate median panel score on a nine point ordinal scale for necessity of that aspect of care.

Letters: (A) mainly convincing scientific evidence, usually consisting of one or more randomised controlled trials, limited added expert opinion; (B) about equal parts scientific evidence and expert opinion; (C) mainly expert opinion, limited scientific evidence.

Practice level data recording

The practice maintains a register of patients with a diagnosis of asthma (C, 9)

The register shows information about admissions to hospital (C, 9)

Diagnostic criteria

The records show one of: a history of symptoms of recurrent wheeze, symptoms of airway narrowing in response to aero-allergens,

cold air, or exercise, diurnal variation of peak flow of at least 15%, or an increase in peak flow of at least 15 % following the inhalation of β2 agonist or a 1 week trial of oral steroids (B,8)

Recording of peak flow
Where a patient has a technical difficulty in using a peak flow meter, this is recorded in the notes (B, 8).

There should be at least one peak flow reading for all patients who can use a peak flow meter (C, 8)

There should be a record of peak flow when well, or best or predicted peak flow (B, 8)

Recording of symptoms
In the last 14 months, there is a record of the level of control of asthma by assessment of day time symptoms or nocturnal symptoms, or activity limitation (B, 9)

Short acting β agonists
The records show that in patients with exercise induced bronchospasm, short acting bronchodilators are prescribed for use before exercise (A, 8).

Inhaled anti-inflammatory drugs
The records show that patients who require three or more doses of short acting inhaled β-agonists daily have also been offered an inhaled anti-inflammatory drug (A, 9)

The records show that patients experiencing daytime symptoms more frequently than 6 days a week have been offered an inhaled anti-inflammatory drug (A, 8)

The records show that patients experiencing nocturnal symptoms more than once a week have been offered an inhaled anti-inflammatory drug (C, 8)

The records show that patients experiencing exacerbations which affect activity have been offered an inhaled anti-inflammatory drug (A, 9)

Long acting bronchodilators, high dose inhaled steroids, oral steroids, and bronchodilators
Where there is evidence from the notes that symptoms are not controlled using β agonists and low dose inhaled steroids, a trial of long acting bronchodilator or high dose inhaled steroid is offered (B, 9)

The records show that all patients on long acting inhaled beta-agonists are taking an anti-inflammatory drug (B, 9)

The records show that a patient on long term maintenance oral steroids should have had a trial of: a long acting bronchodilator, and high dose inhaled steroids and one other step 4 treatment from the British Thoracic Society guidelines (B, 9)

The records show that a patient who is on a maintenance oral beta-agonist has been offered a trial of inhaled β agonist treatment (C, 9)

Treatment in exacerbations of asthma
In consultations associated with an exacerbation (defined as an increase in symptoms of dyspnoea, or wheeze, or limitation of activity), the peak flow is recorded (B, 9)

The records show that in exacerbations of asthma when the peak flow has fallen to less than 60% of best or predicted value, a short acting inhaled β agonist delivered through either a nebuliser or large volume spacer is used (B, 8)

The records show that in exacerbations of asthma when the peak flow has fallen to less than 60% of best or predicted value, oral steroids are offered (B, 9). Treatment with oral steroids is inappropriate unless the predicted peak flow has fallen to below 80% of predicted or best. (B,1).

In an exacerbation requiring immediate inhaled bronchodilator treatment, at least one of the following is recorded: speech, pulse rate, or respiratory rate (B, 9)

Treatment with intravenous β-agonists (C,1) or intravenous aminophylline (C,1) is inappropriate unless there is cotreatment with an inhaled β-agonist.

Patient education and smoking
The records show smoking status (C, 9)

The records show that advice about cessation has been given to smokers (B, 9)

The records show that advice has been given about the recognition of the danger signs of an exacerbation at least once (B, 9)

The records show that advice has been given about the importance of seeking medical advice early in an attack at least once (B, 9)

Inhaler technique
The records show whether the inhaler device can be used properly by the patient (A,9)

The records show that if maintenance treatment is to be increased, inhaler technique has been checked, (unless there is a prior record of satisfactory inhaler technique) (B, 9)

The records show that if maintenance treatment is to be increased, enquiry has been made about compliance with treatment (B, 8)

Self management plan
In all patients on high dose inhaled steroids, the records show evidence of an explicit self management plan detailing action to be taken on deterioration of peak flow and symptoms (B, 9)

In all patients who have had hospital inpatient treatment, the records show evidence of an explicit self management plan detailing action to be taken on deterioration of peak flow and symptoms (A, 9)

Referral
The records show that where oral steroids are used in maintenance treatment the patient has been referred to a respiratory physician

The records show that where occupational asthma is suspected, the patient has been referred to a respiratory physician

Appendix 3: Non-insulin dependent diabetes mellitus

Key
Numbers: indicate median panel score on a nine point ordinal scale for necessity of that aspect of care.

Letters: (A) mainly convincing scientific evidence, usually consisting of one or more randomised controlled trials, limited added expert opinion; (B) about equal parts scientific evidence and expert opinion; (C) Mainly expert opinion, limited scientific evidence.

Practice level data recording
The practice maintains a register of patients with a diagnosis of NIDDM (C, 9)

Information about complications of diabetes is available either in the diabetes register, or otherwise in current aggregated form (e.g. from a practice computer) (C, 9)

Glycaemic control
The records show a recording of HbA1c level in the last 14 months (C, 9)

Where the previous HbA1c was >9, there is a record of HbA1c within the last 7 months (C, 8)

For patients under 75, where the last HbA1c was >9, the records show that the patient has been offered a therapeutic intervention aimed at improving glycaemic control (B,9)

For patients 75 or over, where the last HbA1c was >10, the records show that the patient has been offered a therapeutic intervention aimed at improving glycaemic control (B, 9)

Regular review
The records contain details, recorded within the last 14 months of:

Any comment about general wellbeing (C, 9)

For patients on sulphonylureas, a record of enquiry about symptoms of hypoglycaemia (C, 9)

Recording of peripheral pulses (B, 9)

Recording of vibration sense, or pinprick or monofilament testing (B, 9)

Visual foot examination for ulceration or deformity (B, 9)

Referral to a specialist clinic where there is a foot ulcer (B, 9)

Referral to a chiropodist where feet are "at risk" (age over 65, immobility, nephropathy, retinopathy, < 2 foot pulses) (B, 9)

Serum creatinine (C, 9)

Referral to a specialist where creatinine is >200 mmol/l (B, 9)

A test for proteinuria. (C, 9)

An mid-stream urine where macro-albuminuria is detected (B, 9)

Examination of the fundi through dilated pupils (B, 9)

Referral to an ophthalmologist where there is evidence of significant macular oedema, moderate or severe non-proliferative retinopathy, or any proliferative retinopathy (A, 9)

Visual acuity (B, 9)

Where patients have been diagnosed less than 14 months ago, there should still be a record of fundal examination and visual acuity measurement (B, 9)

Measurement of weight (B, 9)

The patient has been given advice about weight reduction or referred to a dietician where the patient is obese or has a BMI of >27 (B, 9)

Control of hypertension
In patients under 80 years old;

If the average of the last three readings showed a diastolic BP >100, or systolic >150 and diastolic >90, the patient has been offered management for hypertension (B, 8)

In patients over 80 years old,

If the average of the last three readings shown a diastolic BP >110, or systolic >160 and diastolic >100, the patient has been offered management for hypertension (B, 8)

Hypercholesterolaemia
There is a record of cholesterol testing, including total, LDL, and HDL at least once since diagnosis (B, 9)

If the total cholesterol is > 7.8, (or >6.4 in patients who have evidence of ischaemic heart disease), lipid lowering treatment has been offered or intensified.

(B,8 in patients without ischaemic heart disease, B9 where HDL is <1.0 mmol/l. A9 in patients with ischaemic heart disease, B,9 where HDL > 1.0).

Education
The records show that each newly diagnosed patient or carer has been offered education about diabetes management (B, 9)

Home monitoring
The records show whether the patient's glycaemic control is being monitored by urine or blood tests (C, 9)

Where the patient undertakes blood or urine monitoring, there is some record since diagnosis that the technique has been checked (C, 9)

Where the patient undertakes blood or urine monitoring, there is a record of the results of home testing in the past 14 months. (C, 8)

Smoking history
There is a record of smoking status in the notes (B, 9)

In smokers, the records indicate that patients have been given advice on smoking cessation strategies. (B, 9)

Treatment
In patients who have started on ACE inhibitors in the past year, creatinine and potassium have been measured within 1 month of starting treatment. (B, 9)

If the patient is being treated for hypertension, and the patient has proteinuria (macro- but not micro-albuminuria), the patient is on an ACE inhibitor. (B, 8)

Appendix 4: Stable angina
Key
Numbers: indicate median panel score on a nine point ordinal scale for necessity of that aspect of care.

Letters: (A) mainly convincing scientific evidence, usually consisting of one or more randomised controlled trials, limited added expert opinion; (B) about equal parts scientific evidence and expert opinion; (C) Mainly expert opinion, limited scientific evidence.

Initial assessment of angina

Around the time of diagnosis, or within 2 months of diagnosis, the following should be recorded:

Frequency of angina attacks (B, 9)
Exercise capacity (B, 8)
Smoking (A, 9)
Family history of early cardiovascular disease (B, 9)
Current medication (B, 9)
Blood pressure (B, 9)
Peripheral pulses (B, 9)
Auscultation of heart / chest (B, 9)
Weight or body mass index (B, 9)
Resting 12 lead ECG (B, 9)
Total random serum cholesterol (A, 9)
Haemoglobin (C, 9)
Thyroid function tests where there is a clinical suspicion of hyper- or hypo-thyroidism (C, 9)
Blood glucose (B, 9)

Regular review

There should be a record in the previous 14 months of:

Frequency of angina attacks (C, 9)
Exercise capacity (C, 8)
Current medication (B, 9)
Blood pressure (B, 9)
Total random serum cholesterol if previously abnormal (A, 9)

Cholesterol treatment

The records show that the patient was offered cholesterol lowering treatment :

(A) *With no prior myocardial infarction*: All patients under 70 with a serum cholesterol of > 5.5 mmol/l (B, 8.5 for men, B, 8 for women, A, 9 for men and B, 9 for women where cholesterol > 6.5) .

(B) *With a history of previous myocardial infarction*: All patients under 70 with a serum cholesterol of > 5.5 mmol/l (men A, 9; women B, 9), and men over 70 with a serum cholesterol > 6.5. (B,8)

(C) *With a history of diabetes:* All patients under 70 with a serum cholesterol >5.5 mmol/l. (men A, 9, women B,8) and men over 70 if cholesterol > 6.5 (B, 9).

(D) *With a history of diabetes and previous myocardial infarction*: All patients under 70 with cholesterol > 5.5. mmol/l (A, 9) and patients over 70 with a cholesterol of >6.5 mmol/l (men B,9; women B, 8).

Blood pressure treatment

The records should show that the patient has been offered treatment for blood pressure in:

(A) All patients with systolic blood pressure >160 (mixture of A, 9 and B, 9 depending on age, gender, and cholesterol level).

(B) All male patients with cholesterol >5.5 mmol/l and a systolic blood pressure > 140 (B, 8).

(C) All women patients under 70 with a cholesterol > 5.5 mmol/l and a systolic blood pressure >140 (B, 8).

Risk factor recording

For smokers, there should be a record that the patient was offered help to stop smoking (A, 9).

There should be a record that the patient was offered advice on exercise (B, 9).

Drug treatment

The records should show that the patient has been offered treatment with aspirin, in the absence of contraindications (A, 9).

The records should show that the patient has been offered treatment with sublingual glyceryl trinitrate (B, 9).

Where a patient is given regular maintenance treatment, the drug should be a β blocker, in the absence of contraindications to β blockade (B, 8).

Where the records show that a patient is still symptomatic on monotherapy, a second drug should be offered (B, 8).

Records of the following treatments were regarded as inappropriate:

Oral nitrate as regular maintenance treatment without contrindication to β blocker in patients with prior MI (B,2).

Verapamil as second line treatment in patients on β blockers (B,1).

In patients not known to be intolerant of β blockers and not on a β blocker as first line treatment, second line treatment with either a short acting nifedipine (B,1), or a second calcium channel blocker (C,1).

Referral to a cardiologist and for exercise testing

In the absence of major comorbidities *(COPD, dementia, terminal cancer, previous stroke, severe mental illness, neurological disease, and immobility)*, the records show that at onset of diagnosis the patient has been referred for exercise testing or to a cardiologist if:

Any drug has been prescribed for angina, including aspirin and sublingual nitrates (B, 9).

The records show that the patient has been referred to a cardiologist if:

A patient has had a positive exercise tolerance test (A, 9).

The patient has not had revascularisation, has had a negative exercise test, is more than minimally symptomatic, and is on two drug maintenance treatment (B, 8).

The patient has previously been revascularised, has not had an exercise test, is more than minimally symptomatic, and is on two or more maintenance drugs (B, 8).

RECORD OF THE FOLLOWING WAS REGARDED AS INAPPROPRIATE

Patients with angina who are still symptomatic on two drug treatment are offered treatment with a third drug rather than offered referral to a cardiologist (C,1).