New Zealand and United Kingdom experiences with the RAND modified Delphi approach to producing angina and heart failure criteria for quality assessment in general practice

S A Buetow, G D Coster

Abstract

Objectives—**(1) To describe the development of minimum review criteria for the general practice management in New Zealand (NZ) of two chronic diseases: stable angina and systolic heart failure, and (2) to compare the NZ angina criteria with a set produced in Manchester to assess the extent to which use of the same approach to criteria development yields similar criteria.**

Methods—**A modified Delphi approach, based on the RAND consensus panel method, was used to produce minimum criteria for reviewing the recorded management of heart failure and angina in NZ general practice. The criteria for angina were compared with those produced in the UK, including assessment of the extent to which each set describes actions that the other panel agrees are necessary to record.**

Results—**For each condition we report minimum criteria describing actions rated as (a) necessary to record and (b) inappropriate to take but, if taken, necessary to record. Although strong scientific evidence underpins approximately one quarter and one third, respectively, of the final sets of NZ and UK angina criteria for actions necessary to record, the NZ criteria agree strongly with the UK criteria (33 of 39 criteria, 85%) but there is less UK agreement with the NZ angina criteria (28 of 40 criteria, 70%).**

Conclusion—**Despite the lack of scientific evidence for up to three quarters of angina care in general practice, the RAND based approach to criteria development was used in NZ to reproduce most of the UK angina criteria for actions rated as necessary to record in general practice. It is important to make explicit whether ratings of necessity and appropriateness apply to the recording of actions or to the actions themselves.**

(*Quality in Health Care* 2000;**9**:222–231)

Keywords: review criteria; quality of care; general practice; angina; systolic heart failure

Department of General Practice and Primary Health Care, University of Auckland, Auckland, New Zealand
S A Buetow, *research fellow*
G D Coster, *professor of general practice and primary health care*

Correspondence to:
Dr S Buetow
s.buetow@auckland.ac.nz

Review criteria seek "to enable clinicians and others to assess care".[1] More specifically, the Institute of Medicine (IOM) suggests that they are "systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services and outcomes".[2] To permit such assessments, the statements must usually be "suitable for retrospective medical record review of clinical practice"[3] and capable of evaluating key pathways of past care, including guideline implementation.[4]

Although clinicians and others may aim for excellence, review criteria frequently emphasise minimum thresholds of care.[4] Moreover, according to Grimshaw and Russell,[5] they should be "based on mandatory or, at worst, near mandatory elements". Despite the IOM definition of review criteria,[2] it is therefore important that these criteria assess appropriateness *and* necessity in order to show whether inappropriate and necessary care have taken place. Criteria describing appropriate care and unnecessary care are irrelevant to assessing minimum care and identifying service underuse and overuse.

To understand why, we suggest first that only criteria describing necessary care can detect underuse. To be necessary, or crucially important, care must be appropriate and be so likely to benefit patients that failure to offer the care would be improper.[6] Appropriate care maximises net individual health gains (the benefits outweigh the risks) within society's available resources.[7] Since care can be appropriate but not necessary, failure to offer appropriate care is not a sufficient criterion of minimum care and, contrary to Hicks,[8] is not the same as underuse of care. Secondly, we suggest that, to avoid overuse, there is a need for minimum

Table 1 Hypothetical example of how criteria are rated on RAND scales of appropriateness and necessity

Patient with angina is still symptomatic on two drug therapy	Appropriateness of offering third drug	Necessity to offer third drug
1. Not offered referral to a cardiologist	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
2. Offered referral to a cardiologist	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

Table 2 Actual example of how criteria are rated on Manchester scales of appropriateness and necessity

It is **appropriate** and **necessary** for the records to show that the patient with angina, who is still symptomatic on two drug therapy, was offered treatment with a third drug if:	Appropriateness scale	Necessity scale
1. Not offered referral to a cardiologist	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
2. Offered referral to a cardiologist	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

criteria that describe inappropriate care. This is because inappropriate care—where the risks outweigh the benefits[6]—should never be offered. In contrast, care may be unnecessary yet still appropriate.[6 7] In sum, therefore, minimum criteria focus on the inclusionary quality of necessary care and the exclusionary quality of inappropriate care.

The RAND modified Delphi approach (table 1) has been extensively used to produce explicit guidelines for care[9] which have themselves been applied as criteria to assess different degrees of the clinical appropriateness[10 11] and necessity[12] of this care. These criteria have included, but seldom concentrated on, minimum care, although Shekelle *et al*,[13] for example, classified medical procedures as "necessary or not necessary (to evaluate underuse) and inappropriate or not inappropriate (to evaluate overuse)".

Against this backdrop, the RAND-UCLA approach to criteria development has been associated with improved health outcomes for patients.[12 14 15] It has also been reported to offer "one of the only meticulously tested and systematic methods for leavening limited evidence with expert opinion and inference".[16] However, the approach has limitations.[8] Among other things, it has been used mainly in the USA to review indications for clinical procedures in secondary and tertiary care settings, and its rating scales apply to the action under scrutiny, not to recording of the action (table 1).

Other approaches have been used to produce criteria for evaluating primary care for angina[17] and heart failure.[3 18 19] However, the criteria produced, for example, by the former Eli Lilly National Clinical Audit Centre[17 18] have both favourable characteristics[1 20] and their own significant limitations (box 1). The limitations led researchers in Manchester in the UK (including SB) to modify the RAND-UCLA approach to rating criteria in order to produce their own minimum review criteria for the general practice management of stable angina (table 2).[21]

The applicability of these criteria to New Zealand (NZ) general practice is unknown. Also, although 40 of the 44 Manchester criteria for angina describe actions necessary to record, four are based on a scale of appropriateness that is problematic. Although it requires panel-

- Lack of specificity: for patients with heart failure "blood pressure … is in the normal range" and "appropriate advice has been given"; criteria for angina include statements such as "examined for signs of anaemia" and "has had a cardiac examination".
- Criteria ignore potential overprovision of care since there are no "don't do" criteria.
- No published evidence of reliability and validity of approach used to produce criteria.
- Small numbers of criteria limit their comprehensiveness: three of seven "must do" angina criteria refer to actions taken "at diagnosis" so, if the diagnosis is not new, the criteria would say nothing about three-sevenths of current clinical practice.
- Poor relationship between some criteria and preventable morbidity and mortality: for example, angina criteria refer to measuring lipids "at diagnosis" but measuring and *controlling* blood pressure, yet the risk of mortality associated with abnormal lipids is higher than for blood pressure and some anti-angina drugs can control blood pressure.

Box 1 Limitations of the angina and heart failure audit criteria produced by the former Eli Lilly National Clinical Audit Centre.

lists to rate the appropriateness of *recording* an action (table 2), panellists were asked orally to rate the appropriateness of the action itself because an extremely inappropriate action is necessary to record "not inappropriate to record".

This paper therefore has two aims: (1) to describe the development in NZ of minimum criteria for reviewing the general practice management of two chronic diseases: stable angina and systolic heart failure in adults; and (2) to compare the NZ angina criteria with those developed in Manchester to assess the extent to which use of essentially the same modified Delphi method yields similar criteria.[5] We set the latter aim because we were able to use the same necessity rating scale used in Manchester and because the NZ angina panel revised the wording of the appropriateness rating scale without changing the rating task.

Methods

We used a modified Delphi approach based on the RAND consensus panel method (box 2) to produce NZ minimum review criteria for managing heart failure and angina in general practice. The following discussion describes our experience with the four stages of this approach: systematic review of research evidence; generation of preliminary criteria; panel selection; and the rating process. The Manchester study[21] provided a template for the manner in which we defined and implemented these stages. We also note how, subsequent to the development of our heart failure criteria, the

Table 3 Necessity and appropriateness ratings for angina and heart failure by panel

	Manchester (UK)		Auckland (NZ)	
	Round 1	*Round 2*	*Round 1*	*Round 2*
Angina				
Criteria rated for necessity to take and record				
Number of criteria	260	272	332	141
Median 8 or 9 (rated "necessary")	139	120	132	78
Rated "necessary" (%)	54	44	40	55
Final list★	—	39	—	40
Number of criteria in agreement with other set		33		28
Inter-panel agreement (%)†		85		70
Criteria rated for appropriateness of actions (which if taken, are necessary to record)				
Number of criteria	48	51	76	23
Median 1 or 2 (rated "inappropriate")	3	10	25	13
Rated "inappropriate" (%)	6	20	33	57
Final list★		5		3
Number of criteria in agreement with other panel		2		2
Inter-panel agreement (%)†		40		67
Heart failure				
Criteria rated for necessity to record				
Number of criteria	—	—	370	220
Median 8 or 9 (rated "necessary")	—	—	117	77
Rated "necessary" (%)	—	—	32	35
Final list★	—	—	—	56
Criteria rated for appropriateness of actions (which if taken are necessary to record)				
Number of criteria	—	—	156	97
Median 1 or 2 (rated "inappropriate")	—	—	23	11
Rated "inappropriate" (%)	—	—	15	11
Final list★	—	—		3

★The final lists are smaller than the number rated "necessary" or "inappropriate" in round 2 because we were sometimes able to combine very similar criteria.
†The criteria of one panel describe actions that are also necessary to record or inappropriate to take according to the other set of angina criteria.

NZ angina panel clarified the wording of the appropriateness rating task to describe accurately the task completed by all of the panels.

SYSTEMATIC REVIEW OF RESEARCH EVIDENCE
Because ratings of necessity should be based, as far as possible, on panellists' knowledge of the relevant scientific evidence, we first produced a systematic literature review for each condition. We updated the angina review used in the Manchester study, and our reviews for the management of angina and heart failure draw on evidence identified from guidelines,[22] searches of the Internet, two electronic databases (Medline and EMBASE), and reference lists of retrieved articles. Studies published in English from 1990 to June 1998 were assessed for clinical relevance and then against the explicit methodological criteria set by the North of England Guideline Development Group.[22] Independent experts scrutinised the reviews.

GENERATION OF PRELIMINARY CRITERIA
A large number of preliminary criteria were generated from the literature reviews (table 3) and, for angina, from the criteria rated in the second round of the Manchester study. These preliminary criteria were sent to each member of the heart failure and angina panels, respectively, to rate in round 1 of the modified Delphi exercise for each condition. Although the reader has no access to these criteria, panellists had the opportunity to revise them in round 2.

PANEL SELECTION
The heart failure panel comprised six general practitioners (GPs) and four cardiologists, and the angina panel contained five GPs and two cardiologists. Panellists were selected because of their knowledge of, and interest in, heart failure, angina, or both; their community influence (for example, in professional organisations); their ability to represent a range of medical, social, and cultural perspectives; and their diversity of geographical locations within NZ. Approximately three quarters of those invited took part. One GP and one cardiologist did not attend the heart failure meeting and rate criteria for round 2. As in Manchester, we sought no patient input to the generation of the criteria.

RATING PROCESS
The individual panellists were sent the literature review and preliminary criteria for managing heart failure and stable angina, respectively, in general practice. We asked each panellist to rate independently the criteria in a postal

- Can be used to produce guidelines or review criteria
- Uses formal consensus building to integrate scientific evidence and expert opinion
- Involves a multidisciplinary panel of professionals
- Involves an iterative process of literature informed and anonymous ratings, quantitative feedback, and face to face panel discussion
- Gives each panellist equal weight in determining the final result
- Produces specific and comprehensive criteria
- Permits a quantitative description of panellists' final judgements

Box 2 Key characteristics of RAND consensus panel method.

Table 4   Actual example of how criteria are rated on Auckland (NZ) scales of appropriateness and necessity

Action: Offer of third drug to the patient with angina who is still symptomatic on two drug therapy if:	Necessity to record action	Appropriateness of taking action (where action, if taken, is necessary to record)
1. Not offered referral to a cardiologist	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
2. Offered referral to a cardiologist	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

round (round 1) and subsequently to revise and (re)rate the criteria at a face to face panel meeting (round 2). For both conditions, instructions given on the rating sheets in each round repeated the Manchester experience of asking whether "it is appropriate and necessary for the records to show . . ." specified actions.

However, when panellists were asked to assign appropriateness ratings at the UK and NZ panel meetings, they were told orally that it was not the *record* of individual actions they should rate for appropriateness but the *action* to record. The NZ angina panel, which met after the UK study and our heart failure exercise, was the first to notice that this task contradicted the written instruction, and left implicit the necessity to record (in)appropriate actions if they had been taken. This panel therefore modified the scale of appropriateness to assess the appropriateness of actions that, if taken, are necessary to record (table 4).

The revised scale merely makes explicit what UK and NZ panellists actually did. Since the wording of the rating task rather than the task itself was changed, we can compare the small number of appropriateness criteria our panel produced with those misreported in the UK under the heading "Record of the following was regarded as inappropriate".[21]

We only applied our appropriateness scale to the small number of actions that panellists might consider extremely *inappropriate* to take but, if taken, necessary to record. This was because we wanted to produce minimum criteria, and an inappropriate action must never be taken whereas an appropriate action need not always be.[6 7] Meanwhile, panellists mainly rated the necessity to record certain actions (table 4), using the same necessity scale as in Manchester, in order to identify those that are necessary to record as minimum care. In addition, we needed panellists themselves to assign ratings to define minimum care for an average patient with an established diagnosis of the condition under scrutiny.

The rating scales for appropriateness and necessity are ordinal and range from 1 to 9. The necessity scale ranges from 1 (clearly unnecessary to record) to 9 (clearly necessary to record). An action is defined as "necessary" to record when the net benefit to patients is not small and failure to record it would be improper and describe potential underprovision of care. The scale measuring the appropriateness of actions that, if taken, are necessary to record ranges from 1 (extremely inappropriate to take) to 9 (extremely appropriate to take). Inappropriate actions, as noted above, describe potential overprovision of care and offer no expectation of net health benefits. To complement the best scientific evidence available,

panellists were asked to rate criteria on these scales by applying their own clinical judgement rather than their perception of how others might evaluate the criteria, and to include their expectations of variations in patient preferences.

Following our analysis of the postal ratings of preliminary criteria, the seven members of each panel met for one day in Auckland during August 1998 (heart failure) and September 1998 (angina) to discuss, revise, and re-rate the criteria face to face. A computer printout showed panellists their own first round ratings of each criterion in relation to the group distribution. The moderator (SB) knew which panellist(s) had assigned each rating, and used this knowledge to seek to resolve disagreement which occurred when two or more panellists rated a criterion in the top tertile (7–9) and two or more rated it in the bottom tertile (1–3). No attempt was made to force agreement.

Criteria scoring a median of 8 or 9 on the scale of necessity, without disagreement, constitute the final list of actions necessary to take and record. Actions inappropriate to take but, if taken, necessary to record received median ratings of 1 or 2 on the scale of appropriateness, without disagreement. In a third round the cardiologists on each panel assessed whether strong scientific evidence (usually at least one randomised controlled trial) had been published to support the action described by each criterion on the final list.

COMPARISON OF NZ AND UK ANGINA CRITERIA
After describing our review criteria for heart failure and angina, respectively, we compared the extent to which the UK and NZ panels revised preliminary criteria in round 2 and rated them as necessary to record or as inappropriate to take (but, if taken, necessary to record). The extent to which strong scientific evidence supports each final set of angina criteria was then compared. To assess inter-panel agreement we caculated, for each panel in turn, the percentage of its criteria that describe actions which the other panel agrees need recording. We also considered the necessity criteria agreed to by only one panel, rather than both panels.

Results
NZ ANGINA AND HEART FAILURE CRITERIA
Box 3 gives examples of the NZ criteria for reviewing the general practice management of heart failure and angina, respectively. Appendices 1 and 2 list the full sets of criteria.

Table 3 shows that, in round 1, the NZ angina panel rated the necessity to record actions described by 332 criteria. In round 2 this panel reduced the initial set of 332 criteria to 141 and agreed that 78 of these 141 criteria describe actions that are necessary to take and record (median score 8 or 9). A similar process of reduction occurred with the heart failure criteria. The final sets of angina and heart failure necessity criteria contain 40 and 56 review criteria, respectively, because we were sometimes able to combine very similar criteria.

Angina (UK and NZ)

Examples of criteria rated as necessary to record (median score 8 or 9):
- Offer of aspirin in the absence of contraindications.
- Offer of regular symptomatic monotherapy with a β blocker in the absence of contraindications.

Example of a criterion rated as inappropriate to take (median score 1 or 2) but, if taken, necessary to record (median score 8 or 9).
- Combination of a β blocker and verapamil.

Heart failure (NZ)

Examples of criteria rated as necessary to take and record (median score 8 or 9):
- Oral anticoagulant to each patient at high risk of a major thromboembolic event (atrial fibrillation, left ventricular thrombus, previous thromboembolic event).
- Two drug treatment involving an angiotensin converting enzyme (ACE) inhibitor, if not intolerant of an ACE inhibitor, and a diuretic if signs of fluid retention.

Example of a criterion rated as inappropriate to take (median score 1 or 2) but, if taken, necessary to record (median score 8 or 9):
- Offer of monotherapy with a diuretic, nitrate, or cardiac glycoside to each patient with no or mild symptoms of heart failure but ventricular systolic dysfunction; no signs of fluid retention; and no intolerance of an ACE inhibitor.

Box 3   Examples of criteria describing actions necessary to record in NZ, the UK, or both.

Two thirds of the necessity criteria for both conditions describe the initial assessment or annual check up, and drug treatment characterises proportionately more criteria for angina (11/40) than heart failure (8/56). Strong scientific evidence supports 28% (11/40) and 23% (13/56), respectively, of the angina and heart failure criteria describing actions necessary to take and record. Moreover, of criteria describing inappropriate actions, strong scientific evidence supported none of the three angina criteria but two of the three heart failure criteria.

COMPARISON OF NZ AND UK ANGINA CRITERIA

Table 3 also permits comparison of the UK and NZ angina criteria. Over successive rounds necessity ratings were assigned by the UK panel to a slightly increased number of criteria but by the NZ panel to 58% fewer criteria. In round 2 the NZ panel identified proportionately more actions as being necessary to record (median score 8 or 9), but the size of the final list for both panels was almost identical. There is strong scientific evidence to support 35% (10/39) and 28% (11/40), respectively, of the final UK and NZ necessity criteria (using panellists' second round ratings the Manchester group reported a figure of 40% rather than 35%). Nevertheless, there is a high level of NZ agreement with the UK angina criteria describ-

ing actions necessary to record (33/39, 85%); UK agreement with the NZ angina criteria (28/40, 70%) is smaller. The RAND consensus panel approach, as used in NZ, therefore achieved excellent reproducibility of the UK criteria.

For appropriateness ratings the NZ panel was more selective than the UK panel in round 2 in choosing criteria to rate, but was also more likely to rate actions as inappropriate. Because similar criteria could sometimes be combined, the final NZ and UK lists comprise three and five criteria, respectively, describing actions inappropriate to take. Interpanel agreement appeared to be higher for the NZ panel but was unstable because of the small numbers.

AREAS OF UK–NZ DISAGREEMENT

Tables 5 and 6 list criteria agreed to by one panel but not the other. Most of the criteria on which there was disagreement as to necessity to record (table 5) fall into the following areas.

Initial assessment or annual review

Only the NZ panel considered it necessary to record an initial assessment of diet[23][24] and, at an annual review, smoking status if the patient was previously noted to be a current smoker or to have quit smoking within the past six months.[25][26] Based on expert opinion, the NZ but not the UK panel required a record of a previously abnormal body mass index or weight, and the duration of angina attacks. Only in the UK was haemoglobin adjudged necessary to record.

Cholesterol and blood pressure

The NZ panel considered it necessary to record drug treatment in patients with cholesterol levels of ≥5.5 mmol/l after 3–6 months of diet. In contrast, the UK panel favoured a higher cholesterol threshold (>6.5 mmol/l) for recording treatment with a cholesterol lowering drug of patients aged over 70 years with a history of myocardial infarction, diabetes, or both. The panels disagreed on which patient groups need recorded blood pressure treatment for a systolic blood pressure of ≥140 mm Hg.

Symptomatic drug treatment

Unlike the UK panel, the NZ panel recommended recording an offer of three drug treatment for angina patients still symptomatic on two drug treatment. The UK list does not include this criterion presumably because the North of England guideline[22] recommends that patients not adequately controlled on maximal therapeutic doses of two drugs should be referred rather than given a third drug. Table 6 shows that the UK panel even considered it inappropriate to offer treatment with a third drug rather than to offer referral to a cardiologist.

Referrals

There was also disagreement on criteria for referral to a cardiologist and for exercise testing. The NZ panel required recording an offer of referral to a cardiologist for all patients

Table 5   UK-NZ disagreement on actions necessary to record

Actions necessary to record in NZ but not UK	
Initial assessment of angina	Duration of anginal episodes
	Discussion of diet*
Annual review	Duration of anginal episodes
	Auscultation (heart)
	Body mass index or weight, if previously abnormal
	Smoking, if previously noted to be a current smoker or to have quit smoking within the past six months*
Cholesterol treatment	Offer of drug treatment for cholesterol ≥5.5 mmol/l after 3–6 months of dietary therapy*
Blood pressure treatment	Diagnosis of diabetes and, after symptomatic drug treatment, systolic blood pressure ≥140 mm Hg
Regular symptomatic treatment:	
Monotherapy	Verapamil or diltiazem when there is a contraindication to β blockade
Offer as second line treatment	Verapamil, diltiazem or a long acting dihydropyridine if on an oral nitrate and β blockade is contraindicated
Three drug treatment	Offer of third drug to the patient still symptomatic on two drug therapy
Referral to a cardiologist and for exercise testing	Offer of referral to a cardiologist unless not a candidate for revascularisation
Actions necessary to record in UK but not NZ	
Initial assessment	Haemoglobin
Blood pressure treatment	All male patients with cholesterol >5.5 mmol/l and a systolic blood pressure >140 mm Hg
	All women patients under 70 with a cholesterol >5.5 mmol/l and a systolic blood pressure >140 mm Hg
Risk factor recording:	Offer of advice on exercise
Referral to a cardiologist and for exercise testing:	If any drug has been prescribed for angina, including aspirin and sublingual nitrates
	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

*Actions based mainly on strong scientific evidence.

other than those who were not candidates for revascularisation. The UK criteria do not impose this qualification.

The angina criteria common to both studies show good agreement on the strength of the supporting scientific evidence. However, this agreement is independent of the consensus panel method, tending instead to reflect the specificity of the criteria—for example, for cholesterol lowering treatment—and panellists' knowledge of the published scientific literature.

Discussion

MAIN FINDINGS

This paper has described the production of NZ minimum criteria for reviewing the general practice management of stable angina and heart failure. These criteria describe actions rated as necessary to record and actions rated as inappropriate to take but, if taken, necessary to record. They highlight the need to make explicit whether ratings of necessity and appropriateness apply to the recording of actions or to the actions themselves.

We have compared our NZ angina criteria with a UK set to assess the extent to which use of the same approach yields similar criteria. Strong scientific evidence underpins approximately one quarter and one third, respectively, of the final sets of NZ and UK angina criteria describing actions necessary to record. However, the NZ criteria agree strongly with the UK criteria (85%, 33/39). By comparison, UK agreement with the NZ criteria is lower (70%, 28/40).

These results support the internal validity (or quality) of the modified Delphi approach used to produce review criteria in the UK and

NZ, and of the criteria themselves. Moreover, the results on interpanel agreement suggest the generalisability of the UK criteria and, to a lesser extent, the NZ criteria to at least similar general practice settings. This is because scientific evidence operates independently of the health system, and the opinion based criteria tend to be similar because the UK and NZ health systems are alike in many ways. For example, both health systems emphasise the provision by GPs of first contact, comprehensive, family centred care, typically in group practices. Moreover, provider continuity is high in the UK[27] and valued by NZ general practice.[28]

However, differences between the UK and NZ health systems may help to explain why only one panel, more commonly the NZ panel, included certain criteria. For example, funding differences between the two systems may help to explain the different attitudes towards three drug therapy. The fee-for-service system in NZ provides a financial incentive for GPs to offer a third drug before referring to a specialist all patients eligible for revascularisation. In the UK capitation and patient:GP ratios provide no incentive to retain patients. We can also speculate that GPs in NZ have become comfortable over the last decade with, and presumably skilled at, prescribing a third anti-anginal drug, especially to older patients in the absence of restrictions on GP access to these drugs.

Despite such differences at the level of the health system, it is unclear why the NZ panel appeared more likely than the UK panel to hold opinions not shared by the other panel—for example, on the need to record the duration

Table 6   UK and NZ disagreement on actions inappropriate to take but, if taken, necessary to record

Actions inappropriate in NZ but not UK	
Regular symptomatic drug treatment	Offer of short acting nifedipine as regular symptomatic monotherapy or part of two drug therapy
Actions inappropriate in UK but not NZ	
Regular symptomatic drug treatment	Oral nitrate as regular maintenance treatment without contraindication to β blocker in patients with prior myocardial infarction
	In patients not known to be intolerant of β blockers and not on a β blocker as first line treatment, second line treatment with a short acting nifedipine or a second calcium channel blocker
Referral to a cardiologist	Patients symptomatic on two drug treatment who are offered treatment with a third drug rather than offered referral to a cardiologist

of angina attacks. Also apparent is the greater specificity of some UK criteria—for example, on cholesterol treatment—possibly reflecting how the meetings were moderated. Compared with the UK moderator, SB exercised less control over interactions within the NZ panel and was more willing to let it simplify criteria, thus reducing the number of criteria that the researchers themselves combined after the meeting.

STRENGTHS AND LIMITATIONS
Strengths
The RAND modified Delphi approach, on which we based our approach to criteria development, has been shown, mainly in US hospital settings, to be reliable and valid. It offers a systematic and rigorous method of supplementing limited scientific evidence with expert opinion and inference. Before this study SB had experience with using the approach in Manchester and, without changing the rating tasks that the approach requires, we have revised the wording of the appropriateness rating task to make explicit whether ratings apply to the recording of actions or the actions themselves.

Our findings add critical knowledge to the validity of the RAND based approach for quality assessment in primary care outside the USA. As in Manchester, we have shown that this approach is valid in general practice. Our comparison of UK and NZ experiences shows that different panels using the approach can produce similar review criteria for angina in general practice. This finding is consistent with other evidence[13 29] that use of the RAND approach by different panels leads to similar judgements, especially about what should be done. We have also described how our approach was used to produce review criteria for heart failure in general practice.

Limitations
Nevertheless, our study has at least seven limitations. Firstly, the investigators knew the results of the Manchester study in which SB, who moderated the NZ meetings, was involved. However, the NZ panel did not know these results (which had not yet been published) nor did they know that the criteria rated in round 2 in the UK had been used to inform the criteria they were asked to rate in round 1 in NZ. Previous assessments of interpanel reliability have involved panels in rating the same criteria, although from round 1.[13 29]

Secondly, we revised the wording of the appropriateness scale used in the Manchester study. However, we changed the description of the task rather than the task itself, and very few final criteria describe inappropriate actions. Almost all the criteria reported in our study and the UK study result from necessity ratings on a scale that was identical in both studies. Consequently, this paper has focused mainly on the necessity criteria.

Thirdly, we have assumed that the NZ criterion of recording drug treatment for cholesterol levels of ≥5.5 mmol/l in all people does not disagree with the UK panel that it is neces-

sary to record cholesterol treatment for specific patient subgroups such as those under 70 years of age with a serum cholesterol level of >5.5 mmol/l. It would still be necessary for these younger patients to have cholesterol treatment recorded under the NZ criterion. However, an alternative interpretation of the UK criteria is that only the patient subgroups specified need to have their cholesterol treatment recorded. The NZ criterion would then disagree with the seven UK criteria for cholesterol treatment. Overall, the 40 NZ criteria would agree with only 26 of the 39 UK criteria (67%) although, if cholesterol treatment is not considered, agreement with the UK criteria would still exceed 80% (26 of 32). The level of UK agreement with the NZ criteria would be virtually unaffected, with or without reference to cholesterol treatment.

Fourthly, readers may believe that Appendices 1 and 2 omit actions that should be listed. We wish to emphasise that our panels included only actions they considered necessary (crucial), and not merely appropriate, to record in general practice. Panellists described the minimum care to record and criteria most GPs would find acceptable and not "fail".

Fifthly, 12 months elapsed between the UK and NZ angina meetings and new scientific evidence for the management of angina and heart failure has been published since the NZ meeting. However, because all the criteria describe the minimum care that GPs need to record, we believe that these criteria do not easily date and become invalid. For example, there is now convincing scientific evidence[30 31] that, when heart failure is caused by left ventricular systolic dysfunction, β blockers increase the survival benefits of ACE inhibitors, at least in patients aged less than 75 years with mild to moderate heart failure. However, it is not certain that a panel today would require all GPs as part of *minimal* care to implement β blocker treatment for such patients, even with specialist involvement.

Sixthly, most of the criteria lack strong scientific evidence yet some interventions will never undergo experimental evaluation and "it would be unethical to conduct a randomised controlled trial on an aspect of care considered inappropriate by experts".[21] Thus, although it is easy to condemn expert opinion, to produce only criteria for which strong scientific evidence exists is to exempt up to three quarters of care from any comprehensive review or audit. That scientific evidence may be limited or have uncertain external validity in primary care highlights the need systematically to integrate evidence with expert opinion.

Lastly, since some necessary actions may be taken but not recorded, it may be argued that review criteria assess the quality of the recording of care rather than of the care itself. However, in the context of increasing public expectations of, and requirements on, providers to monitor, ensure, and account for the quality of their service delivery, medical records offer the key source of clinical data available for quality assessment.[32] Notwithstanding their limitations,[32] they have medicolegal value, serve as an

aide mémoire of needs and prior care, and facilitate coordinated service delivery by different providers. Moreover, good medical records can at least indicate the quality of care because, for conditions such as heart failure, the quality of recording of measures of process that are sensitive to recording is positively associated with outcomes of care.[33]

Appendix 1: Systolic heart failure

Asterisked criteria describe actions supported by strong scientific evidence (well designed randomised controlled trials, meta-analyses, or systematic reviews). References are available from the authors on request.

ACTIONS NECESSARY TO TAKE AND RECORD
(Median score of 8 or 9 on the scale of necessity in round 2 without disagreement (ratings of 1–3 by two or more panellists)).

Initial assessment (within 2 months of diagnosis of heart failure during past 2 years)
History taking:
- Symptoms at rest or during physical exercise
- Previous heart disease
- Angioplasty, stent, or coronary artery bypass graft★
- Family history of premature heart disease
- Smoking
- Alcohol
- Physical activity★
- Medication(s)
Physical examination:
- Auscultation (heart):
  (a) rate
  (b) rhythm
  (c) sounds
- Auscultation (chest)
- Jugular venous pressure
- Blood pressure
- Abdomen
- Peripheral oedema
- Weight
Blood:
- Glucose
- Thyroid function test if clinical suspicion of thyroid disease
- Complete blood count
- Serum electrolytes
- Serum creatinine
- Fasting total cholesterol fractionated plus triglyceride level
Diagnostic tests:
- Resting 12-lead electrocardiogram
- Offer of an echocardiogram

Ongoing assessment (at least every 3 months during past year)
History taking:
- Symptoms at rest or during physical exercise
- Physical activity★
- Medications
Physical examination:
- Auscultation (heart):
  (a) rate
  (b) rhythm
- Auscultation (chest)
- Blood pressure
- Peripheral oedema
- Weight
Blood (if previously abnormal):
- Complete blood count
- Serum electrolytes
- Serum creatinine
- Fasting total cholesterol fractionated plus triglyceride level

Lifestyle management
Offer of support for:
- Smoking cessation
- Drinking little or no alcohol
- Self-monitoring of weight
- Physical activity within own capabilities
- Vaccination against influenza

Current drug treatment offered
Secondary prophylactic treatment:
- Aspirin to each patient with ischaemic heart failure regardless of whether (s)he is on an ACE inhibitor★
- An oral anticoagulant to each patient at high risk (atrial fibrillation, left ventricular thrombus, previous thromboembolic event) of a major thromboembolic event★
Regular symptomatic treatment:
- ACE inhibitor monotherapy if no signs of fluid retention, no or mild symptoms of heart failure, and not intolerant of ACE inhibitor★
- Two drug treatment involving an ACE inhibitor, if not intolerant of an ACE inhibitor, and a diuretic if signs of fluid retention★
- Three drug treatment
  (a) if still symptomatic on two drug treatment and not offered referral to a cardiologist
  (b) involving a cardiac glycoside or second diuretic if still symptomatic on an ACE inhibitor and a diuretic★
Blood pressure:
- Drug treatment for blood pressure if, after symptomatic drug treatment, systolic blood pressure is ≥140 mm Hg★
Cholesterol:
- Drug treatment for cholesterol to each patient with ischaemic heart disease after ≥3–6 months of dietary therapy according to current guidelines★

Referral to a cardiologist
- Underlying aetiology is unclear in the absence of significant co-morbidity
- Need for prognostic investigation and treatment of concomitant angina in the absence of significant co-morbidity★
- Significant arrhythmias
- Currently more than minimally symptomatic and on three drug treatment
- Rapidly declining renal function

ACTIONS INAPPROPRIATE TO TAKE BUT, IF TAKEN, NECESSARY TO RECORD
(Median score of 1 or 2 on the scale of appropriateness in round 2 without disagreement (ratings of 7–9 by two or more panellists)).

Current drug treatment
Secondary prophylactic treatment:
- Offer of an oral anticoagulant if at low risk of a major thromboembolic event (no atrial fibrillation, no left ventricular thrombus, no previous thromboembolic event)
Regular symptomatic therapy offered:
- Monotherapy with diuretic, nitrate (oral or topical), or cardiac glycoside to each patient with no or mild symptoms of heart failure but ventricular systolic dysfunction; no signs of fluid retention; and no intolerance of an ACE inhibitor★
- Two drug therapy with calcium channel blocker if on a diuretic and not intolerant of an ACE inhibitor★

Appendix 2: Stable angina

Asterisked criteria describe actions supported by strong scientific evidence (well designed randomised controlled trials, meta-analyses, or systematic reviews). References are available from the authors on request.

ACTIONS NECESSARY TO TAKE AND RECORD
(Median score of 8 or 9 on the scale of necessity in round 2 without disagreement (ratings of 1–3 by two or more panellists)).

Initial assessment (within 3 months of diagnosis of angina during past 2 years)
History taking:
- Symptoms:

(a) frequency
(b) duration of anginal episodes
- Physical capacity
- Smoking★
- Medication(s)
- Diet discussed★
- Family history of premature heart disease

Physical examination:
- Auscultation (heart):
  (a) rate
  (b) rhythm
- Auscultation (chest)
- Blood pressure
- Body mass index or weight
- Peripheral pulses

Blood:
- Glucose
- Thyroid function test if clinical suspicion of thyroid disease
- Fasting total cholesterol fractionated plus triglyceride level★

Diagnostic tests:
- Resting 12 lead ECG

One year check up

History taking:
- Symptoms:
  (a) frequency
  (b) duration of anginal episodes
- Physical capacity
- Smoking, if previously noted to be a current smoker or to have quit smoking within past 6 months★
- Medication(s)

Physical examination:
- Auscultation (heart):
  (a) rhythm
- Blood pressure
- Body mass index or weight if previously abnormal

Blood:
- Fasting total cholesterol fractionated plus triglyceride level if previously abnormal★

Exercise testing
- Referral for an exercise tolerance test if no major co-morbidities and no prior revascularisation

Lifestyle management
- Offer to help smokers stop smoking★

Drug treatment offered

Cholesterol:
- Drug treatment for cholesterol levels of ≥5.5 mmol/l after 3–6 months of dietary therapy★

Blood pressure:
- Drug treatment for blood pressure if:
  (a) no diagnosis of diabetes mellitus and, after symptomatic drug treatment, systolic blood pressure ≥160 mm Hg
  (b) diagnosis of diabetes and, after symptomatic drug treatment, systolic blood pressure ≥140 mm Hg

Secondary prophylactic treatment:
- Aspirin in the absence of contraindications★

Initial symptomatic treatment:
- Sublingual glyceryl trinitrate

Regular symptomatic treatment:
- β blocker monotherapy in the absence of contraindications★
- Verapamil or diltiazem when there is a contraindication to β blockade
- Two drug therapy:
  (a) offer of second drug to each patient still symptomatic on monotherapy
  (b) offer as second line treatment β blocker if not prescribed as first line treatment and not contraindicated★
- Verapamil, diltiazem, or a long acting dihydropyridine if on an oral nitrate and β blockade is contraindicated
- Three drug therapy: offer of third drug to patients still symptomatic on two drug therapy

Referral to a cardiologist
- Offer of referral unless not a candidate for revascularisation★

ACTIONS INAPPROPRIATE TO TAKE BUT, IF TAKEN, NECESSARY TO RECORD

(Median score of 1 or 2 on the scale of appropriateness in round 2 without disagreement (ratings of 7–9 by two or more panellists)).

Regular symptomatic treatment
- Offer of short acting nifedipine as regular symptomatic monotherapy or part of two drug therapy
- Combination of a β blocker and verapamil
- Combination of two calcium channel blockers in the absence of β blockade to which the patient is not intolerant.

1 Baker R, Fraser RC. Development of review criteria: linking guidelines and assessment of quality. *BMJ* 1995;**311**:370–3.
2 Field M, Lohr KN, eds. *Institute of Medicine guidelines for clinical practice. From development to use.* Washington DC: National Academy Press, 1992.
3 Hadorn DC, Baker DW, Kamberg CJ, *et al.* Phase II of the AHCPR-sponsored heart failure guideline: translating practice recommendations into review criteria. *J Quality Improvement* 1996;**22**:265–76.
4 Agency for Health Care Policy and Research. *Using clinical practice guidelines to evaluate quality of care.* Rockville, Maryland: US Department of Health and Human Services; 1995.
5 Grimshaw J, Russell I. Achieving health gain through clinical guidelines. I: Developing scientifically valid guidelines. *Quality in Health Care* 1993;**2**:243–8.
6 Kahan JP, Bernstein SJ, Leape LL, *et al.* Measuring the necessity of medical procedures. *Med Care* 1994;**32**:357–65.
7 Buetow SA, Sibbald B, Cantrill JA, *et al.* Appropriateness in health care: application to prescribing. *Soc Sci Med* 1997;**45**:261–71.
8 Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ* 1994;**309**:730–3.
9 Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SA, Siegel RA, eds. *Clinical practice guideline development. Methodology perspectives.* Rockville, Maryland: US Department of Health and Human Services, 1994: 59–70.
10 Kosecoff J, Chassin MR, Fink A, *et al.* Obtaining clinical data on the appropriateness of medical care in community practice. *JAMA* 1987;**258**:2538–42.
11 Kleinman LC, Kosecoff J, Dubois RW, *et al.* The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. *JAMA* 1994;**271**:1250–5.
12 Kravitz RL, Laouri M, Kahan JP, *et al.* Validity of criteria used for detecting underuse of coronary revascularization. *JAMA* 1995;**274**:632–8.
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 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.
15 Shekelle PG, Chassin MR, Park RE. Assessing the predictive validity of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. *Int J Tech Assess Health Care* 1988;**14**:707–27.
16 Naylor D. What is appropriate care? (editorial). *N Engl J Med* 1998;**338**:1918–20.
17 Khunti K, Baker R, Lakhani M. *Management of angina in general practice.* Leicester: Eli Lilly National Clinical Audit Centre, 1995.
18 Eli Lilly National Clinical Audit Centre. *Management of heart failure in primary care.* Leicester: Department of General Practice and Primary Health Care, University of Leicester, 1998.
19 Krumholz HM, Wang Y, Parent EM, *et al.* Quality of care for elderly patients hospitalized for heart failure. *Arch Intern Med* 1997;**157**:2242–7.
20 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.
21 Campbell SM, Roland M, Shekelle P, *et al.* The 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* 1998;**8**:6–15.

22 Centre for Health Services Research. *Evidence-based clinical practice guideline. The primary care management of stable angina*. North of England Evidence-based Guideline Development Project, Newcastle upon Tyne: Centre for Health Services Research, 1996.

23 Stephens NG, Parsons A, Schofield PM, *et al.* Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). *Lancet* 1996;**347**:781–6.

24 Gapinski JP, Van Ruiswyk JV, Heudebert GR, *et al.* Preventing restenosis with fish oils following coronary angioplasty. A meta-analysis. *Arch Intern Med* 1993;**153**:1595–601.

25 Rose G, Colwell L. Randomised controlled trial of anti-smoking advice: final (20 year) results. *J Epidemiol Community Health* 1992;**46**:75–7.

26 Hasdai D, Garratt KN, Grill DE, *et al.* Effect of smoking status on the long term outcome after successful percutaneous coronary revascularisation. *N Engl J Med* 1997;**336**: 755–61.

27 Starfield B. *Primary care. Balancing health needs, services and technology*. New York: Oxford University Press, 1998.

28 Richards JG, ed. *The nature of New Zealand general practice. Occasional Paper Number 1*. Wellington: RNZCGP, 1997.

29 Brook RH, Kosecoff JB, Park RE, *et al.* Diagnosis and treatment of coronary disease: comparison of doctors' attitudes in the USA and the UK. *Lancet* 1988;i:750–3.

30 Cleland JGF, McGowan J, Clark A, *et al.* The evidence for beta-blockers in heart failure. *BMJ* 1999;**318**:824–5.

31 McMurray JJV. Major β-blocker mortality trials in chronic heart failure: a critical review. *Heart* 1999;**82**(Suppl IV):IV14–22.

32 Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;**260**:1743–8.

33 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.

Rapid responses

If you wish to comment on an article published in *QHC*, why not make use of our "rapid response" option?

Log on to our website (www.qualityhealthcare.com), find the paper that interests you, click on "full text" and send your response by email by clicking on "submit a response".

Unless the content seems irrelevant or likely to offend, it will be posted within seven days. You can retrieve it by clicking on "read rapid responses" on our home page.

The editor will decide whether also to publish it in a future paper issue.

It is hoped that "Rapid responses" will enable readers to set up a dialogue about topics in quality of health care.