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

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 researchers and there are a variety of methods available for its measurement including clinical indicators or review criteria. 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.

McGlynn 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, measures should be based on characteristics of best practice such as validity, reliability, and transparency and include instruments created for use with medical records.

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. 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, 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. This study included a clinical audit using the previously developed clinical review criteria 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. 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. 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. Others have described the desirable characteristics of review criteria.

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

<table>
<thead>
<tr>
<th>Condition</th>
<th>(a) Criteria rated valid by expert panels</th>
<th>(b) Criteria for which data were collected for all patients in the sample</th>
<th>(c) Criteria excluded from analyses because prevalence in sample &lt;1%</th>
<th>(d) Criteria excluded from analyses as inter-rater kappa &lt;0.60</th>
<th>(e) Criteria merged with other criteria for analyses</th>
<th>(f) Criteria used in final analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angina</td>
<td>42</td>
<td>26</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Adult asthma</td>
<td>34</td>
<td>26</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>35</td>
<td>31</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>19</td>
</tr>
</tbody>
</table>
METHODS
Sixty practices in England were selected using a three stage process. 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 recruited.

Table 2 Criteria used in the clinical audit

<table>
<thead>
<tr>
<th></th>
<th>Frequency met*</th>
<th>95% CI</th>
<th>Sample†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Angina (n=1048)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past 14 months, record of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood pressure</td>
<td>85.1%</td>
<td>82.9 to 87.3</td>
<td>1048</td>
</tr>
<tr>
<td>Prescribed or advised to take aspirin unless record of contraindication or intolerance</td>
<td>74.0%</td>
<td>71.3 to 76.6</td>
<td>1048</td>
</tr>
<tr>
<td>Prescribed β blocker as maintenance therapy if sole therapy</td>
<td>51.8%</td>
<td>45.6 to 57.9</td>
<td>255</td>
</tr>
<tr>
<td>Frequency or pattern of angina attacks</td>
<td>41.5%</td>
<td>38.5 to 44.5</td>
<td>1048</td>
</tr>
<tr>
<td>Action taken on blood pressure if systolic blood pressure &gt;160, or &gt;140 if cholesterol level &gt;5.5 mmol/l</td>
<td>37.1%</td>
<td>29.4 to 45.6</td>
<td>139</td>
</tr>
<tr>
<td>Exercise capacity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past 5 years, record of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td>82.3%</td>
<td>80.0 to 84.7</td>
<td>1048</td>
</tr>
<tr>
<td>Smoking advice to smokers</td>
<td>72.3%</td>
<td>65.8 to 78.9</td>
<td>181</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>62.3%</td>
<td>59.4 to 65.2</td>
<td>1048</td>
</tr>
<tr>
<td>Weight advice if overweight</td>
<td>59.3%</td>
<td>54.6 to 64.0</td>
<td>425</td>
</tr>
<tr>
<td>Dietary advice</td>
<td>56.9%</td>
<td>53.9 to 59.9</td>
<td>1048</td>
</tr>
<tr>
<td>Action taken if cholesterol &gt;5.5 mmol/l</td>
<td>55.2%</td>
<td>47.3 to 62.6</td>
<td>162</td>
</tr>
<tr>
<td>Ever recorded:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referred for specialist assessment</td>
<td>76.5%</td>
<td>73.9 to 79.1</td>
<td>1048</td>
</tr>
<tr>
<td>Referred for an exercise ECG</td>
<td>41.7%</td>
<td>38.7 to 44.7</td>
<td>1048</td>
</tr>
<tr>
<td><strong>Asthma (n=1133)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past 14 months, record of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record of daily, nocturnal or activity limiting symptoms</td>
<td>40.8%</td>
<td>37.9 to 43.7</td>
<td>1133</td>
</tr>
<tr>
<td>Past 5 years, record of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech rate, pulse rate or respiratory rate during a consultation for an exacerbation of asthma if immediate bronchodilator therapy was used</td>
<td>100%</td>
<td>23.2 to 53.8</td>
<td>39</td>
</tr>
<tr>
<td>Oral steroids prescribed if peak flow &lt;60% of normal/predicted</td>
<td>84.1%</td>
<td>75.8 to 93.4</td>
<td>65</td>
</tr>
<tr>
<td>Smoking status</td>
<td>80.0%</td>
<td>77.5 to 82.2</td>
<td>1133</td>
</tr>
<tr>
<td>Normal or predicted peak flow or record of difficulty using a peak flow meter</td>
<td>74.7%</td>
<td>72.2 to 77.2</td>
<td>1133</td>
</tr>
<tr>
<td>Peak flow during a consultation for an exacerbation of asthma</td>
<td>69.2%</td>
<td>64.5 to 73.8</td>
<td>376</td>
</tr>
<tr>
<td>Action taken if patient experiencing nocturnal symptoms</td>
<td>64.8%</td>
<td>59.1 to 70.4</td>
<td>278</td>
</tr>
<tr>
<td>Smoking advice to smokers</td>
<td>58.4%</td>
<td>52.8 to 64.9</td>
<td>253</td>
</tr>
<tr>
<td>Referral to a respiratory physician where oral steroids are used in maintenance treatment</td>
<td>53.8%</td>
<td>34.7 to 73.0</td>
<td>26</td>
</tr>
<tr>
<td>Action taken if patient experiencing activity limiting symptoms</td>
<td>50.6%</td>
<td>42.4 to 58.3</td>
<td>153</td>
</tr>
<tr>
<td>Inhaler technique</td>
<td>50.1%</td>
<td>47.2 to 53.1</td>
<td>1133</td>
</tr>
<tr>
<td>For patients with recorded exercise induced bronchospasm, short acting bronchodilators prescribed for use before exercise</td>
<td>39.2%</td>
<td>32.3 to 46.0</td>
<td>194</td>
</tr>
<tr>
<td>Self-management plan for those on high dose steroids or who have had inpatient treatment for asthma</td>
<td>37.3%</td>
<td>30.9 to 43.7</td>
<td>217</td>
</tr>
<tr>
<td><strong>Type 2 diabetes (n=1111)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past 14 months, record of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood pressure</td>
<td>92.9%</td>
<td>91.4 to 94.4</td>
<td>1111</td>
</tr>
<tr>
<td>HbA1c</td>
<td>87.1%</td>
<td>85.3 to 89.2</td>
<td>1111</td>
</tr>
<tr>
<td>Weight</td>
<td>82.2%</td>
<td>79.9 to 84.5</td>
<td>1111</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>78.8%</td>
<td>76.2 to 81.1</td>
<td>1111</td>
</tr>
<tr>
<td>Examination of fundi or visual acuity</td>
<td>71.9%</td>
<td>69.3 to 74.6</td>
<td>1111</td>
</tr>
<tr>
<td>Urine proteinuria</td>
<td>69.2%</td>
<td>66.4 to 71.9</td>
<td>1111</td>
</tr>
<tr>
<td>Recording of peripheral pulses or record of vibration sense</td>
<td>64.6%</td>
<td>61.6 to 67.2</td>
<td>1111</td>
</tr>
<tr>
<td>Visual examination of the feet</td>
<td>62.6%</td>
<td>59.7 to 65.3</td>
<td>1111</td>
</tr>
<tr>
<td>Record of hypoglycaemia symptoms if patient on sulphonylurea</td>
<td>21.9%</td>
<td>19.0 to 24.7</td>
<td>818</td>
</tr>
<tr>
<td>Past 5 years, record of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td>86.5%</td>
<td>84.4 to 88.4</td>
<td>1111</td>
</tr>
<tr>
<td>Documentation of education about diabetes if diagnosed &lt;5 years</td>
<td>82.9%</td>
<td>77.6 to 85.6</td>
<td>364</td>
</tr>
<tr>
<td>Serum cholesterol</td>
<td>75.3%</td>
<td>72.8 to 77.8</td>
<td>1111</td>
</tr>
<tr>
<td>Advice given to smokers</td>
<td>62.8%</td>
<td>56.2 to 70.4</td>
<td>1111</td>
</tr>
<tr>
<td>Blood pressure (criteria developed before publication of UKPDS trial)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 80 years: Offered treatment if average of last 3 readings shows diastolic &gt;100 or systolic &gt;150 and diastolic &gt;90</td>
<td>64.6%</td>
<td>55.2 to 74.1</td>
<td>99</td>
</tr>
<tr>
<td>Treatment (criteria developed before publication of UKPDS trial)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral to a specialist where serum creatinine is &gt;200 mmol/l</td>
<td>81.2%</td>
<td>62.1 to 100</td>
<td>16</td>
</tr>
<tr>
<td>For patients under 70, where the last HbA1c was &gt;9, patient offered a therapeutic intervention aimed at improving glycaemic control</td>
<td>75.9%</td>
<td>68.9 to 83.3</td>
<td>134</td>
</tr>
<tr>
<td>For patients over 70, where the last HbA1c was &gt;10, patient offered a therapeutic intervention aimed at improving glycaemic control</td>
<td>64.3%</td>
<td>58.8 to 86.0</td>
<td>31</td>
</tr>
<tr>
<td>If patient is being treated for hypertension and has proteinuria (macro- but not micro-albuminuria), the patient is on an ACE inhibitor</td>
<td>50.4%</td>
<td>46.2 to 54.6</td>
<td>530</td>
</tr>
<tr>
<td>If patient was started on ACE inhibitor, creatinine and potassium were measured within 1 month of starting treatment</td>
<td>37.5%</td>
<td>31.7 to 43.3</td>
<td>272</td>
</tr>
</tbody>
</table>

*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.
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 GLLAMM6 using the STATA program.41

**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. 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.

**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 panels 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.
good agreement.
and five diabetes criteria. Values above 0.6 have good to very

Reliability
Inter-rater reliability, a prerequisite for validity, is the level of
agreement between different users of an instrument for the
same sample.25 26 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.25 26 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 2, 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 cri-
tera 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 crite-
ria 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
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 condi-
tion. Investigations and procedures were more often per-
formed 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
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 proce-
dures and investigations.

### Table 3 Variation in scores for individual patients

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean patient score</th>
<th>Minimum patient score</th>
<th>Maximum patient score</th>
<th>Standard deviation of patient scores</th>
<th>Mean number of items relevant to individual patients</th>
<th>Mean number of items met for individual patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angina (n=1048)</td>
<td>60.48</td>
<td>0</td>
<td>100</td>
<td>22.61</td>
<td>10.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Asthma (n=1133)</td>
<td>58.73</td>
<td>0</td>
<td>100</td>
<td>30.2</td>
<td>5.3</td>
<td>3.2</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>71.28</td>
<td>0</td>
<td>100</td>
<td>21.76</td>
<td>12.32</td>
<td>8.79</td>
</tr>
</tbody>
</table>
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. 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 panels 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. 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. 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. However, others have found that only 24% of audits involved a re-audit to see if care had improved, and that only 35% of audit recommendations are implemented. 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. 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.

Thirdly, we found that practices had significantly different levels of computerisation. The increasing availability and comprehensiveness of electronic information systems such as PRODIGY 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. 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. However, there has been concern about the validity and reliability of using medical records to assess quality of care. 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. 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

---

**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.
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 NPCDRC core funding from the Department of Health.

Conflicts of interest: none.

REFERENCES

BMJ Quality & Safety

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 and M O Roland

Qual Saf Health Care 2002 11: 125-130
doi: 10.1136/qhc.11.2.125

Updated information and services can be found at:
http://qualitysafety.bmj.com/content/11/2/125

These include:

References
This article cites 42 articles, 16 of which you can access for free at:
http://qualitysafety.bmj.com/content/11/2/125#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/