

ORIGINAL ARTICLE

Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes

S M Campbell, M Hann, J Hacker, A Durie, A Thapar, M O Roland

Qual Saf Health Care 2002;11:125-130

See end of article for authors' affiliations

Correspondence to:
Mr S M Campbell,
National Primary Care
Research and Development
Centre, University of
Manchester, Manchester
M13 9PL, UK;
stephen.campbell@man.ac.uk

Accepted for publication
18 February 2002

Objectives: To field test the reliability, validity, and acceptability of review criteria for angina, asthma, and type 2 diabetes which had been developed by expert panels using a systematic process to combine evidence with expert opinion.

Design: Statistical analysis of data derived from a clinical audit, and postal questionnaire and semi-structured interviews with general practitioners and practice nurses in a representative sample of general practices in England.

Setting: 60 general practices in England.

Main outcome measures: Clinical audit results for angina, asthma, and type 2 diabetes. General practitioner and practice nurse validity ratings from the postal questionnaire.

Results: 54%, 59%, and 70% of relevant criteria rated valid by the expert panels for angina, asthma, and type 2 diabetes, respectively, were found to be usable, valid, reliable, and acceptable for assessing quality of care. General practitioners and practice nurses agreed with panellists that these criteria were valid but not that they should always be recorded in the medical record.

Conclusion: Quality measures derived using expert panels need field testing before they can be considered valid, reliable, and acceptable for use in quality assessment. These findings provide additional evidence that the RAND panel method develops valid and reliable review criteria for assessing clinical quality of care.

Quality of care has been defined by a number of researchers¹⁻⁴ and there are a variety of methods available for its measurement including clinical indicators or review criteria.²⁻⁵⁻¹¹ The UK government has developed sets of clinical indicators for the National Health Service (NHS), most of which are focused on secondary care or public health, but there is an increasing focus on primary care indicators.¹²

McGlynn¹³ set six challenges for measuring quality of care, including establishing credible criteria. While it may never be possible to produce an error-free measure of quality of care,¹⁴ measures should be based on characteristics of best practice such as validity, reliability, and transparency¹⁵⁻¹⁶ and include instruments created for use with medical records.⁵⁻¹⁷⁻²³

Previous work by two of the authors (SC, MR) used expert panels to develop evidence based review criteria for angina, adult asthma, and type 2 diabetes.²⁴ These panels developed 42 criteria for angina, 34 for asthma, and 35 for type 2 diabetes (table 1). These criteria only included aspects of care which the panels judged were necessary to record as well as do,²⁵ thus

enabling medical records to be used to judge quality of care against these criteria. From April 1998 to December 1999 a team from the National Primary Care Research and Development Centre (NPCRDC) undertook a detailed multifactorial quality assessment of a nationally representative sample of 60 randomly selected practices in England.²⁶ This study included a clinical audit using the previously developed clinical review criteria²⁴ and a questionnaire survey and interviews with general practitioners and practice nurses to assess their views about the criteria used.

The aim of the clinical audit, postal questionnaire, and interviews was to assess the validity, reliability, and acceptability of the review criteria developed by the expert panels.²⁴ A review criterion is a systematically developed statement that can be used retrospectively to assess the appropriateness of specific health care decisions, services, and outcomes.⁶⁻²⁷ It relates to a measurable aspect of care that is so clearly defined that it is possible to say whether the element of care it relates to occurred or not.²⁸ Others have described the desirable characteristics of review criteria.²⁹⁻³⁰

Table 1 Criteria rated valid by expert panels and used subsequently in quality assessment

	(a) Criteria rated valid by expert panels ²⁴	(b) Criteria for which data were collected for all patients in the sample	(c) Criteria excluded from analyses because prevalence in sample <1%	(d) Criteria excluded from analyses as inter-rater kappa <0.60	(e) Criteria merged with other criteria for analyses	(f) Criteria used in final analyses
Angina	42	26	10	2	0	14
Adult asthma	34	26	6	3	4	13
Type 2 diabetes	35	31	3	5	4	19

Table 2 Criteria used in the clinical audit

	Frequency met*	95% CI	Sample†
Angina (n=1048)			
Past 14 months, record of:			
Blood pressure	85.1%	82.9 to 87.3	1048
Prescribed or advised to take aspirin unless record of contraindication or intolerance	74.0%	71.3 to 76.6	1048
<i>Prescribed β blocker as maintenance therapy if sole therapy</i>	51.8%	45.6 to 57.9	255
Frequency or pattern of angina attacks	41.5%	38.5 to 44.5	1048
<i>Action taken on blood pressure if systolic blood pressure >160, or >140 if cholesterol level >5.5 mmol/l</i>	37.1%	29.4 to 45.6	139
Exercise capacity	36.3%	33.4 to 39.1	1048
Past 5 years, record of:			
Smoking status	82.3%	80.0 to 84.7	1048
<i>Smoking advice to smokers</i>	72.3%	65.8 to 78.9	181
Cholesterol	62.3%	59.4 to 65.2	1048
<i>Weight advice if overweight</i>	59.3%	54.6 to 64.0	425
Dietary advice	56.9%	53.9 to 59.9	1048
<i>Action taken if cholesterol >5.5 mmol/l</i>	55.2%	47.3 to 62.6	162
Ever recorded:			
Referred for specialist assessment	76.5%	73.9 to 79.1	1048
Referred for an exercise ECG	41.7%	38.7 to 44.7	1048
Asthma (n=1133)			
Past 14 months, record of:			
Record of daily, nocturnal or activity limiting symptoms	40.8%	37.9 to 43.7	1133
Past 5 years, record of:			
<i>Speech rate, pulse rate or respiratory rate during a consultation for an exacerbation of asthma if immediate bronchodilator therapy was used</i>	100%	23.2 to 53.8	39
<i>Oral steroids prescribed if peak flow <60% of normal/predicted</i>	84.1%	75.8 to 93.4	65
Smoking status	80.0%	77.5 to 82.2	1133
Normal or predicted peak flow or record of difficulty using a peak flow meter	74.7%	72.2 to 77.2	1133
<i>Peak flow during a consultation for an exacerbation of asthma</i>	69.2%	64.5 to 73.8	376
<i>Action taken if patient experiencing nocturnal symptoms</i>	64.8%	59.1 to 70.4	278
<i>Smoking advice to smokers</i>	58.4%	52.8 to 64.9	253
<i>Referral to a respiratory physician where oral steroids are used in maintenance treatment</i>	53.8%	34.7 to 73.0	26
<i>Action taken if patient experiencing activity limiting symptoms</i>	50.6%	42.4 to 58.3	153
Inhaler technique	50.1%	47.2 to 53.1	1133
<i>For patients with recorded exercise induced bronchospasm, short acting bronchodilators prescribed for use before exercise</i>	39.2%	32.3 to 46.0	194
<i>Self-management plan for those on high dose steroids or who have had inpatient treatment for asthma</i>	37.3%	30.9 to 43.7	217
Type 2 diabetes (n=1111)			
Past 14 months, record of:			
Blood pressure	92.9%	91.4 to 94.4	1111
HbA _{1c}	87.1%	85.3 to 89.2	1111
Weight	82.2%	79.9 to 84.5	1111
Serum creatinine	78.8%	76.2 to 81.1	1111
Examination of fundi or visual acuity	71.9%	69.3 to 74.6	1111
Urine proteinuria	69.2%	66.4 to 71.9	1111
Recording of peripheral pulses or record of vibration sense	64.6%	61.6 to 67.2	1111
Visual examination of the feet	62.6%	59.7 to 65.3	1111
<i>Record of hypoglycaemia symptoms if patient on sulphonylurea</i>	21.9%	19.0 to 24.7	818
Past 5 years, record of:			
Smoking status	86.5%	84.4 to 88.4	1111
<i>Documentation of education about diabetes if diagnosed <5 years</i>	82.9%	77.6 to 85.6	364
Serum cholesterol	75.3%	72.8 to 77.8	1111
<i>Advice given to smokers</i>	62.8%	56.2 to 70.4	177
Blood pressure: (criteria developed before publication of UKPDS trial ³⁸)			
<i>Under 80 years: Offered treatment if average of last 3 readings shows diastolic >100 or systolic >150 and diastolic >90</i>	64.6%	55.2 to 74.1	99
Treatment (criteria developed before publication of UKPDS trial ³⁸)			
<i>Referral to a specialist where serum creatinine is >200 mmol/l</i>	81.2%	62.1 to 100	16
<i>For patients under 70, where the last HbA_{1c} was >9, patient offered a therapeutic intervention aimed at improving glycaemic control</i>	75.9%	68.9 to 83.3	134
<i>For patients over 70, where the last HbA_{1c} was >10, patient offered a therapeutic intervention aimed at improving glycaemic control</i>	64.3%	58.8 to 89.6	31
<i>If patient is being treated for hypertension and has proteinuria (macro- but not micro-albuminuria), the patient is on an ACE inhibitor</i>	50.4%	46.2 to 54.6	530
<i>If patient was started on ACE inhibitor, creatinine and potassium were measured within 1 month of starting treatment</i>	37.5%	31.7 to 43.3	272

*Frequency with which care was provided and recorded for patients for whom the necessary care was clinically indicated; †number of patients to whom the criterion applied. Conditional variables are shown in italics.

Note: these criteria were devised by expert panels using a systematic process to combine evidence with expert opinion.²⁴

METHODS

Sixty practices in England were selected using a three stage process.²⁶ Three out of the eight English NHS regions were selected to be nationally representative in terms of rurality, socioeconomic deprivation, and geographical population dispersion. From each of these three regions, two health

authorities were selected to be representative of their region in terms of rurality and socioeconomic deprivation. Finally, within each of these six authorities a random sample of 10 practices was selected, stratified in terms of practice size, training status, and socioeconomic deprivation. Where a practice refused to participate, another with similar characteristics

was chosen at random and invited to participate; 60 out of 75 practices approached (80%) agreed to take part.

Sampling

Lists of patients with a confirmed diagnosis of angina, asthma, or type 2 diabetes who were also taking regular medication from a list of the most commonly prescribed drugs for these conditions were generated from computerised records in each practice. Patients had to have been registered with the practice for 2 years to enable sufficient time for “necessary” care to be undertaken. From the lists generated, 20 patients were selected using random numbers, with a further 20 reserves. In some small practices fewer than 20 patients were included as the relevant practice population base was too small. Twenty patients per practice per condition were chosen as it was felt that most practices would have this number of relevant patients for each of the three conditions.

Data abstraction

Data were abstracted for up to 20 patients per condition per practice using standardised data abstraction forms (available from the authors). While the time taken to abstract data depended on the complexity of the patient (volume and density of data) and the quality of the medical records (handwriting and whether notes were summarised), data abstraction took on average 20 minutes per patient.

Timescales

Many criteria rated valid by the expert panels related to care provided in a specific time period such as the last year. In order to make the data abstraction practical, criteria with no time period attached were restricted to care provided in the previous 5 years, with the exception of two angina criteria as noted in table 2.

Analysis

Certain criteria were applicable to all patients, whereas others were applicable only to subgroups—for example, action to be taken if blood pressure exceeded a certain value. Each criterion in table 2 was scored on a 0/1 basis depending on whether necessary care was provided and recorded for individual patients as appropriate.

These binary indicators were analysed using an item response model via the procedure GLLAM6 using the STATA program.^{31 32}

Postal questionnaire

A questionnaire asking respondents to rate the validity of the criteria listed in table 2 was sent to a nurse and doctor in 59 practices; one practice was used as a pilot. Respondents were asked to use an ordinal scale of 1–9 where 9 meant an action was considered necessary and valid for delivering quality of care and 1 meant that it was clearly not necessary and invalid. Criteria rated with an overall median of ≥ 7 without disagreement were considered valid measures of quality of care for angina, adult asthma, and type 2 diabetes. Disagreement existed where 33% or more of the overall ratings were 1–3 and 33% or more were 7–9.³³ Respondents were also asked to state on a “yes/no” basis whether they usually recorded the care relating to the criterion in the medical record.

Interviews

A member of the research team visited 59 practices, with one practice being used as a pilot, and conducted interviews with a general practitioner and practice nurse (usually the senior partner and nurse) who had completed the postal questionnaire, as well as the practice manager. Interviews took place in September to December 1999 and followed a semi-structured schedule. Interviewees had been sent feedback relating to their practice’s clinical audit results. The second part of the interview discussed

Box 1 Examples of review criteria which could not be operationalised

- Whether angina patients had been referred to a cardiologist or for an exercise test at the time of their first prescription at initial diagnosis (initial diagnosis or first prescription may be difficult to identify with certainty).
- Whether a diabetic patient whose feet were “at risk” had been referred to a chiropodist (“at risk” could not be easily operationalised).
- Whether an asthmatic patient on long term maintenance oral steroids had previously had a trial of long acting bronchodilator, high dose inhaled steroids, and one other step 4 British Thoracic Society recommended intervention (this criterion was found to be too complicated and practice prescribing records were often inadequate to extract the data).

respondents’ opinions of the criteria used in the audits, their practice’s results, and possible explanations for low scores.

RESULTS

All 60 practices in the main study took part in the clinical audit. Data were collected for 1048, 1133, and 1111 patients with angina, asthma, and type 2 diabetes, respectively. The mean number of patients included in each practice was 18 (range 6–20) for angina, 19 (range 13–20) for asthma, and 18 (range 9–20) for type 2 diabetes.

Doctors and nurses in 56 of the 60 practices completed the questionnaire (response rate 93%) and 55 practices (92%) took part in the interviews. Three practices refused to take part in both parts of the study, and in some practices staff were unavailable for interview or nurse posts were vacant. A total of 161 interviews were conducted (55 with doctors, 51 with nurses, and 55 with practice managers).

Review criteria rated valid by expert panels that were not be included in the audit

Not all the review criteria rated valid by the expert panels²⁴ were included in the quality assessment (table 1, column a). Firstly, some criteria were practice level criteria—for example, diabetic register—and these were excluded as they would have been the same for all patients and therefore not discriminated between patients; this excluded two asthma and two diabetes criteria. In addition, screening and diagnostic criteria—for example, family history of angina—were excluded as the audit focused on patients with confirmed diagnoses; this excluded a further 13 angina criteria and one asthma criterion. In addition, some criteria could not be operationalised using an audit abstraction sheet so these items were excluded (three angina, five asthma, two diabetes; box 1).

Data on quality of care

Data were subsequently collected for all patients in the relevant samples using 26 criteria for angina, 26 for asthma, and 31 for diabetes (table 1, column b). Data were abstracted from both manual records (including clinic cards and hospital letters) and computerised records.

Prevalence

If a criterion applied to fewer than 1% of the relevant condition sample it was not included in any analyses (table 1, column c). This applied to 10 angina criteria, six asthma criteria, and three diabetes criteria. While this cut off point was arbitrary, criteria relevant to fewer than 1% of a condition sample were prone to clustering within practices. These criteria mostly related to combinations of medication—for example, whether verapamil was used in combination with β blockade as second line treatment for patients with angina.

Box 2 Examples of items with poor inter-rater reliability

Asthma

- Record of whether maintenance treatment was initiated or increased if the patient was experiencing daily symptoms (such symptoms proved difficult to distinguish from nocturnal and activity limiting symptoms).

Angina

- Record of whether the patient had been referred for an exercise test or to a cardiologist if the patient had not had revascularisation but had had a negative exercise test, was more than minimally symptomatic, and on two drug maintenance treatment (this criterion was found to be too complicated and "more than minimally symptomatic" proved difficult to operationalise).

Type 2 diabetes

- Record of general well being (data abstractors could not agree about what constituted a recorded comment about general well being).

Reliability

Inter-rater reliability, a prerequisite for validity, is the level of agreement between different users of an instrument for the same sample.^{34–35} Two raters abstracted data separately (but on the same day) for 7.5% of all patient records included in each of the three samples across 23 of the practices in the main study. Items with a Cohen kappa coefficient of agreement value <0.60 were excluded from the analyses (table 1, column d). This applied to two angina criteria, three asthma criteria, and five diabetes criteria. Values above 0.6 have good to very good agreement.^{36–37} Examples of items excluded because of poor inter-rater reliability are listed in box 2.

Other omissions

Some criteria—for example, diabetes criteria relating to recording of vibration sense and peripheral pulses—were combined by the research team to create a single criterion (table 1, column e).

Measuring quality of care

The denominators for angina, asthma, and diabetes are 26, 22, and 27, respectively, as columns b and e in table 1 refer to criteria excluded for reasons which are not related to the process of development—for example, practice level criteria or due to merging criteria. Consequently, 54%, 59%, and 70% of the angina, adult asthma and type 2 diabetes criteria rated valid by the expert panels had further evidence of face validity, were feasible to apply, and could be applied reliably. These criteria cover a broad spectrum of care including prevention, evaluation, treatment, and referral (table 2).

Of the 14 angina, 13 asthma and 19 diabetes criteria rated valid, nine, four and 10, respectively, were unconditional criteria relevant to all patients in the relevant sample, whereas five, nine, and nine, respectively, were conditional variables only relevant to a patient dependent upon the answer to another question—for example, action taken if blood pressure

Box 3 Necessary criteria not recorded routinely in the medical record by more than 20% of doctors

Angina

- Exercise capacity (28%)
- Referred for specialist opinion (35%) or exercise test (25%) some time since initial diagnosis.
- Frequency of angina attacks (22%)
- Dietary advice (22%)

Asthma

- Speech rate, pulse rate, or respiratory rate during a consultation for an exacerbation of asthma if immediate bronchodilator treatment was used (23%)
- Self-management plan for patients on high dose inhaled steroids (21%)

Type 2 diabetes

- Record of hypoglycaemia symptoms if patient on sulphonylurea (21%)

exceeded a given value. While some criteria were discarded because of low prevalence, the mean number of criteria relevant to individual patients within the three samples were 10 for angina, five for asthma, and 12 for diabetes. This showed that, on average, only one, one, and two conditional variables, respectively, were relevant to patients in the three samples.

Table 2 shows how frequently each criterion was met for all patients for whom that criterion was relevant, for each condition. Investigations and procedures were more often performed and recorded than advice and prevention. Table 3 shows the variation in scores for individual patients in the three samples.

Validity

The doctors and nurses in the questionnaire survey rated all of the criteria listed in table 2 as valid measures of quality of care (overall median ≥ 7 without disagreement). All the criteria listed in table 2 were recorded on a consistent basis according to over 80% of all nurses in the questionnaire survey and, with the exception of the seven criteria listed in box 3, by over 80% of all doctors.

Reasons for poor performance

While the doctors and nurses in the study sample agreed with the expert panels that the criteria used to assess their quality of care were valid, there were differences in opinion between these core staff and the expert panellists.

Firstly, there were examples where doctors and nurses felt confident that necessary care had been provided but that it had not been recorded. Doctors and nurses often described a trade off between time spent recording data and time spent with patients. Commonly cited examples included smoking, exercise, diet, and weight advice. This finding was also supported by the questionnaire survey in terms of general practitioner record keeping (box 3). Moreover, table 2 shows that criteria relating to preventive care and the recording of symptoms were less frequently met than criteria for procedures and investigations.

Table 3 Variation in scores for individual patients

Condition	Mean patient score	Minimum patient score	Maximum patient score	Standard deviation of patient scores	Mean number of items relevant to individual patients	Mean number of items met for individual patients
Angina (n=1048)	60.48	0	100	22.61	10.4	6.4
Asthma (n=1133)	58.73	0	100	30.2	5.3	3.2
Type 2 diabetes (n=1111)	71.28	0	100	21.76	12.32	8.79

Secondly, despite agreeing with panellists that all the criteria in table 2 were valid, some respondents disagreed with panel recommendations that all criteria were necessary for all patients. Patient-centred care was seen to be irreconcilable with clinical guideline and protocol based care. For example, while the expert panels explicitly decided that some criteria should be applied to patients of all ages, some interviewees argued that referral for an exercise test or the importance of cholesterol testing were age specific—for example, less meaningful in patients over the age of 80. This reflected differences in perception by the doctors and nurses interviewed and the expert panellists. We have argued elsewhere that quality of care is at its most meaningful when related to individual patients.³ While the panellists focused upon care for individual patients, in practice there may be a difference between taking part in a consensus method which considers care relevant to an “average” patient with a given condition, and sitting in a consulting room with a patient with complex co-morbidities and personal circumstances.

Other reasons given for poor performance were poor recording by doctors rather than nurses, inadequate or inconsistent information technology, poor data recording templates, insufficient computer training, and poor patient compliance and attendance. However, more frequently, doctors and nurses accepted that poor audit results simply reflected the fact that necessary care had not been taken.

Most staff felt that the audit results for their practice painted an accurate picture of their care for angina, asthma, and diabetes or corroborated other assessments such as those by their local primary care group. However, only a few practices reported at interview that they would change their procedures or protocols as a result of the audit. For example, two practices intended using their results as a baseline as part of bids to become Personal Medical Services pilots and to re-audit their care of angina, asthma, and diabetes. Other staff stated at interview that they had discussed their results within their practice and initiated quality improvement initiatives, including an explicit intention to re-audit care.

DISCUSSION

This study shows that some of the criteria developed previously by expert panels²⁴ were unoperationalisable, unreliable, too rare to be useful, or too hard to extract reliably. This finding emphasises the fact that quality measures need field testing before they can be used in quality assessment. Nevertheless, the expert panels produced review criteria which were found, after extensive field testing, to be valid and reliable. These findings therefore provide additional evidence that the RAND panel method develops valid and reliable review criteria for assessing clinical quality of care. However, the audit showed clearly that many patients are not receiving necessary care.

The collection of audit data for clinical care represents a cornerstone of many current initiatives such as National Service Frameworks and many primary care groups/trusts are collecting audit data from their practices as part of clinical governance initiatives.³⁸ These findings have some important implications for the successful implementation of quality improvement in general practice. Firstly, table 2 shows that, while each of the criteria had been rated necessary by both the expert panels and by general practitioners and practice nurses in the practice sample, the care was frequently not provided to patients who needed it. This is perhaps unsurprising as variation in quality of care is endemic in the UK.^{39–40} However, it shows that there is significant room for improvement in the quality of chronic disease management delivered in general practice in the UK.

Secondly, the dominant approach to quality improvement in the UK over the last decade has been audit.^{41–42} However, others have found that only 24% of audits involved a re-audit

Key messages

- After a rigorous process of development in 60 general practices, it was not possible to operationalise some criteria and others were found to be unreliable, invalid, or too rare to be useful. This suggests that measures of quality developed by expert panels require field testing before they can be used in quality assessment/improvement.
- Over 54% of angina criteria, 59% of adult asthma criteria, and 70% of type 2 diabetes criteria passed through tests of face validity, reliability, feasibility, and acceptability. The criteria covered a broad spectrum of care including prevention, evaluation, treatment, and referral.
- While general practitioners and nurses taking part in the study agreed with the validity of the review criteria, they did not agree with expert panellists that all data items should always be recorded.

to see if care had improved,⁴¹ and that only 35% of audit recommendations are implemented.⁴³ In this study, while most staff felt that their results reflected an accurate picture of their care for angina, asthma, and diabetes, these findings confirm that only a few practices were keen to use the data to improve their care and to re-audit. Those charged with improving quality of care in general practice, particularly primary care groups/trusts, need to motivate practice staff to see the value of auditing and reviewing their care, as many practice staff have a negative attitude towards audit. This will require engaging often suspicious practice staff, as well as cultural and behavioural changes in the attitudes of practice staff.^{44–45} The fact that the criteria used were acceptable to doctors and nurses in a representative sample of practices in England is, however, important as shared understanding and ownership of ideas enhances successful implementation of change.^{46–47}

Thirdly, we found that practices had significantly different levels of computerisation. The increasing availability and comprehensiveness of electronic information systems such as PRODIGY⁴⁸ will foster reliable quality assessment, particularly of clinical data. Data need to be reliable, especially if financial incentives or penalties are to result from quality assessments. However, this will require investment to ensure that all practices have comparable data systems.

Limitations

The review criteria were evidence based and they were developed in 1997. The evidence upon which some of the criteria were based is now out of date. For example, diabetes criteria relating to blood pressure control were developed before publication of the UK PDS study.⁴⁹ This emphasises the importance of updating review criteria and the evidence/literature reviews upon which they are based.

There is some evidence that the quality of record keeping is positively correlated with increased quality of care.^{50–52} However, there has been concern about the validity and reliability of using medical records to assess quality of care^{17–19, 53}; in particular, that data abstraction from records underestimates quality of care because records are not sensitive enough to measure all that goes on in a consultation, especially preventive or counselling/advice activities.^{25–24–36} This limitation was often emphasised by the general practitioners and nurses during interviews. Certainly, while most of the criteria used in this study focused on clinical care, table 2 shows that criteria relating to preventive care and symptoms were less frequently met than those pertaining to investigations/procedures. It is not possible to state how frequently care was given but not recorded or simply not given at all; audit does not distinguish between the two. It is important to emphasise that poor audit results can either reflect poor care or poor recording—a fact accepted by general practitioners and practice nurses. The difference between what the expert panels recommended as

valid review criteria and the views of what doctors and nurses in practices think should always be recorded is an important outcome of the study.

In conclusion, the criteria reported in table 2 provide a workable group of review criteria that could be used by primary care organisations and general practitioners for assessment of the quality of care they deliver to patients with angina, adult asthma, or type 2 diabetes.

ACKNOWLEDGEMENTS

The authors thank the staff of all 60 practices and six health authorities which took part in the study and acknowledge Cath Burns, Dianne Oliver, Nicola Mead and Emma Ruff for their contribution to this project.

Authors' affiliations

S M Campbell, M Hann, J Hacker, A Durie, A Thapar, M O Roland, National Primary Care Research and Development Centre, University of Manchester, Manchester M13 9PL, UK

Funding: This project was funded out of NPCRDC core funding from the Department of Health.

Conflicts of interest: none.

REFERENCES

- Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care* 2000;**12**:281-95.
- Roland MO, Holden J, Campbell S. *Quality assessment for general practice: supporting clinical governance in primary care groups*. Manchester: National Primary Care Research and Development Centre, 1998.
- Campbell SM, Roland MO, Buetow S. Defining quality of care. *Soc Sci Med* 2000;**51**:1611-25.
- Howie JGR, Heaney DJ, Maxwell M. *Measuring quality in general practice*. Occasional paper 75. London: Royal College of General Practitioners, 1996.
- Anderson BG, Noyce JA. Clinical indicators and their role in quality management. *Aust Clin Rev* 1992;**12**:15-21.
- Lawrence M, Olesen F, et al. Indicators of quality in health care. *Eur J Gen Pract* 1997;**3**:103-8.
- Leape LL, Hilborne LH, Schwartz JS, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centres. *Ann Intern Med* 1996;**125**:8-18.
- 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.
- Gray D, Hampton JR, Bernstein SJ, et al. Audit of coronary angiography and bypass surgery. *Lancet* 1990;**335**:1317-20.
- Scott EA, Black N. Appropriateness of cholecystectomy in the United Kingdom: a consensus panel approach. *Gut* 1991;**32**:1066-70.
- MacColl A, Roderick P, Gabbay J, et al. Performance indicators for primary care groups: an evidence based approach. *BMJ* 1998;**317**:1354-60.
- Department of Health. *The NHS plan: a plan for investment. A plan for reform*. Cm 4818-1. London: The Stationery Office, 2000.
- McGlynn EA. Six challenges in measuring the quality of health care. *Health Aff* 1997;**16**:7-21.
- Brook RH, McGlynn EA, Cleary PD. Measuring quality of care. *N Engl J Med* 1996;**335**:966-70.
- Campbell SM, Cantrill JA. Research note: consensus methods in prescribing research. *J Clin Pharm Ther* 2001;**26**:5-14.
- Irvine D. *Managing for quality in general practice*. London: King's Fund Centre, 1990.
- Wu L, Ashton CM. Chart review: a need for reappraisal. *Eval Health Prof* 1997;**20**:146-63.
- Allison JJ, Wall TC, Spettell CM, et al. The art and science of chart review. *Jt Comm J Qual Improv* 2000;**26**:115-36.
- Law M, Ryan B, Townsend E, et al. Criteria mapping: a method of quality assurance. *Am J Occup Ther* 1989;**43**:104-9.
- Huff ED. Comprehensive reliability assessment and comparison of quality indicators and their components. *J Clin Epidemiol* 1997;**50**:1395-404.
- McGlynn EA, Asch SM. Developing a clinical performance measure. *Am J Prev Med* 1998;**14**:14-21.
- Eddy DM. Performance management: problems and solutions. *Health Aff* 1998;**17**:7-25.
- Luck J. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *Am J Med* 2000;**108**:642-9.
- Campbell SM, Roland MO, Shekelle PG, 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. *Qual Health Care* 1999;**8**:6-15.
- Kahan JP, Bernstein SJ, Leape LL, et al. Measuring the necessity of medical procedures. *Med Care* 1994;**32**:357-65.
- Campbell SM, Hann M, Hacker J, et al. Identifying predictors of high quality of care in English general practice: a multi-factorial approach. *BMJ* 2001;**323**:784-7.
- Russell I, Grimshaw J, Wilson B. Scientific and methodological issues in quality assurance. *Proc R Soc Edinb* 1993;**101B**:77-103.
- Donabedian A. *Explorations in quality assessment and monitoring. Volume 2. The criteria and standards of quality*. Ann Arbor, Michigan: Health Administration Press, 1982.
- Hearnshaw HM, Harker RM, Cheater FM, et al. Expert consensus on the desirable characteristics of review criteria for improvement of health quality. *Qual Health Care* 2001;**10**:173-8.
- Baker R, Fraser RC. Development of review criteria: linking guidelines and assessment of quality. *BMJ* 1995;**311**:370-3.
- STATA Corporation. *STATA statistical software. Release 6.0*. College Station, Texas, 1999.
- Rab-Hesketh S, Pickles A, Taylor C. Generalised-linear latent and mixed models. *STATA Technical Bulletin* 2000;**53**:44-57.
- 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.
- Kline P. *The handbook of psychological testing*. London: Routledge, 1993.
- Wilkin D, Hallam L, Doggett MA. *Measures of need and outcome for primary health care*. Oxford: Oxford University Press, 1992.
- Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
- Fleiss JL. The measurement of inter-rater agreement. In: Fleiss JL, ed. *Statistical methods for rates and proportions*. New York: John Wiley, 1981.
- Campbell SM, Roland MO, Wilkin D. Improving the quality of care through clinical governance. *BMJ* 2001;**322**:1580-2.
- Seddou ME, Marshall MN, Campbell SM, et al. Systematic review of studies of clinical care in general practice in the United Kingdom, Australia and New Zealand. *Qual Health Care* 2001;**10**:152-8.
- NHS Executive. *Clinical governance: quality in the new NHS*. London: Department of Health, 1998.
- Gnanalingham J, Gnanalingham MG, Gnanalingham KK. An audit of audits: are we completing the cycle? *J R Soc Med* 2001;**94**:288-9.
- Johnston G, Crombie JK, Davies HTO, et al. Reviewing audit: barriers and facilitating factors for effective clinical audit. *Qual Health Care* 2000;**9**:23-6.
- Balogh R, Bond S. Completing the audit cycle: the outcomes of audits in mental health services. *Int J Qual Health Care* 2001;**13**:135-42.
- Campbell SM, Sheaff R, Sibbald B, et al. Implementing clinical governance in English primary care groups/trusts: reconciling quality improvement and quality assurance. *Qual Saf Health Care* 2002;**11**:9-14.
- Firth-Cozens J. Cultures for improving patient safety through learning: the role of teamwork. *Qual Health Care* 2001;**10**(Suppl II):ii26-31.
- Davies HT, Marshall MN. UK and US health care systems: divided by more than a common language. *Lancet* 2000;**355**:336.
- Davies NT, Nutley SM, Mannion R. Organisational culture and quality of health care. *Qual Health Care* 2000;**9**:111-9.
- Purves IN. PRODIGY: implementing clinical guidance using computers. *Br J Gen Pract* 1998;**48**:1552-3.
- UK Prospective Diabetes Study (UKPDS) Group. Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *Lancet* 1998;**352**:837-53.
- Solomon DH, Schaffer JL, Katz JN, et al. Can history and physical examination be used as markers of quality? An analysis of the initial visit note in musculoskeletal care. *Med Care* 2000;**38**:383-91.
- Koseoff 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**:196-210.
- Kahn KL, Rogers WH, Rubenstein LV, et al. Measuring quality of care with explicit process criteria before and after implementation of a DRG-based prospective payment system. *JAMA* 1990;**264**:1969-73.
- Wyatt JC, Wright P. Design should help use of patients' data. *Lancet* 1998;**352**:1375-8.
- Aaronson LS, Burman ME. Use of health records in research: reliability and validity issues. *Res Nurs Health* 1994;**17**:67-73.
- Stange KC, Zyzanski SJ, Smith TF, et al. How valid are medical records and patient questionnaires for physician profiling and health services research? A comparison with direct observation of patient visits. *Med Care* 1998;**36**:851-67.
- Baker R. Are some topics difficult to audit? *Audit Trends* 1997;**5**:69-70.