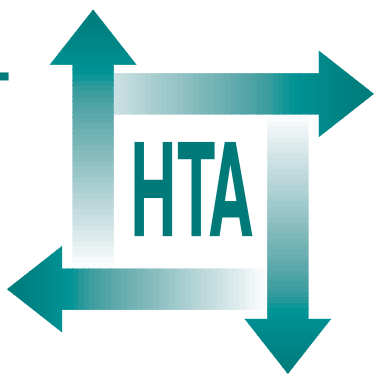
*Health Technology Assessment* 2002; Vol. 6: No. 1

# A study of the methods used to select review criteria for clinical audit

H Hearnshaw
R Harker
F Cheater
R Baker
G Grimshaw

**Health Technology Assessment**
**NHS R&D HTA Programme**





**How to obtain copies of this and other HTA Programme reports.**
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

**Contact details are as follows:**

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

**Payment methods**

*Paying by cheque*
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

*Paying by credit card*
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

*Paying by official purchase order*
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

**How do I get a copy of *HTA on CD*?**

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

---

The website also provides information about the HTA Programme and lists the membership of the various committees.

# A study of the methods used to select review criteria for clinical audit

H Hearnshaw[1*]
R Harker[2]
F Cheater[3]
R Baker[4]
G Grimshaw[5]

[1] Centre for Primary Health Care Studies, University of Warwick, UK
[2] National Children's Bureau, London, UK
[3] School of Health Care Studies, University of Leeds, UK
[4] Clinical Governance Research and Development Unit, University of Leicester, UK
[5] Centre for Health Services Studies, University of Warwick, UK

* Corresponding author

**Competing interests**: none declared

Published January 2002

This report should be referenced as follows:

Hearnshaw H, Harker R, Cheater F, Baker R, Grimshaw G. A study of the methods used to select review criteria for clinical audit. *Health Technol Assess* 2002;**6**(1).

*Health Technology Assessment* is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see opposite).

Part of the work described in this report has also been published as:
Hearnshaw HM, Harker R, Cheater F, Baker R, Grimshaw G. Expert consensus on the desirable characteristics of review criteria for health care quality improvement. *Quality in Health Care* 2001;**10**:173–8.

# NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

Although the National Coordinating Centre for Health Technology Assessment (NCCHTA) commissions research on behalf of the Methodology Programme, it is the Methodology Group that now considers and advises the Methodology Programme Director on the best research projects to pursue.

The research reported in this monograph was funded as project number 95/21/01.

The views expressed in this publication are those of the authors and not necessarily those of the Methodology Programme, HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

**Criteria for inclusion in the HTA monograph series**
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.



The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278



Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

M

# Contents

# List of abbreviations

| | |
|---|---|
| ACQ | audit criteria questionnaire |
| df | degrees of freedom |
| PCAG | primary care audit group |
| SD | standard deviation* |

* Used only in tables

# Executive summary

## Background

Substantial variation is reported in the quality, appropriateness and cost-effectiveness of healthcare services. To reduce such variation, quality improvement initiatives have been actively promoted by many healthcare providers and policy-makers. One potentially powerful method of quality improvement involves establishing the extent to which clinical practice complies with identified criteria.

A possible reason for the incomplete success of activities such as clinical audit stems from the review criteria used. Review criteria have been defined as 'systematically developed statements that can be used to assess the appropriateness of specific healthcare decisions, services and outcomes'. If desirable performance measures are set according to appropriate criteria, the attainment of these targets should result in improved care. In contrast, if quality of care is assessed against inappropriate criteria, attainment of targets may not effect any improvement in care and resources may be wasted in ineffective quality improvement activities

This report describes a programme of research to study the methods used to select the review criteria for clinical audit used in quality improvement activities in the NHS in England and Wales.

## Objectives

- To develop a clear definition of the desirable characteristics of review criteria and their selection.
- To create and use a valid questionnaire to identify the degree to which review criteria that have those characteristics are selected or developed.
- To identify obstacles to the selection or development of review criteria and recommend methods of overcoming such obstacles.
- To advance our understanding of how review criteria for clinical audit are selected.

## Methods

A definition of the important and feasible characteristics of review criteria was created. The definition was developed through an iterative questionnaire process to generate consensus among an international panel of experts in the field of quality improvement in healthcare. Their consensus on the desirable characteristics of review criteria was used to develop a questionnaire to assess how well review criteria were selected or developed. This was then used to measure how well review criteria have been selected or developed for use in clinical audits in the NHS in England and Wales.

After piloting and revisions, the questionnaire was distributed to leads of clinical audits in NHS trusts and general practices. Following the questionnaire study, a sample of respondents was selected for interview. Interviews explored obstacles to using systematic methods to select criteria and methods that had been used to overcome these obstacles successfully.

## Results

The audit criteria questionnaire (ACQ) was created to assess the extent to which systematic methods are used to select review criteria and assess the quality of the review criteria actually used in clinical audit in the UK. The ACQ score was based on the list of desirable characteristics of review criteria derived from expert consensus.

Reported methods of selecting review criteria for clinical audit were often less systematic than is desirable. The mean ACQ score was 0.52 (range 0 to 0.98, $n = 476$) from a possible range of 0 to 1.00.

Seventy-one per cent ($n = 337$) of respondents based their review criteria on the research literature. Of these, 78% used a literature review that was less than 3 years old. Only 27% recorded whether the validity of the research was appraised and 25% recorded the methods used to appraise it. Thus, over 70% of the cases that used evidence as the base for review criteria did not check the validity of the evidence. Furthermore, 29% of respondents had not reported using the research literature to select their review criteria. Only 1% ($n = 3$) of all literature searching respondents used systematic reviews. Of the 305 respondents who

used both literature and expert opinion, 33% ($n$ = 102) reported that the method used to combine evidence and expert opinion was not made explicit. Consultation with colleagues was the most commonly used basis for review criterion selection, as an alternative or supplement to evidence from the literature. However, patients or carers were rarely consulted.

Assessing the validity of review criteria is impeded by the lack of information on how review criteria were developed, even in published audit protocols. The mean score of 0.52 for published review criteria implies that half the desirable characteristics of review criteria are absent. The items were all deemed feasible by expert consensus, and so a perfect score of 1.0 should be possible. Published protocols could improve their development methods, transparency and information on usability.

About half of respondents used audit review criteria that had been piloted. Audits using unpiloted review criteria risk wasting time and resources in discovering that the criteria are unfeasible, contradictory or ambiguous after collecting large amounts of data. Of the respondents, 81% had prioritised their review criteria. Most had used more than one method of prioritisation. Prioritisation according to importance to patients was most often used. Prioritisation according to the quality of the evidence was the choice of the experts, but was used by less than half of respondents.

Creating practical, easy to apply review criteria is more achievable than developing review criteria in a systematic, evidence-based manner.

Clinical and non-clinical audits did not differ significantly on ACQ scores.

The reporting of national or regional audits was extremely rare in the current study. Scores suggested that single organisation audits, the vast majority of audits, were associated with lower ACQ scores than national or regional ones.

The ACQ scores for unpublished review criteria were even lower than for published ones, and 41% of respondents reported using unpublished ones. Thus the review criteria used in many audits do not meet the desirable characteristics of review criteria.

Reports on the process of selection of review criteria should include information on the methods by which they are selected from the literature, consultation with patients and staff, and reference to criteria from previous audits. However, these items are often absent.

There was no difference in the scores for review criteria between audits from general practice and from NHS trusts.

The most commonly noted problems associated with review criteria development focused on organising the audit and gathering literature upon which to base criteria. Some respondents reported that in their particular clinical discipline there was little or no research evidence to guide their practice. Several respondents had difficulty gaining access to the literature through libraries or specialist journals. Some respondents had trouble in narrowing down large review criteria sets to produce a manageable audit protocol.

Although the sample in this study was probably biased, the bias would be towards better ACQ scores. Thus the conclusion that ACQ scores need to be improved is reinforced.

The interview study identified many barriers to using effective, systematic methods of developing review criteria, but was also able to identify ways in which these may be overcome. Levels of skill in literature searching and critical appraisal are important for ensuring that relevant evidence is considered. The organisation has an important role to play in ensuring that adequate training is provided and taken up and that library facilities are of a high standard and individuals are assisted by electronic or staff services in searching for relevant evidence.

To ensure that review criteria are valid, it is essential to have details of the evidence they are based on, the quality of the evidence, the reasons behind any prioritisation and so on. It is important that published audit protocols include a detailed and transparent account of how the review criteria were selected.

## Conclusions

This study has shown that review criteria selections often omit many of the desirable characteristics of review criteria. A significant proportion of review criteria were not based on research evidence. Even where review criteria development did involve reference to research literature, only a limited number of respondents had attempted to assess the quality of the literature, in terms of either its recency or its validity. The higher scores

on usability show that creating practical, easy to apply review criteria is more achievable than developing review criteria in a systematic evidence-based manner. Nevertheless, piloting or providing information on consultation with staff or patients involved were often omitted.

The most commonly noted problems associated with review criteria development focused on organising the audit and gathering literature upon which to base criteria. Audit leads interviewed in this study identified ways in which these barriers may be overcome. Training to enhance levels of skill in literature searching and critical appraisal are important. Furthermore, it is important that all published audit protocols include a detailed and transparent account of how the review criteria were selected, in order that informed choices can be made.

## Implications for further research

There is potential for improving the selection of clinical audit criteria, which will directly and immediately increase the effectiveness of clinical audits.

Recent, high-quality evidence is rarely used to select review criteria. The skills in using the literature to create review criteria are lacking and are difficult to acquire. A national resource of review criteria for clinical audit, which has all the desirable characteristics of review criteria, would overcome some of those difficulties. The criteria would be based on informed assessment of the literature, and kept up to date, with full provenance reported. They would also be based on consultation with patients and experts, both on the importance of criteria and the demands made by collecting the relevant data.

A simple tool with which review criteria could be assessed for quality could be used by those starting an audit, in order to make an informed selection from published criteria. Another use would be for those developing their own criteria, to assess the quality of the criteria they have created.

## Recommendations for further research

- Trials of interventions designed to improve the selection of review criteria for clinical audit. The questionnaire (ACQ) developed in this study could be used as an outcome measure for such trials. One such intervention could be the creation of a library of review criteria that have all the desirable characteristics.
- The development and validation of a simple tool by which review criteria can be assessed. This should be based on the expert consensus view of the desirable characteristics of review criteria.
- Testing the relative effects on the quality of patient care of national or regional audits compared with local audits.
- Case studies of organisations, where the selection of review criteria is given appropriate importance and resources, would identify the organisational policies that enable and maintain this.

# Chapter 1
# Background

Substantial variation is reported in the quality, appropriateness and cost-effectiveness of healthcare services.[1–3] To reduce such variation, quality improvement initiatives designed to improve the consistency and quality of clinical practice have been actively promoted by many healthcare providers and policy-makers.[4–6]

One potentially powerful method of quality improvement involves establishing the extent to which clinical practice complies with identified criteria. The extent of compliance can be used to identify areas where improvements in practice are necessary and to promote beneficial change through feedback to participants.[7,8] Quality improvement exercises such as clinical audits and clinical utilisation reviews follow this pattern of comparing clinical practice against agreed criteria to identify and facilitate areas of change. Unfortunately, such exercises do not always result in the intended improvements in patient care.[9–11]

A possible reason for the incomplete success of activities such as clinical audit stems from the review criteria they use. Review criteria have been defined as "systematically developed statements that can be used to assess the appropriateness of specific healthcare decisions, services and outcomes".[12] If desirable performance measures are set according to appropriate criteria, the attainment of these targets should result in improved care. In contrast, if quality of care is assessed against inappropriate criteria, attainment of targets may not effect any improvement in care and resources may be wasted in ineffective quality improvement activities.[13]

Limited information is available to direct the development and evaluation of review criteria. Although some insight is given into desirable characteristics of review criteria,[12–14] there is no clear indication of how appropriate criteria might be developed. A number of alternative methods are expounded in the literature. Review criteria have been generated from guidelines[15–17] based directly on high-quality research evidence[7,18,19] or drawn from expert opinion.[20,21] Although some authors recommend that criteria should not be developed at all if research evidence is lacking,[7] high-quality research evidence is not readily available for all clinical topics. An alternative approach is to synthesise the two methods, asking experts to identify review criteria from available research evidence, and drawing on expert opinion when evidence is lacking.[22,23]

Without a clear method of defining appropriate review criteria it becomes difficult accurately to appraise the effectiveness of criteria-based quality improvement exercises. Although some guidance is available regarding methods of assessing the utility and effectiveness of clinical audit,[24,25] these publications adopt a generic view of the overall quality improvement process, giving limited attention to specific details, including the selection of review criteria. This is a significant omission if the purpose of such literature is to increase the likelihood that quality improvement exercises will facilitate improvements in care. A clinical audit that follows good procedural practice (e.g. has clear objectives relating to an important clinical issue, and is well managed, organised and documented) may still fail to achieve projected improvements in care if practice has been reviewed against inappropriate criteria.

Little is known about the methods used to select review criteria in UK healthcare quality improvement. Therefore, a major aim of this study was to produce a tested method to assess how well review criteria are actually selected in practice and to determine the extent to which systematically selected, evidence-based criteria are used. This assessment was then used to identify examples where the selection of criteria was not at all systematic. From these examples, possible obstacles to using a systematic approach and recommendations on ways to overcome such obstacles were identified. Examples where a systematic method had been used were also identified and individuals' experience on overcoming obstacles and their views on supportive factors for criteria selection were elicited.

The results from this project should benefit health service providers through the provision of:

- a clear definition of desirable characteristics for quality in review criteria
- a tested method to assess how well audit review criteria have been selected

- improved knowledge of how review criteria are actually selected in practice
- identification of obstacles to using systematically developed, evidence-based criteria and strategies to overcome such obstacles.

The study should therefore identify to what extent current investment in quality improvement, such as clinical audit, is being used in appropriate activities, and provide recommendations on how to promote better practice.

# Chapter 2

# Expert consensus on the desirable characteristics of review criteria

## Introduction

The first stage of the project involved creating a definition of desirable characteristics for quality review criteria. The objectives were to:

- disseminate the definition to a wider audience
- use the definition to create a valid questionnaire to measure the extent to which methods of selecting review criteria reflect such desirable characteristics.

There is no generally accepted method of defining appropriate review criteria. In the absence of clear agreement on an issue, consensus methods can be a useful methodological tool. The Delphi technique[26,27] is a consensus method that gathers expert opinion through an iterative questionnaire process. The method involves researchers communicating, in writing, with a panel of experts comprising between ten and 50 members. Experts are anonymous, to the extent that other panel members do not know their identity at the time of data collection.

It is recommended that the expert panel include both **advocates** and **referees**.[27,28] The expertise of advocates stems from their participant involvement in the area under study (e.g. clinicians, quality managers). The expertise of referees is derived mainly from the study of the topic rather than from direct involvement (e.g. academic researchers). Therefore, advocates could also be called **practitioner experts** and referees **academic experts**.

Initial contact with the expert panel serves to formulate issues and identify options. This preparatory work is used to develop a questionnaire to circulate among the experts. Delphi questionnaires usually ask respondents to rate statements along the dimensions of importance, feasibility, desirability or confidence in making a judgement, with responses being recorded on seven-point rating scales.[28] The researchers collate the experts' responses to each questionnaire and summarised results are fed back to them.[29] Further response is then invited in light of the results of the previous round. The process continues until a previously agreed level of group consensus is reached. Views are unlikely to change after two or three rounds and participation may fall off beyond that.[28] The number of rounds required to reach consensus largely depends on the degree of refinement of the initial questionnaire.[30]

Modifications to a 'pure' Delphi process are fairly common. The preparatory stage of formulating issues can be supplanted by reference to existing research,[31] and subsequent rounds can be used to develop, rather than directly reiterate, the concerns of previous rounds.[28] A modified Delphi technique was suitable to identify the desirable characteristics of review criteria. This information would inform and enable those who develop and select review criteria to make quality improvement in healthcare more effective.

## Method

A two-round modified Delphi technique was used to generate consensus amongst an international panel of experts. A decision was made to restrict the Delphi process to two rounds, since the initial questionnaire was based on a careful review of the available literature. Two rounds were considered sufficient to reach adequate consensus while minimising the workload for participants.

### The expert panel

We identified an international group of experts in quality improvement in healthcare, from a variety of professional disciplines. Three sources of information were used to identify experts: publication records, membership of quality improvement groups in relevant organisations (e.g. Royal Colleges in the UK) and recommendations from researchers in the field. Forty-nine experts were contacted, mostly by email, and asked to contribute to the study. The expert group was categorised by the researchers into 26 academic experts (referees) and 23 practitioner experts (advocates).

### Round 1

MEDLINE and EMBASE were searched from 1990 to March 1999 using the topic headings

- clinical audit
- medical audit
- clinical utilisation
- quality assurance
- guidelines

and the text words

- review criteria
- appropriateness criteria
- clinical indicators
- performance indicators.

The abstract of each citation was reviewed and all studies concerned with the development of review criteria were retrieved. In addition, publications of expert panel members on the development of clinical guidelines were reviewed. Based on the definition of review criteria as "systematically developed statements that can be used to assess the appropriateness of specific healthcare decisions, services and outcomes",[12] the literature review was used to compile a list of identified desirable characteristics of review criteria. From this, the questionnaire for round 1 of the Delphi process was constructed. The questionnaire contained 40 items in three sections:

- the process of developing review criteria (18 items)
- attributes of review criteria (11 items)
- the usability of review criteria (11 items).

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

Round 1 responses were aggregated and used to identify aspects to retain for round 2. A definition of disagreement based on a published appropriateness method[22] was used to exclude items from round 2 if their ratings were polarised to the extreme points of the scale (i.e. if three or more experts gave a high rating of 6 or 7 while, in addition, three or more gave a low rating of 1 or 2). Cumulative percentage scores were then used to determine which of the remaining items met the inclusion criteria of at least 80% of the expert panel providing an importance rating of 5 or more and a feasibility rating of 4 or more. Where experts provided comments, these were carefully considered in project team discussions. Some comments resulted in a refinement of item wording for round 2 of the Delphi process; others led to the inclusion of additional items where experts felt significant omissions arose. The aggregation of round 1 results, and the subsequent development of round 2, occurred over a 2-week period. The round 2 questionnaire was ready 7 weeks after the round 1 questionnaire was sent out.

### Round 2

The round 2 questionnaire informed the experts of the method used to identify items to be included or excluded for round 2. Experts were asked to re-rate each item for round 2, and provide additional comments if they wished. The questionnaire reminded experts of their own round 1 rating for each item, and presented the expert group's mean rating for that item. If the wording of items had been altered, ratings for the original item were provided and the initial wording was shown below the altered item. Some new items were added to section 1 in response to expert comment. These were clearly labelled as being new. All excluded items were shown separately at the end of each section. Experts could alter their ratings for these items and comment on their exclusion, if they wished. The same processes, as in round 1, for distribution, reminding and analysis were used in round 2. Items retained after round 2 identified the desirable characteristics of review criteria and the method of selecting them.

## Results

### Participants

Thirty-eight of the 49 experts invited to take part agreed to do so. The participating experts are listed in appendix 1. The number of experts responding to each round of the Delphi process is shown in *Table 1*. The table also gives details of the number of practitioner and academic experts included in each round. There were no significant differences in the proportion of practitioners and academics responding to the initial participation request ($\chi^2 = 0.3$, $p > 0.05$), and the experts' status as a practitioner or academic did not

***TABLE 1*** *Experts involved in each stage of the Delphi process*

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

***TABLE 2*** *Items included or excluded after each round of the Delphi process*

| | No. of items | Included | Excluded |
|---|---|---|---|
| ***Round 1*** | | | |
| Section 1: development | 18 | 17 | 1 |
| Section 2: attributes | 11 | 7 | 4 |
| Section 3: usability | 11 | 5 | 6 |
| **Total** | 40 | 29 | 11 |
| ***Round 2*** | | | |
| Section 1: development | 19 (2 new) | 14 | 5 |
| Section 2: attributes | 7 | 7 | 0 |
| Section 3: usability | 5 | 5 | 0 |
| **Total** | 31 | 26 | 5 |

relate to their likelihood of completing round 1 ($\chi^2 = 1.5$, degrees of freedom (df) = 1, $p > 0.05$) or round 2 ($\chi^2 = 0.5$, df = 1, $p > 0.05$).

### Included and excluded items

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

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

The round 2 questionnaire also allowed experts to reconsider the 11 items excluded at round 1. The round 2 responses confirmed that all these items should be excluded. In addition a further five items were excluded after round 2 (including one of the items newly introduced at round 2). All the 16 excluded items are shown in *Table 4*, and the reason for exclusion is given. The final lists of included and excluded items (*Tables 3* and *4*) were given to all expert participants. There was no dissent.

In summary, the expert consensus view is that review criteria should be developed through a well-documented process involving:

- consideration of valid research evidence, possibly combined with expert opinion
- prioritisation according to health outcomes and strength of evidence
- pilot testing.

Review criteria should also be accompanied by clear, full information on how they might be used and how data might be collected and interpreted.

## Discussion

The desirable characteristics of review criteria were defined by use of a modified Delphi process. The Delphi method refined and validated the list of characteristics initially based on literature alone. The use of expert judgement enabled us to identify which of the literature-based characteristics of review criteria are both important and feasible.

Our original list of important aspects of review criteria consisted mainly of items mentioned in publications by expert panel members. While the Delphi process confirmed the importance of most of these items, it also excluded some. For instance, although the process retained items such as 'Criteria are based on a systematic review of research evidence' and 'Expert opinion is included in the process of developing review criteria', it excluded specifying the search strategy used or the names of the experts involved. This could reflect the effectiveness of the Delphi process in excluding unimportant or extreme views to arrive at a more centralised, accepted definition. However, the items that survive a Delphi process reflect a compromise position. It is possible that the desire to achieve consensus overrides a resolution of the tensions between conflicting views. New ideas and opinions frequently require time to

***TABLE 3*** *The retained desirable characteristics of the review criteria and the mean importance and feasibility ratings (presented in order of importance)*

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

become widely assimilated or demonstrated as valid. Some of the items excluded by the experts in this study may be innovative items, the importance and feasibility of which are not yet recognised. However, the panel of experts provided both expertise and commitment to achieving the aim of this study. They are unlikely to have been strongly affected by tensions or by unawareness of the value of innovative views.

The value of such collective opinion is dependent on the composition of the expert panel used in the Delphi method. This study aimed to include the views of referees (academic researchers) and advocates (quality improvement practitioners) to balance insight from theoretical understanding with practical experience. However, the eventual composition of our expert panel was marginally biased towards an academic research perspective, as slightly more referees than advocates agreed to participate in the Delphi process. This may have caused practical aspects of review criteria to be underrepresented in the final definition, and could explain the exclusion of some items from the final definition. The expert panel did not include any health economists or patient representatives. Were such individuals included it is possible that the perceived importance of items reflecting resource allocation and patient considerations would be enhanced.

***TABLE 4*** *Excluded items (those not included in the final list of desirable characteristics)*

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

## Summary

A set of desirable characteristics of review criteria emerged from the expert consensus process used in this study, although the problems associated with the nature of consensus, and the structure of our expert panel, should be acknowledged. Nevertheless, this result represents a considerable advance in our understanding of what appropriate review criteria are and how they might be developed. Previous published literature alone did not directly translate into a definitive list of the desirable characteristics of review criteria. Inconsistency was found in the literature on the relative importance that individual researchers assigned to different aspects of review criteria.

Our expert consensus process has provided information on how review criteria can be effectively identified and presented so that data can be collected on justifiable, appropriate and valid aspects of care.

Since review criteria are the statements against which specific aspects of healthcare are assessed, the knowledge gained from this study should be of relevance to all those involved in the development of review criteria for use in quality improvement activities. This can also be used to guide the assessment of review criteria for those who need to decide on which criteria are appropriate for their particular quality assurance or quality improvement project.

# Chapter 3

# Methods of selecting audit review criteria: a survey using the audit criteria questionnaire

## Introduction

The second aim of this study was to operationalise the definition created by expert consensus, as outlined in chapter 2, into a valid questionnaire that could assess how well review criteria are actually selected in practice. Application of this measure would enable us to estimate the degree to which systematically selected evidence-based review criteria are used in actual clinical audit activities. The leaders of a range of ongoing clinical audits, drawn from both primary and secondary care sectors of the NHS, would be invited to complete this questionnaire to provide information on how they had selected their audit review criteria. The questionnaire study also aimed to recruit a sample of clinical audit leads who could be contacted for follow-up interviews.

The data from respondents would enable investigation into the relationships between the quality of audit review criteria and the methods of selection, sources of evidence used, the type of the audit, or the organisational setting. Problems in review criteria selection and potential solutions would also be examined.

## Pilot study

The project team used the list of desirable characteristics for quality in review criteria (see chapter 2) to create a pilot version of the **audit criteria questionnaire** (ACQ pilot). This was used to test the feasibility of the instrument and test the effectiveness of the proposed method of distributing questionnaires to audit leads.

### The questionnaire

In addition to questions based on the desirable characteristics of review criteria, there were questions on who carried out the literature search, to provide information on who was currently doing searches. This would inform how the use of literature could be improved, should it be found necessary. Questions on whether the criteria were prioritised on cost implications or importance to patients were added for completeness of the response options list, to maintain credibility of the questionnaire for respondents. The item of the list of desirable characteristics 'Similar criteria should emerge if other groups review the same evidence' was not included in the ACQ. It was felt that the respondents would not have the knowledge to answer this question accurately and so responses would not add any value to this study.

To confirm the respondent was answering about one specific clinical audit, the title of the audit was requested. For subgroup analyses, information was requested about the source of the criteria (i.e. whether the criteria were drawn from published protocols or guidelines, developed with the help of audit support staff, or developed by the individual audit lead). Opportunity was given for comments about any problems experienced in the development or use of the review criteria and strategies employed to overcome these problems. Finally, they were invited to participate in a follow-up interview study, and to provide their contact details.

Each item which was derived from the list of desirable characteristics was scored. The questions were answered 'Yes', 'No', 'Don't know' or 'Not applicable (N/A)' for each question, as shown in *Figure 1*. Some questions did not have the option of the N/A response, since it was deemed these items should always be applicable. The items had all been deemed feasible by the expert panel, and therefore the maximum score was possible. Further description of the properties of the final version of the questionnaire used in the main study are given in a later section in this chapter (*Scoring the ACQ*, p. 14). The final version of the ACQ is given in appendix 2.

### The method of questionnaire distribution

The initial plan was to use clinical audit coordinators from NHS trusts and primary care audit groups (PCAGs) as intermediaries in distributing questionnaires to audit leads. The audit coordinators were to be contacted and asked to provide a list of the leads for all ongoing audits within their trust or in general practices in a PCAG area. An 'ongoing audit' was defined as

Please tick the relevant box in response to each question

**Section 1: The development of the audit criteria**

| 1.1 Were the audit criteria based upon: | Yes | No | Don't know | N/A |
|---|---|---|---|---|
| a. evidence from the research literature? | | | | |
| b. consultation with experts? | | | | |
| c. consultation with local experts? | | | | |

***FIGURE 1*** *Sample items from the pilot version of the ACQ*

one in which the first data collection has begun, or is complete, but the follow-up data collection has not finished, and it is less than 12 months since the beginning of the first data collection. The researchers would then select a random sample of audit leads from each list. To pilot this method, draft letters were sent to three audit coordinators from NHS trusts and one from a PCAG. The pilot coordinators were asked for comments on whether they would produce positive responses.

Comments suggested that the intended method would be unacceptable to many audit coordinators, either because their area covered so many audits it would be impractical to provide a full list of names and addresses, or because issues of confidentiality prevented them from giving out details. Hence, the option was added that the audit coordinators could themselves select a random sample of ten clinical audit leads in their trust or PCAG area and distribute the questionnaires on the project team's behalf.

## The sample

The sampling frame for this project was all the NHS trusts and PCAGs in England and Wales. A list of NHS trusts was obtained from the NHS Direct website (http://www.nhs50.nhs.uk) and details of PCAGs were provided by the Clinical Governance Research and Development Unit, University of Leicester. Participation request letters for the pilot study were sent to a random sample of ten audit coordinators (eight in trusts, two in PCAGs). The coordinators were asked to complete a response sheet to indicate whether they would be able to participate and, if so, what was their preferred method of questionnaire distribution. If coordinators were unable to participate they were still asked to complete and return a response sheet in order to provide the project team with some indication of why they could not take part. Pre-paid envelopes were provided for the return of completed response sheets.

A total of seven audit coordinators agreed to participate, six of these were from NHS trusts and one was from a PCAG. Of the remaining three coordinators contacted, one declined to take part due to difficulties associated with trust reorganisation, while the remaining two coordinators did not reply and could not be reached by telephone.

## Questionnaire distribution

Of the seven participating coordinators, only one, from an NHS trust, provided a list of all ongoing audits. In this case the project team sent questionnaires directly to a random sample of ten audit leads drawn from the list provided. The remaining six coordinators (five NHS trusts, one PCAG) preferred to distribute the questionnaires themselves, on behalf of the project team. In NHS trusts, ten questionnaire packs (each containing a questionnaire, covering letter and freepost return envelope) were sent to the audit coordinators who agreed to pass these on to audit leads. Each questionnaire was given a unique code number and coordinators were given a form on which to keep a record of the recipient of each questionnaire. Maintaining an accurate record of who received a questionnaire was important for follow-up purposes.

The PCAG coordinator agreed to distribute a questionnaire pack to each of the 125 practices in the area covered by that PCAG. Although questionnaires distributed via PCAGs were given an identifying code number, PCAG coordinators were not asked to maintain a record of questionnaire recipients. This decision was based on the considerations that: keeping a distribution record would be an onerous task for PCAG coordinators due to the large number of practices involved; and the number of responses needed from the PCAG would be reached without recourse to reminders.

In total, 185 pilot questionnaires were sent out, 60 to NHS trusts and 125 to general practices. Reminders for non-responders were organised 3 weeks after the distribution of the questionnaire. In the one trust that provided a list of audits, postal reminders were sent directly to non-responding audit leads. In the remaining five trusts audit coordinators were contacted and asked to chase up non-responders on behalf of the project team.

### Follow-up interviews

Sixteen pilot respondents indicated that they would be willing to be contacted for a brief follow-up interview. A random selection of six of these volunteers were contacted by telephone and asked for their opinion on the questionnaire study.

### Results

A total of 37 pilot questionnaires were returned, a response rate of 20%. The low response rate alerted us to the need for revision of the distribution and follow-up methods. Respondents answered between 70% and 100% of the questions, suggesting that the questionnaire was feasible and acceptable. Interviews showed that the questionnaire was not seen as onerous to complete, taking a reported average of 5 minutes to complete, and was considered clear and straightforward.

Improvements to the questionnaire design were suggested. Two changes were accepted. First, an intermediary response option 'Partly' was requested as the yes/no dichotomy was seen as too rigid. Secondly, it was felt that details of the scope of the audit (e.g. national, regional, single organisation) would enable comparison across these different audits.

Information drawn from pilot follow-up interviews also led to changes to the covering letter that accompanied the questionnaire. A date by which completed questionnaires were to be returned and the expected time to complete the questionnaire (5 minutes) were included.

## Main study

The various alterations to the questionnaire and covering letter, as well as to the study design regarding the method of questionnaire distribution, were incorporated in the main questionnaire study. The final version of the ACQ is given in appendix 2.

### Participant recruitment

Questionnaires were completed by the lead people in ongoing audits in a random sample of NHS trusts and PCAGs. As in the pilot study, audit leads were contacted via the audit coordinators (or equivalent posts) in each of these organisations. Participation request letters, for both PCAGs (appendix 3) and NHS trusts (appendix 4), offered audit coordinators a choice of two ways to assist with the project.

PCAG coordinators could either:

- provide the project team with a list of all practices attached to their audit group, giving the name and contact details of someone involved in audit, and the project team would then write to each individual asking them to complete a questionnaire; or
- agree to distribute questionnaires to each practice on behalf of the project team.

NHS trust coordinators could either:

- provide the project team with a list of all ongoing audits in the organisation, so that the team could select a random sample of ten audit leads to send questionnaires to; or
- agree to give out questionnaires on behalf of the project team to the lead people in ten ongoing audits in the trust.

The definition of an ongoing audit used in the pilot study was retained (i.e. an ongoing audit was defined as one in which the first data collection has begun, or is complete but the follow-up data collection has not finished, and it is less than 12 months since the beginning of the first data collection). Thus, responses would be related to recent, salient activity, and therefore were likely to be accurate.

Audit coordinators were provided with a response sheet to indicate whether they would be able to participate and, if so, what was their preferred method of questionnaire distribution. (For PCAG coordinator forms, see appendix 5; for NHS trust coordinator forms, see appendix 6). Coordinators

who were unable to participate were nonetheless asked to complete and return the response sheet to indicate why they could not take part. Pre-paid, addressed envelopes were provided for the return of completed response sheets.

## The sample

A sample for the main study of 70 audit coordinators in NHS trusts and 20 in PCAGs, each providing responses from five out of ten audit leads invited, would provide a sample of 450. This would provide a power of at least 80% ($\alpha = 0.05$) for expected response values of 50%, to within 4%. However, the pilot study revealed that audit coordinators in PCAGs preferred to distribute questionnaires to all practices in their area. This allowed them to avoid differentiating between practices engaged in audit and those that were not. If such a preference were evident in the main study, a large number of questionnaires (typically over 50) would need to be sent to each PCAG. Hence, a decision was taken to reduce the number of PCAG coordinators included in the study from 20 to ten. Therefore, a new target of recruiting 70 coordinators in NHS trusts and ten in PCAGs was set. An expected response rate from audit coordinators of 50%, meant that 140 trusts and 20 PCAGs needed to be contacted. The sampling frame for the main study was all PCAGs ($n = 118$) and NHS trusts ($n = 391$) in England and Wales.

### *PCAGs*

A database of all PCAGs was obtained from the Clinical Governance Research and Development Unit, at the University of Leicester, from which a random sample of 20 PCAGs was selected.

### *NHS trusts*

A database of NHS trusts was obtained from the NHS Direct website (http://www.nhs50.nhs.uk), with information about the type of trust being derived from the Department of Health website (http://www.doh.gov.uk/tables98). Both sets of information were incorporated in an SPSS® file. This enabled stratified sampling, reflecting the distribution of trust types in England and Wales (acute, mental health, community, etc., and combinations of these). The trusts were divided, according to the type of service they provided, into eight blocks. *Table 5* shows the frequency of NHS trusts in each category and the percentage of overall trusts this represents. The percentages were used to calculate the frequency of trusts of each type needed for a sample of 140 trusts to reflect the actual incidence of trust type. Once the necessary frequencies were determined, the SPSS was used to select randomly the appropriate quota of trust in each block, arriving at a total of 140 NHS trusts.

***TABLE 5*** *The number of NHS trusts according to the type of service in the NHS and in the sample: recruitment round 1*

| Type of NHS trust | Total | Sample frequency |
|---|---|---|
| Acute | 130 (33%) | 46 (33%) |
| Acute and community | 18 (5%) | 7 (5%) |
| Acute and mental | 13 (3%) | 4 (3%) |
| Acute, community and mental | 67 (17%) | 24 (17%) |
| Community | 16 (4%) | 6 (4%) |
| Community and mental | 107 (27%) | 38 (27%) |
| Mental | 15 (4%) | 6 (4%) |
| Teaching and acute | 25 (6%) | 9 (6%) |
| **Total** | 391 (100%) | 140 (100%) |

### *Recruitment round 1*

Invitation letters were sent to audit coordinators in 140 NHS trusts and 20 PCAGs in April 1999, with reminders for non-responders sent after 1 month. After 6 weeks, 73 replies (47 positive, 26 negative) were received from coordinators in NHS trusts and 12 (six positive, six negative) from PCAG coordinators. This meant that a further 23 trusts and four PCAGs were needed to meet the target of 70 participating NHS trusts and ten PCAGs. Further random selection of NHS trusts and PCAGs was necessary to reach target levels.

### *Recruitment round 2*

Invitations were sent to a further 50 randomly selected NHS trust audit coordinators and ten PCAG coordinators, to reach a total of 50 NHS trusts (*Table 6*). The additional NHS trusts and PCAGs increased the sampling frame to 190 NHS

***TABLE 6*** *The number of NHS trusts according to the type of service in the NHS and in the sample: recruitment round 2*

| Type of NHS trust | Total | Sample frequency |
|---|---|---|
| Acute | 130 (33%) | 17 |
| Acute and community | 18 (5%) | 3 |
| Acute and mental | 13 (3%) | 1 |
| Acute, community and mental | 67 (17%) | 9 |
| Community | 16 (4%) | 2 |
| Community and mental | 107 (27%) | 14 |
| Mental | 15 (4%) | 1 |
| Teaching and acute | 25 (6%) | 3 |
| **Total** | 391 (100%) | 50 |

trusts and 30 PCAGs. Extending the sample in this manner proved worthwhile, as 4 weeks later the sample targets were met. From the 190 NHS trusts contacted 122 responses were received (73 positive, 49 negative), and from the 30 PCAGs contacted 18 responses were received (ten positive, eight negative).

### *Recruitment round 3*

Although it was expected that recruitment of 70 NHS trusts and ten PCAGs should result in 500 completed questionnaires being returned, the response rate from audit leads within these organisations was very low, despite reminders. Therefore, at a later stage of the questionnaire study a decision was taken to contact a further 20 NHS trusts and five PCAGs in order to recruit a further ten trusts and two PCAGs (*Table 7*).

***TABLE 7*** *The number of NHS trusts according to the type of service in the NHS and in the sample: recruitment round 3*

| Type of NHS trust | Total | Sample frequency |
|---|---|---|
| Acute | 130 (33%) | 7 |
| Acute and community | 18 (5%) | 1 |
| Acute and mental | 13 (3%) | 1 |
| Acute, community and mental | 67 (17%) | 3 |
| Community | 16 (4%) | 1 |
| Community and mental | 107 (27%) | 15 |
| Mental | 15 (4%) | 1 |
| Teaching and acute | 25 (6%) | 1 |
| **Total** | 391 (100%) | 20 |

The additional NHS trusts and PCAGs increased the sampling frame to 210 NHS trusts and 35 PCAGs, while maintaining the distribution across types of trust. This resulted in a total of 139 responses from audit coordinators in NHS trusts (83 positive, 56 negative), while in PCAGs 21 coordinators provided responses (11 positive, ten negative).

## Reasons for non-participation

Audit coordinators who were unwilling to participate in the study were asked to complete a response sheet indicating the reasons underpinning their decision. Coordinators were asked to indicate whether they could not participate because: they could not spare the time to contact audit leads, they did not believe audit leads in their area could spare the time to take part, they were not provided with enough information to make an informed decision, or other reasons (to be specified). The frequency of responses provided by PCAG and NHS trust coordinators is given in *Table 8*.

***TABLE 8*** *Frequency of PCAG and NHS trust audit coordinators' reasons for non-participation*

| Reason for non-participation | PCAG coordinators | NHS trust coordinators |
|---|---|---|
| (a) Not enough time to contact audit leads | 2 | 18 |
| (b) Audit leads would not have time to participate | 7 | 32 |
| (c) Insufficient information for informed decision | 1 | 6 |
| (d) Other | 7 | 27 |
| Failed to reply | 14 | 71 |

It can be seen from *Table 8* that, in addition to the suggested reasons given by the researchers, seven PCAG coordinators and 27 NHS trust coordinators gave 'Other' reasons for non-participation. The seven PCAG coordinators all stated that the magnitude of recent changes in primary care meant it was inappropriate to take up any additional practice time for audit leads. The reasons provided by the 27 audit coordinators in NHS trusts are shown in *Table 9*. All the trust audit coordinators who mentioned 'Other' reasons had also ticked one of the response options (a), (b) or (c) (see *Table 8*).

***TABLE 9*** *Frequency of other reasons for non-participation given by audit coordinators in NHS trusts*

| Reason for non-participation | Frequency |
|---|---|
| Trust undergoing reorganisation | 10 |
| Not enough audits meet study inclusion criteria | 9 |
| Audit department understaffed | 5 |
| No central register of audit activity kept | 2 |
| Project has no obvious benefit to trust | 1 |

## Questionnaire distribution

The first wave of questionnaire distribution began in June 1999, when questionnaires for audit leads were sent to 47 NHS trusts and six PCAGs. A total of 801 questionnaires were distributed at this stage. The remaining questionnaires were distributed as soon as agreement to participate was received from NHS trusts or PCAGs belonging to the second and third rounds of recruitment.

Of the 83 audit coordinators in NHS trusts who eventually agreed to take part, 22 provided a list of audit leads for all ongoing audits and 61 preferred to act as an intermediary and distribute ten questionnaires on behalf of the project team. In the PCAGs, only one coordinator provided a list of audit leads, the remaining ten preferring to distribute questionnaires to all practices in their area on behalf of the project team.

Where a list of audit leads was provided, the project team selected ten individuals at random from each organisation, and questionnaires were mailed directly, along with covering letters and freepost envelopes. Each questionnaire was given a unique code number so a record of respondents could be kept. In organisations where coordinators were to pass on questionnaires, the appropriate number of questionnaire packs (each containing a questionnaire, covering letter and freepost envelope) were forwarded. All questionnaires were given a code number, and trust coordinators were also given a form to record the recipient of each questionnaire for follow-up purposes. The coordinators were also provided with a copy of the questionnaire for their own reference. In total, 1384 questionnaires were sent out, ten in each of 83 NHS trusts and 554 to all general practices in 11 PCAG areas.

Reminders for non-responders were organised, using the same delivery methods, 3 weeks after the distribution of the questionnaire. Subsequent reminders were sent after 6 weeks.

## Results

All (100%) of the 83 NHS trusts and 11 PCAGs, provided at least one completed questionnaire response. Audit leads in NHS trusts returned 391 questionnaires (47% of 830), and audit leads in general practices returned 85 (15% of 554) questionnaires.

The ACQ was scored to produce global indices of how well review criteria had been developed or selected and how easy they were to use in an actual audit. Only the questions directly related to the list of desirable characteristics of review criteria, from the expert consensus (see chapter 2), were used to contribute to this score.

### Scoring the ACQ

Scores were calculated for all the questions that related directly to an item on the list of desirable characteristics developed earlier (see chapter 2). These questions are listed in *Box 1*. A score of 1 was given to each 'Yes' response, 0.5 for 'Partly' and 0 for 'No'. The response option 'Don't know' was also scored 0, because if the information was not known it could not contribute to the review criteria selection. A single ACQ score was calculated by summing the scores for each item and, because for some items the response option 'Not applicable' was available, the score was then divided by the number of applicable questions. Each item was thus given equal weighting, to reflect the high importance given to each item by the expert group. This provided an index between 0 and 1, with 1 being a perfect score of 'Yes' to all questions and 0 indicating that all items yielded the response 'No'. Thus the ACQ score indicated the degree to which the review criteria met the requirements to be deemed fit for their purpose.

#### *Properties of the ACQ*

**Internal consistency**

Using the data from the main study, Cronbach's $\alpha$ coefficients were calculated to assess the internal consistency of the questionnaire. For the 25 items, $\alpha = 0.78$, an acceptable value, demonstrating internal consistency. In addition, $\alpha$ values were calculated for the two subscales of 'development' (the first 14 questions), for which $\alpha = 0.75$, and 'usability' (the last 11 questions), for which $\alpha = 0.64$. These values suggest that the subscales were each also internally consistent. These subscales were derived from the layout of the questionnaire, reflecting the categorisation of items used in the consensus process. The subscales were used for selected analyses of the study data in order to identify more accurately where desirable characteristics were lacking, and to link the quantitative data findings with the qualitative data. All items were answered by at least one respondent.

**Content validity**

The aspects of review criteria contained in the questionnaire were based on a list validated by expert opinion, and thus had high content validity.

**Criterion validity**

Criterion validity was assessed by testing whether the scores for published review criteria, which should be of the highest available standard, were higher than the scores for unpublished review criteria. This should be particularly true for the questions related to development. Mean scores (*Table 10*) suggested that the instrument did generate higher scores for published than for unpublished review criteria.

**BOX 1 The questions from the ACQ that were used to calculate the ACQ score**

**1.1** Were the audit criteria based on:
(a) Searching the research literature?
(b) Consultation with experts?
(e) Consultation with patients or carers?
(f) Criteria used in previous audits?

**1.2** (d) How up to date was the literature review?

Was the following information recorded (by you or the authors of the review):
(f) The sources/databases used to identify the literature?
(g) Whether the validity of the research was appraised?
(h) The methods used to assess validity?

**1.3** Is the method of combining evidence from the literature and expert opinion made explicit?

**1.4** Is the method used to select the audit criteria described in enough detail to be repeated?

**1.5** Were the audit criteria pilot tested for practical feasibility?

**1.6** Were the audit criteria prioritised on:
(a) Impact on health outcome?
(b) Quality of supporting evidence?

**1.7** Were the relative values of harms and benefits associated with treatment options considered in selecting criteria?

**2.1** Do the criteria:
(a) State the patient populations to which they apply?
(b) State the clinical settings to which they apply?
(c) Give clear definitions of the variables to be measured?
(d) Use unambiguous terms?

**2.2** Are the criteria linked to improving health outcomes (rather than, say, to reducing costs or increasing throughput)?

**2.3** Do the criteria enable you to differentiate between appropriate and inappropriate care?

**3.1** Did the criteria have information on:
(d) How the demands of the audit on patients might be minimised?
(e) How the demands of the audit on staff might be minimised?

**3.2** Did the criteria have clear instructions for using them?

**3.3** Were patients consulted about the acceptability of these criteria for them?

**3.4** Were all relevant staff consulted about the acceptability of these criteria for them?

***TABLE 10*** *Mean ACQ subscale scores for published and unpublished criteria*

| ACQ subscale | Published criteria | | Unpublished criteria | |
|---|---|---|---|---|
| | Mean ACQ score | No. of responses | Mean ACQ score | No. of responses |
| Development | 0.51 | 83 | 0.39 | 390 |
| Usability | 0.69 | 82 | 0.68 | 387 |
| **Total ACQ** | 0.59 | 81 | 0.50 | 385 |

The difference between scores for published and unpublished criteria was significant for the total ACQ score ($F(1, 464) = 16.88$, $p = 0.000$) and the development subscale ($F(1, 471) = 29.2$, $p = 0.000$). However, the usability subscale showed no significant difference ($F(1, 467) = 0.18$, $p = 0.669$). Thus the discriminatory power of the development scores demonstrated acceptable criterion validity.

***TABLE 11*** *Mean, range and standard deviation (SD) of ACQ scores and ACQ subscale scores*

| ACQ subscale | *n* | Minimum | Mean | Maximum | SD |
|---|---|---|---|---|---|
| Development | 473 | 0 | 0.41 | 0.96 | 0.20 |
| Usability | 469 | 0 | 0.68 | 1.00 | 0.21 |
| **Total ACQ** | 466 | 0 | 0.52 | 0.98 | 0.16 |

**Distribution of scores**
*Table 11* shows the distribution descriptives for the scores calculated from the main study responses.

**Reliability**
Reliability was confirmed informally during the interview study (see chapter 4). During interviews with 40 responders, each was invited to confirm whether their responses still seemed appropriate when given the additional information of the mean ACQ score over all respondents and their own individual ACQ score. All agreed they were still appropriate.

## Development of review criteria

The development section of the questionnaire (see appendix 2) concentrated on the selection of review criteria and gave particular emphasis to whether the development criteria involved searching research literature and how good the quality of any search was.

### *Evidence-based review criteria*

A series of questions in the development section aimed to establish whether individuals had adopted an evidence-based approach to developing review criteria.

**Searching the literature**
A total of 337 respondents (71%) reported that criteria were totally ($n$ = 222, 47%) or partly ($n$ = 115, 24%) based on a search of research literature. Of those who searched the literature, 260 (77%) performed the literature search themselves, while 58 (17%) used some form of library information service to conduct the search. Only three respondents (1%) reported basing criteria on a published systematic review. Ten respondents (5%) failed to provide details of who performed the literature search.

**Age of literature review**
Respondents were asked to indicate how up to date the literature review was. The review was reported to be less than 1 year old by 131 respondents (39%), between 1 and 3 years old by a further 131 respondents (39%), while 19 respondents (6%) reported that the review was over 3 years old. Twenty-four respondents (7%) did not know how old the literature review was and 26 (7%) failed to answer the question. The 150 respondents who reported that the literature review was over 1 year old were asked whether they had searched for more recent literature. Eighty (53%) of these individuals reported attempting to update the literature search, 68 (45%) had not and four (2%) failed to answer the question.

**Details of the literature review**
The questionnaire also addressed issues surrounding the level of detail provided with the literature review upon which criteria were based. Respondents were asked whether any record was made of: the sources used to identify the literature, whether the validity of the research was appraised, or the methods used to appraise validity. *Table 12* shows the frequency of responses for each of these questions.

### *Alternative bases for review criteria*

In addition to asking whether review criteria were based on searching the research literature, respondents were also questioned about a number of alternative sources on which they may have based their review criteria. We investigated whether review criteria had been based on: consultation with experts, consultation with local experts, consultation with colleagues, consultation with patients or carers, or criteria used in previous audits. These options were not mutually exclusive.

***TABLE 12*** *Frequency of responses to questionnaire items concerning details of the literature search*

| Records made of: | Yes | No | Missing | Total |
|---|---|---|---|---|
| The sources/databases used to identify the literature | 186 (55%) | 110 (33%) | 41 (12%) | 337 |
| Whether the validity of the research was appraised | 92 (27%) | 190 (56%) | 55 (16%) | 337 |
| The methods used to assess validity | 83 (25%) | 191 (57%) | 63 (19%) | 337 |

*Table 13* shows the frequency of responses to questions about alternative bases. Consultation with colleagues was the most frequent alternative basis for criteria, with consultation with local experts the second most frequent basis. Consultation with patients or carers was least likely to have been used to develop review criteria.

It was also possible to compare the results for those who had used a literature search and those who had not. This would determine whether the absence of a literature search led to greater reliance on alternative bases. The frequency of reported use of alternative bases for review criteria are shown for those who had (*Table 14*) and had not (*Table 15*) searched the literature.

Both literature searchers and non-searchers reported consultation with colleagues and local experts as the two most commonly used methods. Consultation with patients or carers was the least used method. When the distributions of responses across answers were compared, using a $\chi^2$ test, respondents who based criteria on a search of the research literature reported using alternative methods more often than did respondents who had not searched the research literature ($p = 0.000$ in all cases).

### *Piloting of review criteria*

It emerged that 224 respondents (47%) had pilot tested their review criteria for practical feasibility in some manner, 177 (37%) had not carried out any piloting, while 34 (7%) respondents did not know whether the criteria had been piloted. Forty-one (5%) respondents felt that pilot testing of review criteria was not applicable for the audit they were reporting on.

### *Prioritising the review criteria*

The ACQ asked whether the review criteria had been prioritised in any way and, if so, whether prioritisation was based on: impact on health outcome, quality of supporting evidence, cost implications, or importance to patients. *Table 16* displays the frequency of responses for each method of prioritisation. Response options concerning methods of prioritisation were

***TABLE 13*** *Frequency of responses to questions concerning alternative bases for review criteria (n = 476)*

| Were the audit criteria based on: | Yes | Partly | No | Don't know | Missing |
|---|---|---|---|---|---|
| Consultation with experts? | 143 (30%) | 73 (15%) | 172 (36%) | 12 (3%) | 76 (16%) |
| Consultation with local experts? | 194 (41%) | 88 (18%) | 118 (25%) | 8 (2%) | 68 (14%) |
| Consultation with colleagues? | 268 (56%) | 83 (17%) | 61 (13%) | 8 (2%) | 56 (12%) |
| Consultation with patients or carers? | 62 (14%) | 41 (9%) | 266 (56%) | 16 (3%) | 91 (19%) |
| Criteria used in previous audits? | 137 (29%) | 72 (15%) | 171 (36%) | 16 (3%) | 80 (17%) |

***TABLE 14*** *Frequency of responses to questions concerning alternative bases for review criteria by respondents who had searched the literature (n = 337)*

| Were the audit criteria based on: | Yes | Partly | No | Don't know | Missing |
|---|---|---|---|---|---|
| Consultation with experts? | 128 (38%) | 62 (18%) | 100 (30%) | 10 (3%) | 37 (11%) |
| Consultation with local experts? | 152 (45%) | 71 (21%) | 76 (23%) | 7 (2%) | 31 (9%) |
| Consultation with colleagues? | 195 (58%) | 67 (20%) | 47 (14%) | 8 (2%) | 20 (6%) |
| Consultation with patients or carers? | 49 (15%) | 35 (11%) | 187 (56%) | 15 (5%) | 51 (15%) |
| Criteria used in previous audits? | 100 (30%) | 57 (17%) | 126 (37%) | 12 (4%) | 42 (13%) |

***TABLE 15*** *Frequency of responses to questions concerning alternative bases for review criteria by respondents who had not searched the literature (n = 139)*

| Were the audit criteria based on: | Yes | Partly | No | Don't know | Missing |
|---|---|---|---|---|---|
| Consultation with experts? | 15 (11%) | 11 (8%) | 72 (52%) | 2 (1%) | 39 (28%) |
| Consultation with local experts? | 42 (30%) | 17 (12%) | 42 (30%) | 1 (1%) | 37 (27%) |
| Consultation with colleagues? | 73 (53%) | 16 (12%) | 14 (10%) | 0 (0%) | 36 (26%) |
| Consultation with patients or carers? | 13 (9%) | 6 (4%) | 79 (57%) | 1 (1%) | 40 (29%) |
| Criteria used in previous audits? | 37 (27%) | 15 (11%) | 45 (32%) | 4 (3%) | 38 (27%) |

***TABLE 16*** *Frequency of respondents reporting use of methods of prioritisation (n = 476)*

| Audit criteria prioritised on: | No. of responses |
|---|---|
| Importance to patients | 324 (68%) |
| Impact on health outcome | 320 (67%) |
| Quality of supporting evidence | 200 (42%) |
| Cost implications | 139 (29%) |

not mutually exclusive and most respondents reported using more than one method. In total, 384 respondents (81%) had used at least one method of prioritisation: 64 (16%) of prioritising respondents used only one method, 115 (30%) used two, 129 (34%) used three and 76 (20%) used all four methods. Clearly, prioritisation according to cost implications was the least reported method of prioritisation, with impact on health outcome and importance to patients being the most frequently reported methods.

### *Transparency of developing review criteria*

Two questions on the development of criteria related to the degree to which the development process was transparent. In total, 288 respondents (61%) reported that the development method was described in sufficient detail to be repeated, 72 respondents (15%) felt that the development was only partly described, while 116 (24%) did not feel that the criteria development was described in enough detail to be repeated.

A further question on transparency applied to respondents ($n = 305$) who had used a combination of searching the research literature and consulting with experts to develop their review criteria. Of these, 102 (33%) reported that the method had been made fully explicit, 82 (27%) that it was only partly explicit, and 121 (40%) did not feel the method was made explicit.

## Subgroup analyses

Information asked for on the covering page of the questionnaire was designed both to confirm that the type of audit that the criteria related to, the scope of the audit and the setting were representative, and to allow comparison of the ACQ scores for each of these subgroups and the source of the review criteria.

### *Type of audit*

Respondents were asked to provide details of the title of the audit they were considering when completing the questionnaire, as well as the clinical areas it covered. This information was used by the researchers to classify audits as clinical or non-clinical (e.g. service or organisational audit). Non-clinical audits were those where a service pathway or organisation structure was being examined, rather than the implementation of care in a named clinical setting or an identified clinical condition. This classification was done independently by two of the researchers, and their classifications were then compared. There was disagreement on only two audits, and this was resolved by discussion.

**Clinical versus non-clinical audits**

A total of 465 respondents gave the title of their audit. The researchers classified 257 audits (55%) as clinical and 208 audits (45%) as non-clinical. There was little difference between mean ACQ scores for development and usability according to whether audits were classified as clinical or non-clinical (*Table 17*). Between subjects univariate analyses of variance showed no difference between scores for clinical or non-clinical audit topics.

***TABLE 17*** *Mean (SD) total ACQ, 'Development' and 'Usability' scores for clinical and non-clinical audits*

| | *n* | Mean (SD) | *p* |
|---|---|---|---|
| ***Total ACQ*** | | | |
| Clinical | 204 | 0.53 (0.16) | |
| Non-clinical | 251 | 0.51 (0.17) | 0.466 |
| ***Development*** | | | |
| Clinical | 207 | 0.43 (0.20) | |
| Non-clinical | 255 | 0.39 (0.20) | 0.072 |
| ***Usability*** | | | |
| Clinical | 205 | 0.67 (0.20) | |
| Non-clinical | 253 | 0.69 (0.21) | 0.358 |

### *Scope of audit*

Respondents were asked to indicate whether the audit they reported on was national, regional, run with other trusts or practices in the area, or limited to a single organisation. This question was answered by 473 respondents. *Table 18* displays the number of respondents in each category. The majority of audits were restricted to a single organisation.

***TABLE 18*** *Frequencies of the reported scope of audit (n = 473)*

| Scope | *n* |
|---|---|
| National | 34 (7%) |
| Regional | 21 (4%) |
| With other trust or practice | 78 (16%) |
| Single organisation | 340 (71%) |

***TABLE 19*** *Mean development, usability and total ACQ scores according to audit scope*

| Scope | *n* | Development | Usability | Total ACQ |
|---|---|---|---|---|
| National | 33 | 0.52 | 0.71 | 0.59 |
| Regional | 21 | 0.44 | 0.62 | 0.51 |
| With other trust or practice | 77 | 0.42 | 0.69 | 0.53 |
| Single organisation | 335 | 0.39 | 0.68 | 0.51 |
| All | 466 | 0.41 | 0.68 | 0.52 |

The mean development and usability subscale scores associated with audits of each kind of scope are displayed in *Table 19*. The development and usability scores from national audits are higher than for any other kind. However, the disparity between the sizes of subgroups precluded the use of inferential statistics to evaluate these differences.

### *Source of review criteria*

Respondents were asked whether the review criteria they used were drawn from published audit protocols, published guidelines, developed with the help of audit support staff, or developed by the individual audit lead. *Table 20* displays the frequency of responses to this question, with the mean scores, according to source of criteria, for development, usability and total ACQ. Although the usability scores appear relatively similar for each type of criteria source, there is clearly greater variation among development scores. Scores given for review criteria drawn from published protocols are higher than for any other source, while scores given for individually developed criteria are lower than those from any other source.

***TABLE 20*** *Mean development, usability and total ACQ scores according to source of review criteria* (n = *476)*

| | Mean | *n* |
|---|---|---|
| ***Total ACQ*** | | |
| Published protocol | 0.59 | 84 (18%) |
| Published guideline | 0.54 | 121 (25%) |
| Support staff | 0.51 | 77 (16%) |
| Developed own | 0.48 | 194 (41%) |
| ***Development*** | | |
| Published protocol | 0.51 | 82 |
| Published guideline | 0.44 | 120 |
| Support staff | 0.40 | 76 |
| Developed own | 0.34 | 191 |
| ***Usability*** | | |
| Published protocol | 0.69 | 83 |
| Published guideline | 0.69 | 120 |
| Support staff | 0.68 | 77 |
| Developed own | 0.68 | 193 |

A significant difference according to criteria source was observed for development scores ($F(3, 469) = 16.44$, $p = 0.000$) and total ACQ scores ($F(3, 462) = 9.99$, $p = 0.000$). *Post hoc* analysis using Tukey's highly significant difference tests revealed that respondents who developed their own criteria had significantly lower development scores and total ACQ scores than did those who used published protocols ($p = 0.000$) or published guidelines ($p = 0.000$). There were no significant effects for usability scores ($F(3, 465) = 0.44$, $p = 0.725$).

Review criteria developed from published protocols had less than perfect ACQ scores, and so it was investigated which items were absent. *Table 21* lists the questionnaire items that were absent for more than 20% of respondents whose review criteria were derived from published audit protocols. These indicate where changes in the methods of review criteria selection would be most likely to produce improvements.

### *NHS trust or general practice setting*

The sample of 476 respondents identified 390 (82%) from NHS trusts and 86 (18%) from general practices. This classification does not fully distinguish between primary and secondary care settings, as some of the trusts were community trusts and trusts combining community with acute or mental care. Nevertheless, the distinction can be made between review criteria used in general practice audits and those in trusts. *Table 22* shows the development, usability and total ACQ scores for respondents from general practice and trust settings. There was no significant difference found in any of the scores (development, $F(1, 467) = 5.08$, $p = 0.254$; usability, $F(1, 471) = 1.30$, $p = 0.25$; total ACQ score, $F(1, 464) = 3.75$, $p = 0.53$).

## Qualitative data on problems and their solutions

The ACQ gave respondents the opportunity to comment on any problems they may have experienced in selecting, developing or using review criteria, and strategies to deal with them.

***TABLE 21*** *Questionnaire items that were absent for more than 20% of respondents (n = 82) whose review criteria were derived from published audit protocols*

| Questionnaire item | | % of respondents with item absent |
|---|---|---|
| 1.1 | (e) Were the audit criteria based on consultation with patients or carers? | 45 |
| 1.1 | (f) Were the audit criteria based on criteria used in previous audits? | 29 |
| 1.2 | (f) Was the following information recorded (by you or the authors of the review): the sources/databases used to identify the literature? | 21 |
| 1.2 | (g) Was the following information recorded (by you or the authors of the review): whether the validity of the research was appraised? | 41 |
| 1.2 | (h) Was the following information recorded (by you or the authors of the review): the methods used to assess validity? | 40 |
| 1.3 | Is the method of combining evidence from the literature and expert opinion made explicit? | 21 |
| 3.1 | (d) Did the criteria have information on: how the demands of the audit on patients might be minimised? | 39 |
| 3.1 | (e) Did the criteria have information on: how the demands of the audit on staff might be minimised? | 56 |
| 3.3 | Were patients consulted about the acceptability of these criteria for them? | 48 |
| 3.4 | Were all relevant staff consulted about the acceptability of these criteria for them? | 23 |

***TABLE 22*** *Mean development, usability and total ACQ scores in general practice and NHS trust settings*

| | *n* | Mean |
|---|---|---|
| ***Total ACQ*** | | |
| General practice | 83 | 0.49 |
| NHS trusts | 383 | 0.53 |
| ***Development*** | | |
| General practice | 84 | 0.38 |
| NHS trusts | 389 | 0.41 |
| ***Usability*** | | |
| General practice | 84 | 0.63 |
| NHS trusts | 385 | 0.69 |

A total of 150 respondents provided comments regarding problems with the development of the review criteria, 124 commented on problems with the usability of the criteria, and 66 respondents provided general comments at the end of the questionnaire.

### *Qualitative data coding*

Coding schemes were developed and used independently by three researchers to categorise the qualitative data. Three-way Cohen's κ statistics were calculated to estimate the degree of agreement between raters. The κ statistics showed high inter-rater reliability for all codes used (see appendix 7). The qualitative data were then entered into a NUD*IST NVivo™ data file and re-coded according to the developed schemes. The NVivo programme assisted in organising the various coding categories around eight main themes, running through development problems, usability problems and other comments. The distribution of comments associated with each theme is shown in *Table 23*. Further details of the themes, and the individual codes they cover

***TABLE 23*** *Frequency of comments from each section of the questionnaire associated with each coding theme*

| Theme | Development problems | Usability problems | Other comments | Total |
|---|---|---|---|---|
| Validity issues | 21 | 91 | 2 | 114 |
| Organisational issues | 40 | 13 | 9 | 62 |
| Demand issues | 20 | 16 | 11 | 47 |
| Literature issues | 35 | 3 | 4 | 42 |
| Audit focus | 25 | 11 | 0 | 36 |
| Practical issues | 22 | 2 | 11 | 35 |
| Attitudes and perceptions | 2 | 21 | 5 | 28 |
| Standards issues | 11 | 5 | 7 | 23 |
| **Total** | 176 | 162 | 49 | 387 |

are provided in the following sections. Sample comments are given to illustrate the properties of each theme.

### *Validity issues*

The theme of validity issues covered the clarity of the review criteria, whether the criteria were viewed as appropriate, the sample used in the audit, and the quality of data drawn from the audit. Strategies for dealing with validity issues were offered by 64 respondents.

**Data problems**

All comments relating to this problem ($n = 44$) were reported in the usability section of the questionnaire. Many of the data validity comments focused on missing or incomplete data:

- "Staff unsure about completion of service deficiency forms – thought they were filed in case notes – none were."
- "People not completing audit questionnaire properly – 'other' for things that fitted criteria, missing section."

In addition, some respondents referred to problems in collating data from different sources. There was suspicion that data may have been biased:

- "Team of auditors took ten sets of notes each – each auditor tended to put their own spin on the questions."

*Solutions*: strategies centred around using an alternative source to verify gathered information (e.g. telephoning clients to confirm details, having checks in place while data are being gathered or entered).

**Clarity of review criteria**

Problems encountered in developing clear, unambiguous criteria were mentioned by 37 respondents. Clarity of criteria was considered a validity issue, as comments implied that ambiguity influenced the quality of data collected:

- "Definitions and instructions not always clear for audit criteria – could not answer some questions due to ambiguity."
- "Some definitions of the variables to be measured were too imprecise, so we weren't sure that the correct information was collected."

*Solutions*: strategies suggested to deal with the clarity of the review criteria noted the need to make criteria more explicit (e.g. "More explicitness for criteria on repeat protocol", "Revised protocol to ensure all clinicians clearly understood criteria").

**Sampling issues**

Some respondents reported being aware that the sample used was non-representative:

- "Wording of capture form excluded/included some appropriate/inappropriate groups."

Others were less certain but voiced concerns that there may be problems with the sample:

- "Difficulty selecting patient group from computer system – unsure whether complete group identified."

*Solutions*: these focused on ensuring that staff involved in the audit understood which cases should be included (e.g. "Issuing guidance to clinicians about conditions that qualify/disqualify patients for audit inclusion").

**Appropriateness of review criteria**

This concerned whether criteria were viewed as appropriate by those using them. Examples of comments include:

- "Audit tool used was developed for acute hospital unit – difficult to adapt to mental health, district nurses and learning disability units. Some criteria not appropriate."
- "Autonomy of professionals presented problems with criteria applicability, i.e. some staff were not happy with them in practice."

*Solutions*: suggestions centred on piloting the audit tool to test appropriateness (e.g. "'Piloted on small group and added questions relating to their assessment of what's important").

### *Organisational issues*

The theme of organisational issues covered the areas of setting up the audit and coordinating different staff groups.

**Coordinating different staff**

Coordinating different groups of staff was the most frequently mentioned organisational issue ($n = 42$). Problems concerning coordination of staff included logistical issues of getting staff from different departments or different organisations together:

- "Difficult getting staff from 2 trusts and 3 departments within trust together, plus getting agreement of guidelines."

There were also issues concerning coordination of staff from different disciplines:

- "Main consideration – to get health and social services staff to work together, involved discussion of what was feasible."

*Solutions*: these focused on establishing regular formal meetings (e.g. "Regular multi-disciplinary meetings to discuss and improve communication, practice").

**Setting up the audit**
Problems relating to setting up the audit were mentioned by 19 respondents. Problems tended to concern practicalities such as the time required to organise audit teams:

- "Effective liaison with ENT. Time required to set up audit was greater than anticipated."

*Solutions*: no strategies to deal with problems setting up the audit were offered by any respondents.

### *Demand issues*
The theme of demand issues related to problems due to time and funding limitations.

**Time limitations**
Time factors were mentioned by 38 respondents:

- "Time to get team members together to discuss criteria etc."
- "Time taken to develop acceptable/applicable criteria was very demanding."
- "No 'time-tabled' time to conduct audit – not enough time to do it properly."

*Solutions*: it was suggested that the allocation of protected time would greatly enhance the quality of audits. Others said they sought assistance for the audit (e.g. "Requested secretarial support", "Excellent support from trusts audit department") or relied on the goodwill of the auditor (e.g. "Done in non-work time on own goodwill", "Exceeded the EU working directive for hours at work").

**Funding issues**
Eight respondents forwarded comments on audit funding. Lack of adequate budgets for audit activity were reported:

- "Audit has cost implications – some services felt could not take part due to resistance from management for extra funding for audit."
- "Finding funds from already fixed and tight budget."

*Solutions*: strategies to deal with funding issues centred on seeking grants to assist audit activity (e.g. "Got a grant from [the] Health District", "Applied for funding from clinical audit department").

### *Literature issues*
Literature issues concerned the level of access to and availability of relevant research evidence.

**Availability of literature**
Most comments ($n = 34$) on the theme of literature issues concerned a lack of available literature upon which to base criteria. The problems identified concerned the scarcity of literature examining a particular clinical topic:

- "Main problem was lack of research into the effectiveness or otherwise of challenging behaviour training."

and the lack of an evidence based approach for a given clinical discipline:

- "As we work in rehab, and largely non-medical environments, we have very little research findings to guide our audit criteria."

*Solutions*: the most common suggestion centred on consultation with colleagues and/or experts in the area to overcome gaps in the evidence base (e.g. "Took a consensus of opinion of local experts"). However, some respondents reported developing review criteria without such consultation (e.g. "Used my own common sense", "Made up our own").

**Access to literature**
Eight respondents reported problems with accessing literature. Problems associated with access to the literature included physical access to libraries:

- "Access to a library. Our nearest library is 23 miles away and we don't have any virtual library connections."
- "Hard to access electronic databases at hospital library."

Even where access to libraries was available, further problems locating identified publications were reported:

- "Difficult to obtain some of the references once identified."

*Solutions*: no solutions were offered to deal with these problems.

### *Audit focus*

The theme of audit focus covered issues of refining review criteria sets and adapting national protocols to local use.

**Refining review criteria sets**

The issue of refining criteria sets was mentioned by 29 respondents. Being able to refine criteria sets was seen as important in ensuring that the audit was carried out easily:

- "Problems in choosing too many criteria for each audit and not being selective enough, makes audit extensive + time consuming."
- "Trying to reduce the number of measurable criteria to aid compliance for staff."

*Solutions*: these concentrated on producing more specific audit protocols (e.g. "Changed the audit tool to be more specific which reduced the number of criteria").

**Adapting national protocols**

Seven respondents mentioned difficulties experienced adapting national audit protocols to local use.

- "Audit tool very specific about infection control criteria – many mattresses failed due to this. Local criteria not in agreement."
- "Criteria not enough detail for local purposes."

*Solutions*: these problems tended to be dealt with by altering protocols to suit local needs (e.g. "Shortened and simplified version of protocol used", "Re-wrote some of the criteria to meet local needs").

### *Practical issues*

Comments on practical issues covered access to data required for the audit and the availability of skills and training to assist with audit activity.

**Data access**

The most frequently cited ($n = 28$) practical problem concerned access to the data required for the audit. Typical problems included:

- "Availability of info/data tortuous to extract."
- "More difficult than expected to retrieve data for the criteria that determined what constituted adequate A&E [accident and emergency] assessment."

*Solutions*: the only solutions offered to deal with such problems were to "Persevere" and "Make time".

### *Skills and training*

Seven respondents commented on the difficulties from a lack of adequately skilled/trained staff:

- "Biggest problem is lack of time and expertise to critically appraise the literature."
- "We needed further training for staff. Skills required/experience/qualification of the auditor."

*Solutions*: these involved organising training sessions for staff involved in audit (e.g. "Education session from clinical practice research unit").

### *Attitudes to audit*

The theme of attitudes to audit concerned the degree of compliance with the audit and levels of staff motivation.

**Compliance with audit**

The most common ($n = 21$) attitudinal issue centred on individuals failing to acknowledge the importance of the audit and comply accordingly:

- "GPs refusing access to surgeries claiming the audit needed to pass ethical review. Our local ethics committee disagreed."
- "Medics are not very helpful. They are wary of giving information about their prescribing and monitoring procedures."

*Solutions*: strategies to deal with such problems focused on attempts to alter perceptions of the audit process (e.g. "Explained in a non-threatening way that audit was not to check up on individuals", "Explained the advantages of audits").

**Staff motivation**

Seven respondents made explicit reference to staff motivational issues:

- "Motivating staff to carry out audit."
- "Problems in staff motivation. Some of them forgot to fill in pro-formas."

*Solutions*: motivational issues were dealt with by attempts to generate enthusiasm for the audit project (e.g. "Identify a lead person to maintain communication and enthusiasm").

### *Standards issues*

The theme of standards issues dealt with uncertainty regarding whether the standard for an audit was realistic. Twenty-three comments were made, including:

- "Because we developed them locally, setting the standard was probably most difficult – i.e. was 100% compliance unrealistic?"
- "Setting standards that clinicians accept."

*Solutions*: no solutions were offered to deal with the issue of standard setting.

### *Summary*

The problems and solutions are summarised in *Table 24*. Many of the themes, which emerged from the qualitative comments provided in the questionnaire, were revisited in the follow-up interview study. Details of this study and its findings are provided in chapter 4.

## Discussion

The ACQ was created to meet the main aims of this study: to assess the extent to which systematic methods are used to select review criteria, and to assess the quality of the review criteria actually used in clinical audit in the UK. The items in the questionnaire were the desirable characteristics of review criteria, as agreed by expert consensus. The questionnaire was a valid measure that could discriminate between the quality of methods used to develop audit review criteria. Since audit against review criteria is an integral part of many quality improvement activities, the creation of this instrument has important implications for those evaluating quality improvement programmes, such as the clinical governance framework in the UK health services. Identification of good practice in review criteria selection for audit should enable strengths to be built upon, and identification of practice that is less than ideal may afford remedial measures for future quality improvement activities.

### Use of evidence from the literature

This study has shown that reported methods of selecting review criteria were often less systematic than is desirable, and that review criteria used often had few of the desirable characteristics. The observation that 29% of respondents had not based their review criteria on searches of the research literature is cause for concern. Reviewing quality of care against evidence-based criteria means that any changes in practice to reach standards associated with those review criteria should result in improved care. In contrast, if review criteria are not backed up by sound research evidence, reaching target standards may simply result in a change in practice rather than improvements in care, and it will not be known which is the case.

It is encouraging that 71% ($n = 337$) of respondents did report basing their review criteria on a

***TABLE 24*** *Problems and suggested solutions in selecting review criteria from the qualitative answers to the ACQ*

| Theme | Problems | Solutions |
|---|---|---|
| Validity | Poor data | Checking validity; checking data errors |
| | Ambiguous criteria | Explicitness in criteria |
| | Biased samples | Clarity in inclusion/exclusion criteria |
| | Inappropriate criteria | Piloting and using feedback from staff |
| Organisational | Coordinating staff | Regular formal meetings |
| | Time needed for set up | None |
| Demand | Time limitations | Protected time for audit |
| | Funding | Seeking external funding for audits |
| Literature | Lack of evidence base | Consult experts and colleagues |
| | Poor access to libraries, databases or papers | None |
| Audit focus | Too many criteria | Focus the audit |
| | National protocols not suitable | Amend to suit local needs |
| Practical issues | Data difficult to extract | Persevere; allow time |
| | Lack of skills | Training |
| Attitudes to audit | Doctors not complying | Rational explanation |
| | Lack of staff motivation | Enthusiastic leader |
| Standards | Setting an appropriate target | None |

search of the research literature. However, searching the research evidence is a necessary, but not sufficient, condition for developing appropriate review criteria. The literature search undertaken should also consider the quality of the literature identified, in terms both of its recency and of the validity of the reported findings.

Of those 337 respondents who had based their review criteria on a search of the literature, 78% used a literature review that was less than 3 years old. Also, 53% of the respondents using literature more than 1 year old ($n$ = 150), reported trying to identify more recent literature. In such cases it may have been that more recent publications were not available. Nevertheless, this leaves a substantial proportion of respondents relying on evidence that may be out of date. With regard to whether the validity of the evidence was considered, 27% of the literature searching respondents recorded whether the validity of the research was appraised, and 25% recorded the methods used to appraise it. This means that in over 70% of cases where audit review criteria were reported to be evidence based, there was no report of checking whether that evidence was recent and valid.

Perhaps the less than ideal findings regarding recency of literature and validity checking can be explained by the observation that most respondents reported searching the literature themselves. Indeed, only 1% of all literature searching respondents used a published systematic review, and only 18% had used a library information service. Individual practitioners may be less experienced in critical appraisal skills than are professional researchers carrying out systematic reviews or experienced librarians. In addition, individual audit leads may have inadequate time available to them to conduct thorough searches and identify the most recent available evidence.

## Use of experts and colleagues

The survey results suggest that the practice of searching for evidence from the literature may be associated with other aspects of good practice regarding the development of review criteria. Literature searching respondents were more likely to consult with experts when developing review criteria than were those who had not searched the research literature. The expert consensus (chapter 2) was that appropriate review criteria development involves combining research evidence and expert opinion. Therefore, it is encouraging that literature searching respondents tended also to consult experts. However, 33% of the 305 respondents who used both literature and expert opinion to develop review criteria reported that the method of combining evidence and expert opinion was not made explicit.

Consultation with colleagues (wholly or partly) was also reported by 78% of literature searching respondents. This was not an item on the list of desirable characteristics, but was included in the questionnaire for completeness of the possible responses. Such consultation may be viewed as good practice where it involves a discussion of research findings. However, where it becomes a substitute for examining research evidence, based on the convenience of accessing colleagues in the midst of everyday activities, there may be implications for the quality of review criteria development. Although clinical colleagues are likely to have implicit knowledge of research evidence, given the overwhelming volume of publications that need to be read to keep abreast of the most current research, there is a risk that they are not fully cognisant of all relevant literature. Use of clinical common sense may be appealing, but it is far from objective. Therefore, it is cause for concern that 75% of respondents who had not searched the research literature reported fully or partly basing review criteria on consultation with colleagues.

A further major cause for concern was the rarity of consultation with patients or carers. This was the least often used source of information on which review criteria were based, and reinforces the low recognition of the value of service users in quality improvement activities.

## Piloting

About half of respondents used review criteria that had been piloted. Piloting allows ambiguities, contradictions and impracticalities to be discovered at low cost, and therefore avoided in the higher cost full audit. Audits using unpiloted review criteria risk wasting time and resources in discovering that the criteria are infeasible, contradictory or ambiguous. These can lead to audits producing invalid or unusable data, which is not only wasteful of resources, but is also discouraging to those whose time and effort have been used in the audit.

## Prioritisation

Both the literature on review criteria development and expert consensus (chapter 2) identify prioritisation of criteria as important. Hence it was encouraging to find that 81% of respondents had prioritised their criteria in some way. Most of these respondents, 84% ($n$ = 320) had used more

than one method of prioritisation. Prioritisation according to importance to patients was used most often. This finding is somewhat surprising given that such a small proportion of respondents (see *Table 13*) reported consulting patients or carers when developing their review criteria. It is possible that criteria may be better described as prioritised according to clinicians' estimations of importance to patients, rather than according to actual patients' perceptions of importance.

Criteria were also often prioritised on impact on health outcome, but less than half of responders who prioritised their review criteria did so on the quality of evidence upon which the criteria were based. This is a further cause for concern, as the expert consensus was that this was the most important reason for prioritisation.

## Transparency

Almost one-quarter of respondents felt that criteria development was not described in enough detail to be repeated. This means that it is not possible to assess the validity of the method, and hence the validity of the review criteria. Thus, selection of review criteria is impeded by lack of information, which could easily be provided.

## Scores

An analysis of individual items for the development section highlighted many areas where the method of selecting review criteria was open to improvement. This is reflected in the fact that none of the respondents achieved a perfect score, despite all the items being deemed feasible by the expert panel and all items being achieved by at least one respondent. The mean ACQ score was 0.52 (range 0 to 0.98). Many respondents were far from achieving perfection in their review criteria development scores (mean 0.41, range 0 to 0.96), although usability scores were higher (mean 0.68, range 0 to 1.0). It appears that creating practical, easy to apply review criteria is more achievable than developing criteria in a systematic, evidence-based manner.

There are risks associated with this. If review criteria drawn from published protocols are applied with relative ease, this might lead to the assumption that they are valid and appropriate. However, to say something is easy to measure does not necessarily mean it is the right thing to measure. Ease of criteria application is no doubt an important consideration, but it should not override the importance of using systematically selected sets of evidence-based review criteria.

## Subgroup comparisons

Between-groups analyses determined whether particular topics, organisation, setting or criteria bases had relationships with the quality of the review criteria.

### *Clinical versus non-clinical audits*

The discovery that scores provided for audits classed as clinical and non-clinical did not differ significantly was perhaps surprising. Given the evidence-based focus of the development subscale, one might have expected audits on non-clinical topics to be associated with lower development scores than those examining clinical issues. The label 'non-clinical' was applied to audits where a service pathway or organisational structure was being examined. It seems intuitively plausible to assume that there is less relevant research literature available to guide these activities than there would be for clinically related topics dealing with a particular treatment option or procedure, but this was not found.

### *Single organisation versus multi-organisation audits*

The majority of audits (71%) were restricted to the scope of being within a single organisation. The reporting of national or regional audits was extremely rare. Only 7% of respondents reported on national audits, 4% on regional audits and 16% on audits run in conjunction with other trusts or practices in their area. Scores suggested that single organisation audits, the vast majority of audits, were associated with lower development scores.

### *Self-developed versus published audit review criteria*

A significant effect was observed regarding the source of the review criteria. Those who developed their own review criteria scored significantly lower on the ACQ scale than did those who had used criteria drawn from a published audit protocol. Given that 41% of respondents reported developing their own criteria, this suggests that a substantial proportion of clinical audits in England and Wales use poorly developed review criteria. This increases the likelihood that audits using such criteria will fail to facilitate changes that lead to improved care.

In addition, although questionnaire scores for review criteria based on published protocols were significantly higher than those developed by individuals, the mean score of 0.52 for published review criteria is far from the maximum possible score of 1. This score suggests that half of the desirable characteristics of review criteria are

absent. This does not necessarily reflect that the published protocols were based on a poor method of criteria selection, but it could be that the protocol did not include any accompanying information about its development. This could make it difficult for questionnaire respondents to answer some of the development questions. However, this is still an important issue to be addressed. If published audit protocols do not provide complete details of how their review criteria were selected it is almost impossible for individuals carrying out quality reviews to assess their appropriateness. In order to ensure that review criteria are valid, it is vital to have knowledge of the evidence base they were drawn from, the quality of that evidence, the reasons behind any prioritisation, and so on. Published audit protocols should include a detailed and transparent account of how their review criteria were developed and so demonstrate all the desirable characteristics of review criteria. The results of this study suggest that this is not the case for many of the published audit protocols used in England and Wales.

This conclusion is supported by the exploration of missing items from published protocols. Those relating to the development of review criteria covered consultation with patients or carers, reference to review criteria from previous audits, transparency of the methods used to gather data from the literature, or combine that with the views of experts. The missing items relating to the usability of review criteria covered the presentation of information on minimising the demands of audit on staff and patients, and consulting with staff and patients in the selection of review criteria.

### *Setting*

The scores for review criteria used in general practice audits and in NHS trust audits were not different. This is an encouraging finding, since the development of clinical audit and, recently, clinical governance in these two settings has been very different. The support of audits in general practice has been mediated by the PCAGs in each area, which has allowed the independent status of general practitioners to be maintained while encouraging and expecting as much audit activity as in trusts.

### *Identified problems*

The most commonly noted problems associated with review criteria development focused on organising the audit and gathering evidence from the literature upon which to base criteria. Organisational problems tended to be in the coordination of meetings to discuss criteria development, which were possibly solved by establishing regular meeting and good communication patterns. The amount of time taken to set up the audit and develop the review criteria was a frequent problem. If individuals do not have sufficient time to develop review criteria in a systematic, evidence-based manner, shortcuts may be taken, which could explain why scores for the development of the review criteria were rarely close to the maximum possible. Protected time for audit is one possible solution.

Comments concerning problems with literature-related issues provided further insight into why development subscale scores were often relatively low. For some respondents an evidence-based approach was not apparent in their clinical discipline or topic and there was little, or no, research evidence to guide their practice. It is cause for concern that some individuals accepted reliance on their own common sense to select criteria in the absence of published evidence. While the common sense of individual practitioners could be in line with the latest evidence, it is not a certainty.

Even where evidence-based specialities and well-researched topics are concerned, barriers to basing review criteria on research evidence can persist. Several respondents commented on difficulties gaining access to the literature: libraries were geographically distant, facilities for literature searching were unavailable, or resources were not available to subscribe to, or locate copies of, specialist journals. These issues could be addressed by improving library and information searching facilities. Funding issues were mentioned by a number of respondents, although in general terms these were issues related to the cost of the whole audit process rather than to specific literature-related issues.

The difficulties experienced in using a systematic method to develop review criteria may explain why the most frequently reported problem with the usability of the criteria related to validity. Validity problems related to ambiguous criteria (confusing those who conducted the audit), incomplete or missing data, and criteria being viewed as inappropriate by staff involved in the audit. It seems likely that problems such as these would be avoided by a more systematic, evidence-based, approach to criteria development. Review criteria are less likely to be viewed as inappropriate if they are backed by up to date, valid research

evidence. A systematic approach, including some form of piloting of criteria, should allow any ambiguities in the criteria to be identified and rectified at an early stage. This should, in turn, reduce the prevalence of missing or incomplete data in the audit. Piloting could also help address the practical issues that respondents mentioned by, for example, flagging areas where access to data may be problematic or where specific data handling skills are required. This could allow contingency plans to be developed, such as the seeking of alternative data sources and skills training for relevant individuals.

Adopting a systematic approach to selecting review criteria could also be useful in dealing with another identified problem with developing and using review criteria, focusing the audit. Some respondents had trouble in narrowing down large criteria sets to produce a manageable audit protocol. A systematic approach to criteria development is likely to involve some form of prioritisation of criteria, for instance into 'essential' and 'recommended' categories. Prioritising in this manner could help individuals manage larger criteria sets and limit the focus of their audit.

### Critique of the study

This survey using a self-completion questionnaire, depended on the cooperation of audit coordinators to distribute the questionnaires, and on audit leads to complete the questionnaires. Every participating NHS trust and PCAG returned at least one completed questionnaire. Nevertheless, there were probably biases in our sample of respondents, even though the sample was large enough to provide power to justify generalising from the results. The potential biases came from the selection by audit coordinators of to whom to give the questionnaire. They probably chose the cooperative audit leads, who were easily accessible to them, even if they were on a randomly selected list. Those who completed and returned the questionnaires would, in turn, be biased towards those who were enthusiastic about audit, were cooperative enough to complete the questionnaire, and who thought their audit review criteria were good enough to be reported on. In primary care, the questionnaires were sent to all practices by the PCAG, so respondents would also be the cooperative, enthusiastic audit leads. Thus, our results may be biased towards the responses of the highest quality audit leads, using the highest quality review criteria. Thus we might deduce that the results from the rest of the auditing population will probably be worse than those obtained in our study. This underlines the concern that the conclusions give. There is a need for great improvement in the review criteria selection of the sample reported in this study, and there is probably even greater need in the rest of the auditing population. The implications of these results are generalisable, because the probable bias does not detract from the implications of the conclusions.

## Summary

The ACQ was created to assess the extent to which systematic methods are used to select review criteria and to assess the quality of the review criteria actually used in clinical audit in the UK. The ACQ score was based on the list of desirable characteristics of review criteria derived from expert consensus.

Reported methods of selecting review criteria for clinical audit were often less systematic than is desirable. The mean ACQ score was 0.52 (range 0 to 0.98). A perfect score of 1 was possible, since each item was scored by at least one respondent, and all items had been rated as highly feasible by expert consensus, but many scores were very low.

Seventy-one per cent ($n$ = 337) of respondents based their review criteria on the research literature. Of these, 78% used a literature review that was less than 3 years old. Only 27% recorded whether the validity of the research was appraised and 25% recorded the methods used to appraise it. Thus over 70% of the cases that used evidence as the base for review criteria did not check the validity of the evidence. Furthermore, 29% of respondents had not reported using the research literature to select their review criteria. Only 1% ($n$ = 3) of all literature searching respondents used systematic reviews. Thirty-three per cent ($n$ = 102) of the 305 respondents who used both literature and expert opinion to develop review criteria reported that the method they used to combine evidence and expert opinion was not made explicit. Consultation with colleagues was the most commonly used basis for review criteria selection, as an alternative or supplement to evidence from the literature. However, patients or carers were rarely consulted.

About half of respondents used audit review criteria that had been piloted. Audits using unpiloted review criteria risk wasting time and resources in discovering, after collecting large amounts of data, that the criteria are unfeasible, contradictory or ambiguous. Of the respondents 815 had prioritised their review criteria. Most

had used more than one method of prioritisation. Prioritisation according to importance to patients was most often used. Prioritisation according to the quality of the evidence was the choice of the experts, but used by less than half of respondents.

Assessing the validity of review criteria is impeded by the lack of information on how review criteria were developed, even in published audit protocols.

Creating practical, easy to apply review criteria is more achievable than developing review criteria in a systematic, evidence-based manner.

Clinical and non-clinical audits did not differ significantly on ACQ scores.

The reporting of national or regional audits was extremely rare in the current study. Scores suggested that single-organisation audits, the vast majority of audits, were associated with lower ACQ scores than were national or regional ones.

The mean score of 0.52 for published review criteria implies that half the desirable characteristics of review criteria are absent. The items were all deemed feasible by expert consensus, and so a perfect score of 1 should be possible. Published protocols could improve their development methods, transparency and information on usability. The ACQ scores for unpublished review criteria were even lower than for published ones and 41% of respondents reported using unpublished criteria. Thus, the review criteria used in many audits do not meet the desirable characteristics of review criteria. This is reported by the review criteria developers themselves, thus reinforcing the validity of this conclusion.

Information on the methods by which review criteria are selected from the literature, consultation with patients and staff, and reference to criteria from previous audits are items that should be included in the process of selecting review criteria, but often are absent.

There was no difference in the scores for review criteria between audits from general practice and from NHS trusts.

The most commonly noted problems associated with review criteria development focused on organising the audit and gathering literature upon which to base criteria. Some respondents reported that an evidence-based approach had not yet taken hold in their particular clinical discipline and there was little, or no, research evidence to guide their practice. Several respondents had difficulty gaining access to the literature through libraries or specialist journals. Some respondents had trouble in narrowing down large review criteria sets to produce a manageable audit protocol.

Although the sample in this study was probably biased, the bias would be towards better ACQ scores. Thus the conclusion that ACQ scores need to be improved is reinforced.

## Next steps

The comments of respondents about problems experienced in developing and using review criteria and potential strategies to deal with them, are useful in filling out the picture provided by quantitative scores from the questionnaire scores. The results suggested that the development of review criteria left something to be desired in a number of cases. The qualitative comments provide a partial explanation as to why this might be the case by outlining why individuals did not search research evidence, pilot or prioritise criteria, and so on. However, the restricted space available to provide comments on the questionnaire and the restricted medium of the written word meant that the qualitative data from the questionnaire study was limited in scope. More detailed data exploring barriers to developing systematic, evidence-based review criteria, and strategies to overcome those barriers, were required. The final stage of the project aimed to provide such evidence from in-depth interviews with a sample of questionnaire respondents.

# Chapter 4

# Interview study

## Introduction

The aim of the interview study was to provide detailed information about barriers encountered in developing review criteria and potential strategies to overcome such barriers. Barriers to using systematic methods of criteria development would be identified by interviewing individuals whose questionnaire responses received low scores for the development of review criteria. Low scores suggest the presence of barriers for these individuals. The opinion of these individuals would also be sought regarding supports that might enable them to develop more appropriate criteria. In addition, insight from individuals whose questionnaire responses produced high development scores would allow further information regarding supportive factors to using systematic methods of criteria selection to be identified. Therefore, the final stage of the project involved in-depth interviews with a sample of questionnaire respondents from the upper and lower quartiles of the questionnaire development scores (i.e. those with scores in the top 25% and those with scores in the bottom 25%).

## Method

This interview study involved the use of semi-structured interviews with a sample of audit leads. The interviews covered aspects relating to obstacles and barriers to using systematic selection of review criteria. All interviews were taped, with the consent of participants, and subsequently transcribed.

### Participant recruitment

Participants were drawn from a pool of questionnaire respondents who indicated a willingness to take part in a follow-up study. The final section of the ACQ asked respondents to provide contact details if they were willing to participate in a brief follow-up interview. The majority of respondents (76%, $n = 362$) agreed to be contacted for the purpose of a follow-up interview. It is recognised that there may be barriers experienced only by those who were not willing to be interviewed, which would not be captured from this sample.

Random quota sampling was used to identify a sample of 40 participants, which included:

- both those using systematic methods of criteria selection and those using non-systematic methods
- a range of audit topics covering audits on both clinical non-clinical topics
- proportionate representation from general practice, acute, combined, mental health and community trusts
- respondents from both single organisation and multi-organisation audits.

One researcher (RH) contacted potential participants by telephone and thanked them for their cooperation in returning a questionnaire. These individuals were reminded of their expression of a willingness to take part in the interview study. They were then given a brief description of what the interview would involve and asked if they would still consider participating. All respondents contacted in this manner agreed to be interviewed. A convenient appointment for the interview was then scheduled, on the understanding that the interview would not last more than half an hour. All interview appointments were confirmed in writing by the researcher.

### The sample

The interview sample of 40 individuals comprised:

- 20 participants with criteria development scores in the upper quartile and 20 from the lower quartile
- 22 participants reporting audits on clinical topics and 18 on non-clinical topics
- seven participants from general practice, seven from mental health trusts, eight from acute trusts, nine from combined trusts and nine from community trusts
- 23 participants from single organisation audits and 17 from multi-organisation audits.

### The schedule

The project team created a structured interview schedule (see appendix 8). The interview schedule was piloted using five local questionnaire respondents who had not been selected for the main interview sample. The pilot interviews

indicated that the schedule worked well and was viewed as appropriate by respondents. Therefore the piloted schedule was used in the main interview study.

Permission to tape record the interview was requested from all participants before proceeding with the schedule. The interview schedule reminded participants of the nature of the questionnaire study and its main findings. Respondents were informed that the interview would concentrate on the development of review criteria, and were told the mean and range of development scores, for all returned questionnaires. Next, they were given the development score for their own questionnaire and asked to indicate whether this seemed a fair reflection of the review criteria for which they completed the questionnaire. The interview then focused on responses to individual questions in the development section, before soliciting general views on obstacles and barriers to using a systematic method of criteria selection. At the end of each interview, respondents were thanked for their contribution to the study and provided with a £5 book token as a gesture of goodwill.

### Data coding and analysis

The interview tapes were transcribed and a coding scheme developed by researchers. The reliability of the coding scheme was assessed by comparing how two researchers (RH and HH) applied the scheme independently to ten interview transcripts. The two raters showed perfect agreement regarding the codes they applied to each transcript. Thus the definition of each code in the scheme enabled reliable coding of data to be achieved. The ten coded interview transcripts and the remaining 30 transcripts were subsequently entered into a NUD*IST NVivo qualitative data analysis package. One member of the project team (RH) then coded all transcripts within the NVivo package, in accordance with the coding scheme.

## Results

The main aim of the interview study was to provide information on barriers, and supports, to using systematic methods of criteria selection. The information gathered is presented in the following sections. However, before presenting such findings it is worth briefly noting that the interview study also provided a means of testing the reliability of the ACQ. All interview participants were asked to reflect on whether their score seemed a fair assessment for the criteria they had reported on. All 40 interview participants believed the scores they received were a fair reflection of the criteria referred to.

The remainder of the interview focused on eliciting problems and supports experienced in developing review criteria. The coding of the data gathered allowed participants' responses to be grouped around a number of themes. The major themes that emerged from the interview study are shown in *Table 25*. They included and enlarged upon those themes highlighted by the qualitative questionnaire data.

***TABLE 25*** *Frequency of statements, related to themes, selected and included in the analysis*

| Theme | Frequency |
|---|---|
| Organisational | 50 |
| Work demand | 45 |
| Literature | 40 |
| Attitudes | 37 |
| Audit focus | 34 |
| Validity of audit | 33 |
| Practicalities | 30 |
| Standards | 23 |
| Others | 50 |
| **Total** | 342 |

### Literature issues

Issues surrounding the use of literature to base criteria on were mentioned by all interview respondents, irrespective of whether they had high or low ACQ scores.

#### *Searching the literature*

Obstacles to using research literature to guide criteria selection began at the very start of the literature search process. A number of respondents commented that literature searching was difficult without the availability of electronic search engines and databases:

- "We don't have access to things like MEDLINE, CINAHL and what-have-you here in the practice, so we can't really do any comprehensive literature searches."

The time taken to conduct proper searches was also viewed as problematic by a number of respondents:

- "Really it could take weeks to do a proper, thorough search, and sadly I just haven't got the time to do it."

Solutions to overcome such obstacles were readily suggested. They centred around improving access to search engines and providing literature searching services:

- "Obviously life would be much easier if we could just sit at our desks and log onto MEDLINE."
- "You need support to be able to say to somebody look, this is the question, could you please go and find me the evidence."

These suggestions may well be effective, since all interview respondents who had development scores in the upper quartile mentioned access to search engines as contributing to their ability to develop appropriate review criteria:

- "Well yeah, it helped to have such good library facilities. We had access to MEDLINE, PsycLIT, CINAHL."

In addition, many of these individuals had access to literature searching services:

- "But it's generally the literature search availability in the library here which helps, which is very useful. There are medical staff who will, who have the time and skills to search the library here, and we have librarians who are well into it."
- "We couldn't have done that in-depth search without having somebody doing, somebody in, in the background helping us."

### *Access to the literature*

Obstacles of access to research literature included limited access to libraries, either due to restrictive opening hours

- "The library's actually not bad, but it's just that by the time I've finished for the day it's shut, so you have to rely on days off to make it in."

or to the geographical distance to libraries

- "I'm doing a Masters at ... University so I can get in and use the library there, which is great for me – but if I had to rely on what's available from our practice library or even the nearest public library I'd be stumped. The problem is it's a good hour and a half drive over to ... so it's rare that I can make it, other than on days off."

In addition, even when libraries were close at hand and open at convenient hours, some considered their library stock was inadequate:

- "The things that came up in the literature search weren't available in the library, a lot of them. So we were either working from abstracts or we just didn't use those and we used other sources. So it was quite limited in the end."

Lack of access to literature appeared to have impinged on the quality of the review criteria developed. Good access to well-stocked libraries was mentioned, by those with high criteria development scores, as a factor facilitating the development of appropriate review criteria:

- "Umm, the British Library isn't far up the road. Umm, so if we want anything we tend to go over there, anything, you know, telephone, make an appointment, and you can go in, and they know who you are, and you can look through more or less anything you want. So yeah, I, I've gone up there quite a few times and got bits and pieces of research, or books."
- "Err, the library has access to Internet and Cochrane database, and if they stock most of the journals, but if not we can get any reprints we want to, the library can get it to us."

### *Adequacy of literature*

Access to literature-searching facilities and library services does not necessarily mean that pertinent literature will be available on which to base criteria. A common complaint from interview respondents was that there was little available evidence surrounding the specific topic they wished to carry out an audit for:

- "There is a dearth of evidence, a dearth of research around young people, and child mental health generally."
- "But basically, if you really look at the literature, there's no evidence there to actually evidence base what we do."

Such a shortage of literature was largely seen as due to particular disciplines and clinical areas failing to embrace an evidence-based culture:

- "Physiotherapy is not a very evidence-based field, professionally, as a whole."
- "In some, some sort of specialisms it's, it's hard, because not a lot of what they do is evidence-based. Particularly in the therapies."

As well as complaints about the quantity of available literature, reference was also made to the quality of the literature. A number of respondents were concerned that the standard

of research employed in available publications meant that conclusions may not have been valid:

- "I mean there's a lot of literature around and about physiotherapy, but it's just not very good. There's poor samples, you know, there's not a lot of randomised, but mainly sort of non-RCT stuff."
- "Fifty percent of what gets published is absolute rubbish. The problem is you often don't know which fifty percent ... absolute watertight evidence is very difficult to find."

There were also concerns about the inconclusive nature of research findings. Some respondents felt that reference to research evidence did not necessarily highlight the appropriate criteria to review care by. The following quote illustrates such sentiments quite graphically:

- "In a subject like neonatal sepsis the literature is very grey ... there is no consensus as to how do you diagnose sepsis early. There is work from The States, where they tried to find how often a particular sign, or a particular symptom was present. But that practice is very different than UK practice. Now if you have that evidence and you say well, in, in the States people who diagnose sepsis, let us say in the new-born, pallor was seen in 50%. You could have pallor checked as a criteria, but here we might not even take pallor into account. We might take other things into account. So it becomes very difficult."

## Resource issues

Issues surrounding the provision of resources were mentioned by the majority of interview respondents. Resources issues covered time in which to develop criteria, available skills and financial support.

### *Time factors*

A commonly cited obstacle to developing appropriate review criteria concerned the amount of time individuals had available to them. Time limitations were frequently mentioned by interview respondents drawn from the lower quartile of scores as a means of explaining their low scores:

- "The time that is required for literature search and critical review means that if we were doing that to the extent which would be ideal there would never be any audit being done."
- "Whereas actually what you need is a huge amount of time, far longer than the audit actually takes usually, to do the development work."

Such individuals suggested that some form of protected time for audit activity would be a major support in developing more appropriate criteria:

- "Things don't get done unless they're done within the working day. So we need sort of protected time to do it. It's got to be protected."

Indeed, respondents drawn from the upper quartile of scores all mentioned being given some form of protected time for quality improvement activities. This time was seen as extremely helpful in enabling individuals to adopt a thorough, systematic approach to developing review criteria:

- "The trust were very good in giving us time to go away and look at things, and they supported us management wise."
- "Yes, I, I take time. Half a day a week, or sometimes half a day in two weeks. This is a great help."

### *Available skills*

A further obstacle under the theme of resource issues concerned the degree of skill individuals had to draw upon in applying a systematic, evidence-based approach to selecting review criteria. In some cases this was described as a lack of knowledge about the audit process in general:

- "I think basically it was down to the fact that I really didn't have a great deal of knowledge what I was doing. And the person I was doing it with was in the same position. Neither myself or my colleague really had been involved in designing and doing audit before ourselves. And really hadn't had any training, I guess that's probably the biggest barrier."

Other individuals were more specific and mentioned a lack of critical appraisal skills as a major obstacle:

- "... and each of us deciding what the actual evidence is, and we're going to do that variably, because it's very difficult to look properly at the literature and weigh it up. You need people with knowledge to be able to do that, and so on and so forth."

These were obstacles that individuals felt could be overcome by provision of training:

- "There's so much research stuff out there, isn't there, that if they're going to use that then they must appraise it properly, and not just take it as

verbatim. But, we don't teach appraisal skills. I think that's in the pipeline for people who are not very confident about appraising literature. I think we're going to try and put that on in the future."
- "It could all so easily be remedied if they'd give some sort of training courses in basic audit techniques, critical appraisal and the like."

Indeed individuals who were able to apply a relatively systematic approach to selecting review criteria did espouse the virtues of training programmes:

- "What's helped here is that we've tried to, within our directorate, to incorporate into our audit meetings training, both from outside people. Also that a section of the audit meeting is not just for presenting audits, but is for planning audits as well. That's done largely in group work, so there is a training aspect there from those that have more experience than others."
- "The trust runs basic audit awareness courses continuously. So everybody's encouraged to attend those."

Other individuals commented that skills learned whilst training for external qualifications were transferable to aspects of selecting review criteria:

- "I'm doing a Masters at the moment and as part of the course we learnt the skills how to look at literature critically."
- "I did an MSc in Health Services Research. I'm particularly well read, umm, unusually well read, umm, as compared to the majority in my sort of province."

#### *Financial support*

All interview respondents acknowledged that providing training courses and protected time for audit activity had financial implications. Inadequate financial support was reported as an obstacle to the development of review criteria by preventing protected time being allocated:

- "They're doing a lot of it out of their own time, because there isn't additional funds to support clinical governance."

In addition, financial restrictions could prevent the provision of training in audit activity:

- "The other one is lack of funding. To do any of these courses I only have 25 pounds per person per year for my training budget. So it's resources as well."

Where adequate finances were available this was seen as beneficial in enabling time to be freed up for audit activity:

- "In addition to that, if they choose a specific topic then they can apply for funds to support that audit, and then they can actually have time off their work, and get in agency staff to replace them."
- "The clinical effectiveness department paid for me at a staff nurse level, for 15 hours a week. Which I think was okay. Hopefully it promoted the idea of other staff nurses taking on audit."

### Attitudes

The theme of attitudes covered instances where respondents believed obstacles or supports, to a systematic method of criteria selection, could be attributed to the beliefs and attitudes of individual practitioners undertaking audits.

#### *Resistance to change*

A commonly mentioned obstacle towards using a systematic approach to developing review criteria lay in the observation that some practitioners believe that their own practical experience and knowledge is sufficient to bypass any recourse to research evidence:

- "And I think it is tough to do science once you have dogma or religion. And we have a lot of dogma. We have a lot of dogma in our British system of medicine. You know, it is very difficult to be able to challenge somebody, and say why are you doing this? There is no evidence it works. But they'll say, oh well, we don't have evidence for everything, you know? I have done this for 20 years, and I'll continue to do this for the rest of my living days, end of story."
- "... people aren't keen on change, are they? And sometimes they're not keen to look into the evidence base, because it may mean a huge change in the way they go about their work I suppose."

Solutions offered to this kind of obstacle centred around the use of gentle persuasion to alter the perceptions of others:

- "So it's a conflict in cultures. Conflict between a couple of very old peoples, I would use the word, ritualistic culture, to one that needs to move on and look at the evidence first hand. But at the end of the day it's still about getting the best care for your patients. So it's pointing out that the background philosophy is exactly the same."

### *Negative view*

A number of respondents felt that individuals may be reluctant to undertake audit adequately, as they viewed audit in a negative manner and failed to see the benefit of it:

- "The mind set of audit has been put as 'Oh, yet another thing to do'. I think now it is changing, people are realising it's a good thing. But they haven't crossed the boundaries of doing a meaningful audit."
- "I think there is, as in many NHS, district hospitals and teaching hospitals, there's quite a degree of scepticism to audit, in what it can achieve."

These comments were more a reflection on the overall audit process, rather than direct references to review criteria development. However, it seems likely that individuals who are disposed to think ill of the audit process may not give serious consideration to the need systematically to select their review criteria. It is worth noting that negative attitudes were not felt to be impenetrable, and some respondents believed they could change over time if individuals became involved in effective audits:

- "So I think, apart from having support from the top, we're also getting the support from the bottom, because they can see all this is actually making a change. It takes time, but it actually does make change. It brings about improvements for both the clinicians and the patients."

In addition, individuals with a positive view and enthusiasm for audit were seen as instrumental, not only in lessening the influence of negative perspectives, but also by inspiring others to audit well:

- "Some of the doctors hold a very negative view of audit, so it's hard getting them to take it seriously and do it well. But again some of them are more enthusiastic. And you just hope that people's enthusiasm will drag the reluctant ones along."
- "I think without a good leader, someone who is really motivated, you know, you wouldn't get the cooperation of your colleagues."

## Organisational issues

A large proportion of respondents noted that factors associated with the institution or organisation for which they worked could impinge on the audit process, in particular review criteria development.

### *Organisational culture*

The main obstacle identified was that the culture of the institution concerned did not adequately support audit activity:

- "The problem here, I think, is the same thing all over the country, any part of the country from the North down South, is that there's a certain anti-clinical audit feeling. I don't know if you've come across it. You say audit, they say 'Fine, go ahead and just do what you want. Don't worry if it's wasting time'. Or they say 'Don't bother about it', because, it's c onsidered to be a tedious process."
- "Within our trust I think there's a big problem, I mean this is going to sound really horrible. I don't think it's supported from the top. But senior management, I don't think that's quite one of their priorities. A lot of our areas have never been pushed into doing audit. We still have specialities that don't do it."

Such comments were in stark contrast to those provided by individuals from the upper quartile of scores, who viewed their institutional culture as a supportive factor:

- "We were supported by the trust. The trust were very good in that they wanted this audit to happen as well."
- "The organisation have a commitment as far as they have a department of clinical effectiveness that believe in grounded theory, starting from the ground. And networking, and are looking at other aspects. So we've got a good supportive department who are committed to the audit process. I think the culture within the organisation is paramount."

### *Audit support staff*

Part of having a supportive institutional or organisational culture appeared to be provision of efficient audit support staff:

- "We're quite lucky here, we have an audit department, who actually are very helpful in actually helping you develop criteria."
- "There's a strong culture of a very effective audit department. There's actually an audit assistant allocated to each directorate. And, you know, helps anybody who wants to look at a project."

In some instances clinical audit departments directly encouraged the use of systematic methods of criteria selection by refusing to approve audit proposals that relied on a less thorough approach:

- "That's on our forms now for audit proposals. It's have you looked at the literature, and who's done it and so on and so forth."
- "They gave us this ideal process map, thinking how each stage related to the criteria and standards. And I mean it was just so valuable in helping you to think. You know, is each stage based on the evidence?"

#### *NHS policy*

Some individuals felt that a supportive culture had been fostered by NHS policy initiatives. The introduction of the concept of clinical governance was commonly cited:

- "The trusts themselves are very into research and audit, very much so. And you can actually prove, under clinical governance now, that the quality of service that you're giving is a good one, because you've got the evidence there."

Some individuals drawn from the lower quartile appeared confident that the clinical governance initiative would improve their likelihood of being involved in high-quality audits:

- "I think the culture is changing, and clearly with clinical governance then audit is being forced up the agenda."
- "If you asked me to fill that questionnaire out again now, I think I'd score much higher. A lot's changed in our practice now that clinical governance had been brought in. People are much more aware of needing to look at the evidence, judge the quality of the care we give and be accountable. I've been to some audit and research training classes recently and they're getting a lot more people interested and involved."

### Available protocols

A number of the individuals interviewed had not developed their own protocols, but had either used previously published protocols for their own local use or had been provided with such protocols as part of a national or regional audit. In addition, some of the individuals who had developed their own protocols forwarded opinions on involvement in national audits.

#### *National and regional audits*

A number of interview participants identified the need for more national, or at least regional, audit activities to be carried out. Concentration on small-scale, local audits was judged, by some participants, as rather futile:

- "Local ones, I think they have a downside to it. You could actually get demotivated if you carry it out and you don't see any progress. It can have a very bad demoting effect, effect on a team. So I'd be very cautious about that."
- "I did a local one to start with, but it wasn't until I got involved with the national audit, that it had more impact and helped to move things. Especially when you're going to present this data to our managers in the trust, it has a lot more impact."

Such individuals were keen to see greater reliance on national audit programmes:

- "I think there is going to be a shift to more national based audits now, anyway, so, yes, well I'll be more for that."
- "A lot of what's going on, is being part of the national picture. Working, using audits with a national focus probably has more impact than I would say local level audits can."

However, some individuals who had participated in national audits were concerned that nationally developed audit protocols did not always fully apply to local circumstances:

- "One of the problems was that we felt that some of the nationally based questionnaire items didn't really apply so we needed to refine it for local use."
- "Even when you've got consensus from national organisations, there is going to be local interpretation, and local circumstances."

#### *Convenience*

For those using published protocols, for either local use or as part of national audits, there appeared to be considerable advantages in terms of the time and effort saved developing criteria and protocols. Many respondents felt it would be inefficient to develop their own criteria if pre-existing protocols were available:

- "There's no point re-inventing the wheel when good work's been done already. And a lot of time is wasted generally I think within audit, when people are repeating other people's work."
- "So, an audit about communication and using a very similar questionnaire had been done elsewhere by our audit manager. So he had piloted that previously, had found that it worked reasonably well, so I used that. It wasn't appropriate to go into a lot of effort to developing another. Because it's been done before, I've used it."

Whilst the need to keep the time and effort required for audit to a minimum is understandable, it appeared that some individuals may have been prepared to sacrifice the quality of the audit for the sake of convenience. A number of respondents who had used previously developed protocols were unaware whether they were based on valid evidence:

- "You mean the information that their audit was based on, the research that the criteria were based on, was that specified in the pack that came? And, appraisal of the validity? I don't recall there being anything on that."
- "What we did there was to use the protocol that the ... produce. And what I did realise, I mean the audit has been done, but for some of those things it wasn't clear where they were getting their evidence from."

Although protocols may fail to include supplementary information about criteria development, for some respondents this was not seen as problematic. Some individuals preferred to place their trust in the organisation responsible for the criteria and assume they would be valid:

- "You might not be sure exactly what the evidence base was but you know that it's been done by a reputable organisation so you agree to follow it."
- "Because you know the MAAG deal with audit all the time, and, obviously they encourage people to do it. So I think you feel that, 'Oh it's come from them, it's going to be good', and perhaps it's going to save time."

### Overview

An overview of the various barriers and supports to using a systematic method in review criteria development is given in *Table 26*. The results match well with those from the quantitative data from the questionnaire survey summarised in *Table 20*. This provides confirmation for the validity of our interpretation of the data and the findings extracted from the data.

## Discussion

The interview study enabled the themes introduced by qualitative data from the questionnaire to be explored in more detail. The study provided further insight into the reasons why individuals may not have been able to apply a systematic approach to developing their review criteria. In addition, it helped to identify supportive elements that individuals believed facilitated the production of systematically selected, evidence-based criteria. Barriers and solutions were identified for these themes.

### Literature

A major element of a systematic method of criteria selection involves incorporating relevant literature. Therefore, as might be expected, all respondents

***TABLE 26*** *Interview respondents' identified barriers and supports to developing systematically selected review criteria*

| Theme | Barriers | Supports |
|---|---|---|
| Literature | Lack of literature search engines | Access to literature search engines |
| | No assistance with literature search | Assistance with literature searching |
| | Restricted access to literature | Adequate access to literature |
| | Poor quality literature | |
| | Inadequate quantity of literature | |
| Resources | Time limitations | Protected time |
| | Lack of skills | Skills training provided |
| | Financial shortage | Financial support |
| Attitudes | Resistance to change | Effective communication persuasion |
| | Negative view | Meaningful audits |
| | | Enthusiastic audit leads |
| Organisational | Non-supportive organisation | Supportive organisation |
| | Lack of audit support staff | Effective audit support staff |
| | | NHS policy initiative |
| Available protocols | Ignorance of protocol development | Convenience of published protocols |
| | Difficulties applying to local audit | National audits |

mentioned issues surrounding the literature/evidence base available for audit topics. It seems that barriers to developing evidence-based review criteria can begin at the very point of attempting to search for evidence. Limited or non-existent access to literature searching engines and databases can represent a considerable barrier. Searching for evidence is still possible without such technological aids, but it becomes a substantially more onerous task and is likely to be impracticable if it must be performed alongside an individual's clinical workload. It is worth noting that all interviewees who had high ACQ scores reported having access to literature search engines and databases.

In addition, a number of individuals who achieved high ACQ scores reported they had been given assistance in searching for relevant literature. This could be a significant factor in enabling criteria to be based on relevant evidence. Mere access to search engines does not guarantee that the search will be exhaustive and highlight all relevant evidence. Several individuals whose audits received low ACQ scores believed that they lacked either the time or skills required to perform a sufficient literature search. This implies that the introduction of labour-saving devices, such as electronic literature search engines, is of limited use without adequate training in how to make the most efficient use of such systems. In addition, even where individuals are competent in conducting literature searches they may not have sufficient time free from their clinical schedule to perform a comprehensive search. In such instances support from library staff or audit departments can prove invaluable in ensuring that an adequate literature search reveals pertinent evidence.

#### *Access*

Gaining access to search engines and staff with literature-searching experience was, in many cases, connected to access to libraries themselves. A number of respondents complained that the availability of relevant literature was restricted by the limited scope of their organisation's library. Individuals who did not have convenient access to a well-stocked library could be forced to travel in excess of 30 miles to find an adequate alternative. Even where libraries are close at hand, restrictive opening hours often rendered them inaccessible to staff. Relying on the goodwill and enthusiasm of practitioners by expecting them to travel excessive distances or to use holiday time to visit libraries is likely to blunt enthusiasm. If organisations wish to encourage evidence-based practice it seems that attention must be given to ensuring that adequate access to the evidence, as well as methods of searching it comprehensively, is facilitated.

#### *Evidence*

Even with the ability to search the literature and with access to libraries, some individuals complained that there was little or no literature concerning their area of interest. This was either attributed to a particular clinical area failing to attract research interest, or to the lack of an evidence-based approach in certain specialities. However, the lack of evidence may be overestimated. Perhaps comments about an absence of evidence are more a reflection of an individual's inability to conduct comprehensive literature searches, be it through a lack of skills or the absence of decent library facilities, rather than a genuine dearth of evidence. It may well be the case that certain topics and disciplines suffer from less research interest than others, but it seems implausible that areas exist where there is absolutely no available evidence.

It is possible that some areas exist where the quality of available literature is weak, or the literature fails to provide a clear indication of what best practice might be. Some participants did comment that some of the research published in their area involved limited sample sizes, non-randomised groups or inadequate controls. Such comments suggested that individuals might be aware of methods of critically appraising research. However, not all interview participants appeared to be quite so *au fait* with assessing the quality of research evidence. Comments such as "Fifty percent of what gets published is absolute rubbish ... absolute watertight evidence is very difficult to find" imply that individuals lack a true understanding of the process of research and its purpose in guiding clinical practice. The production of 'absolute watertight evidence' is likely to be an elusive ideal. However, insight derived from consideration of the relative strength of research evidence can effectively guide and consolidate clinical judgement and practice. Individuals who hold unrealistic expectations (e.g. that research evidence is only valuable if it provides absolute proof) risk ignoring important literature that could in fact improve the quality of care they provide for patients.

#### *Skills*

Some interview participants did acknowledge that they lacked the necessary skills to enable them to successfully appraise the relevance of research evidence. Mention was commonly made of the need for provision of training courses in critical

appraisal. In addition, individuals who scored well on the development scale of the questionnaire believed that education and training in literature searching and appraisal had enabled them to develop appropriate review criteria. Unfortunately, provision of such training has financial implications, not only for providing the necessary education, but also for freeing up the time of healthcare professionals to attend training sessions. Some individuals acknowledged that they required additional training in literature searching and evidence appraisal, but bemoaned the fact that the organisation for which they worked lacked the funds for such training. Other individuals found that external education, such as studying for higher degree programmes, had positive benefits regarding critical appraisal skills that could improve audit activity. Nonetheless, as with the provision of access to library facilities, overreliance on the goodwill of staff to organise further education and training seems unlikely to ensure that training is provided for all who would benefit.

### Resources

It is not only training programmes that require resources. Provision of information technology packages, support for literature searches and availability of adequately stocked libraries all require financial support. Unfortunately, this is not always readily available. Many respondents pointed out that inadequate financial support had implications for the development of their review criteria. This was frequently noted in connection with the provision of protected time to carry out audit activities. Many respondents considered protected time essential to the development of appropriate, evidence-based review criteria. The time taken to develop criteria systematically can be considerable if it incorporates a thorough search of the literature, critical appraisal of identified evidence, consultation with experts or colleagues, prioritisation and piloting. Some interview participants were aware that they should be using systematic methods, but reported it was impracticable for them to do so because they lacked the necessary time for audit activity. Indeed, a common response from individuals who had been able to apply a systematic approach was that funding, provided to free up their time for audit, was of immense value.

### Organisational issues

Collectively, the previously discussed barriers and supports can be seen to relate to the organisation or institution in which interview respondents worked. Certain organisations appear better equipped with the appropriate facilities and staffing levels to encourage audit and research. Indeed, a number of interview participants drawn from the upper quartile of scores believed their organisation was effective in supporting audit activity. Mention was made of organisational commitment to audit, which often included the provision of efficient audit support staff. Clinical audit departments were frequently noted as effectively facilitating high-quality audit by ensuring that an evidence-based approach was adopted in the development of criteria and by providing expertise in the planning and execution of audit activities. Conversely, a number of interview respondents drawn from the lower quartile of development scores lamented the fact that their organisational culture did not adequately support audit activity. However, many individuals felt that improvements were likely to arise from NHS policy initiatives such as clinical governance.

### Individual characteristics

As well as organisational factors that may impinge on the audit process, a number of interview participants observed that individual characteristics could serve as barriers or supports to effective audit. It was noted that individuals may be resistant to considering research evidence when selecting criteria as they prefer to rely on their own experience and clinical judgement. This may relate to individuals' lack of confidence or ability in appraising research literature. Alternatively, it may reflect resistance to change in general, rendering a reliance on habit more attractive than seeking out information that could discredit existing practice. In addition, it was noted that individuals can be resistant to the audit process because they view it as an additional burden on their workload rather than as an integral part of clinical practice. Opportunities to overcome such barriers were seen to arise when individuals were given the opportunity to participate in meaningful audits that bring about improvements in quality of care. Meaningful audits were seen to be facilitated either by the support of enthusiastic individuals or organisational backing for audit as an important element of overall clinical practice.

### National influences

Participation in a meaningful audit was viewed by some individuals as more likely to arise from involvement in national audit programmes. These respondents believed that national audits had more impact and credibility than locally based projects. They may also be more likely to involve systematically selected review criteria than local audits, as protocols are produced by a national organisation with the available time and expertise

to develop appropriate evidence-based criteria. However, other respondents were less convinced and believed that nationally developed protocols did not always apply to local conditions. It is difficult to determine if this was genuinely the case or whether individuals would prefer to believe that local idiosyncrasies were responsible for criteria-based performance rather than view it as a reflection of substandard practice. What did emerge from many of the interviews is that a major attraction of national audits is that the review criteria are developed by an outside body, thereby limiting the degree of effort required by the actual auditor. This argument was also used in justifications for the use of published protocols. However, the need for convenience may also, unfortunately, be associated with complacency. Many of the individuals using published protocols, either locally or nationally, were unaware how the criteria had actually been developed. Assumptions were made that because protocols were produced by a reputable organisation they must inevitably involve a systematic method of criteria selection. Often these assumptions were not grounded on any evidence, as protocols did not include supplementary information about criteria development. If individuals are driven by convenience to use published protocols then it seems vital that such documents are as transparent as possible about their development process, so that judgements can be made about their validity and their likelihood to lead to improvements in practice.

## Summary

The interview study identified many barriers to using effective, systematic methods of developing review criteria, but was also able to identify ways in which these may be overcome. There is no simple solution to ensuring that a systematic approach to criteria selection occurs. However, there does appear to be a general pattern of factors associated with individuals who, according to questionnaire-based assessment, were able to develop appropriate review criteria. These factors arise at both an individual and organisational level, and indeed influence each other. At the level of the individual it appears that level of skill in literature searching and critical appraisal are important elements in ensuring that relevant evidence is considered. Acquiring the required knowledge to search literature effectively and consider its relative strength also arises at an individual level where people are motivated to apply for training places or register for further education courses. However, not all personnel may be so highly motivated. Therefore, it seems that the organisation also has an important role to play in ensuring that adequate training is provided and uptake is encouraged. The organisation can also provide support by ensuring that library facilities are of a high standard and individuals are assisted by electronic or staff services in searching for relevant evidence. Organisationally, a view of audit as an integral part of clinical practice can also ensure that individuals are given the necessary time carefully to select criteria and to carry out a meaningful audit. Of course meeting each of these needs requires considerable financial support, but perhaps costs incurred could be offset by adopting a view of audit as part of the whole process of healthcare provision. Monies directed towards improving library facilities, training staff and supporting audit and research activities are likely to reap long-term benefits in terms of the effectiveness and efficiency of healthcare.

# Chapter 5

# Conclusions, implications and recommendations

The overall aims and objectives of this research programme have been achieved. The expert consensus process was successful in producing a definitive list of desirable characteristics for quality in review criteria. This definition should be of relevance to all those involved in the development of review criteria for use in quality improvement activities. It represents a considerable improvement in understanding the characteristics of appropriate review criteria and how they might be developed. Therefore, the definition has practical implications, as using it to guide the selection of review criteria should ensure that appropriate criteria are effectively identified and presented so that data can be collected on justifiable and valid aspects of care.

The conclusions are presented in the following section in categories ordered by their importance. The implications of the conclusions are then discussed followed by further research.

## Conclusions

The use of the ACQ in the main survey phase of the project revealed that methods of review criteria selection in England and Wales often leave much to be desired. A significant proportion of audits did not appear to involve any consideration of research evidence in the process of developing criteria. Even where review criteria development was reported to involve reference to research literature, only a limited number of respondents had attempted to assess the quality of the literature, both in terms of its recency and in terms of the validity of the reported findings.

However, it is worth noting that protocols provided for national audits, and published protocols in general, often fail to include accompanying information about the development of review criteria. This makes it almost impossible for individuals carrying out quality reviews to assess the appropriateness of the review criteria involved. To ensure that review criteria are valid it is essential to have details of the evidence on which they are based, the quality of the evidence, the reasons behind any prioritisation, and so on. Therefore, it is of paramount importance that the developers of published audit protocols are encouraged to include a detailed and transparent account of how their review criteria were selected.

A national resource of available protocols, which can be shown to demonstrate a high proportion of the desirable characteristics of review criteria, would meet the needs of audit leads in many instances. This would also improve the effectiveness of audits if high-quality review criteria were readily available for use.

Organisationally, important factors were the provision of adequate library and information technology services, skill training in audit techniques and critical appraisal, financial support for protected time to audit by, and a supportive organisational culture that encourages a view of audit as a central part of clinical practice. Funds directed at improving library facilities, training, and supporting audit and research activities were seen as a worthwhile investment to improve the quality of healthcare.

Interviewees proposed a move to greater reliance on large-scale national audit programmes rather than localised audits. This suggestion is supported by our finding that small-scale, local audits were more likely to involve audit leads developing their own criteria, which generally received significantly lower development scores than those drawn from published sources. In addition, a trend was observed for single-organisation audits to be associated with lower development scores than those given to multi-organisation audits. This adds further weight to the view that large-scale national audits may be more desirable than local audits, the insular focus of which could have negative implications for the quality of clinical audit carried out.

Interview results indicated that a number of factors at both individual and organisational level could contribute to the development of appropriate, systematically selected review criteria. Individual factors associated with the development of appropriate review criteria included the

enthusiasm of audit leads and their motivation to organise audits effectively and register for relevant training.

Further aspects of a systematic approach to developing review criteria, such as piloting and prioritising criteria and consulting with experts, were also frequently neglected. Therefore, it was scarcely surprising that most questionnaire responses produced low scores for the development of review criteria. Qualitative data drawn from the questionnaire suggested that many individuals were aware that they could have been more systematic in their approach to developing review criteria, but certain obstacles had prevented them from being so. In many cases, respondents were aware of strategies that should enable them to overcome these obstacles.

No differences were found between review criteria used in audit in general practice and in those used in audit in trusts. No differences were found between review criteria used in audits of clinical or non-clinical topics.

In conclusion, the main objectives of the project were achieved. A definition of a systematic method of selecting review criteria now exists and has enabled the creation of a valid questionnaire to identify the degree to which systematic methods are used to select review criteria. The questionnaire enabled the project team to identify the extent to which systematic methods are used to select review criteria in clinical audits in England and Wales. Qualitative questionnaire data also provided some insight into obstacles to using systematic methods and strategies to overcome such obstacles. This insight was consolidated and strengthened by detailed interviews with a selection of questionnaire respondents. The overall findings should enhance our understanding of how review criteria are selected and provide recommendations for improved practice.

## Implications

We now have an estimate of how good the review criteria are that are in daily use in clinical audits in England and Wales. The mean ACQ score of 0.52 indicates that almost half the desirable characteristics of review criteria are missing from those used in clinical audits. There is clearly room for improvement. The respondents in this study have identified the problems in selecting review criteria, and provided solutions to those problems. The respondents are the experienced, enthusiastic auditors in the NHS and they have given here the potential solutions to the clearly identified obstacles. This shows the huge potential for improving the selection of clinical audit criteria, which will directly and immediately increase the effectiveness of clinical audits. There is a high risk that audits which are conducted using inappropriate criteria may be wasting time and resources in measuring and changing things that do not actually improve the quality of care.

The major problems lie in the quality of the process by which review criteria are developed. In particular, recent high-quality evidence is rarely used to select review criteria. This is possibly because the literature is difficult and the skills in using the literature to create review criteria are lacking and difficult to acquire. A national resource of review criteria for clinical audit, which has all the desirable characteristics of review criteria, would overcome some of those difficulties. The criteria would be based on informed assessment of the literature, and kept up to date, with full provenance reported.

These review criteria would also be based on consultation with patients and experts, both on the importance of criteria and on the demands made by collecting the relevant data. It could be argued that users do not need to know all the details of the development and selection of review criteria, as they only need to know how to use them. Of course, if users are to make informed decisions on the criteria they are to select, then this information is essential.

This study has shown that practitioners can identify the real obstacles to selecting review criteria systematically, and others have found successful ways to overcome those obstacles. The obstacles frequently identified were lack of access to the evidence, lack of skills in using the evidence to derive review criteria, lack of time to undertake this task, and lack of support either to develop these necessary skills or to have the service provided by specialists.

The solutions identified were mostly provision of the items shown as lacking in the list of obstacles. These would include libraries, information technology, information technology support, review criteria developers, transparency of published criteria, and training in review criteria development, including literature review skills. These supports would make the recognition of enthusiastic leaders more likely and more tenable, which,

our expert respondents tell us, would enhance the quality of audits. Organisations such as NHS trusts have the opportunity to support clinical governance by valuing the selection of good review criteria for clinical audits, and thus encouraging the improvement and maintenance of high-quality healthcare provision. A further potential solution would be a simple tool with which review criteria could be assessed for their quality. Such a tool could be used by those starting an audit, in order to make an informed selection from published criteria. Another use would be for those developing their own criteria to assess the quality of the criteria they have created.

The task of developing, assessing or selecting audit review criteria may be a responsibility for organisational clinical governance staff, who would have the necessary skills and resources to perform these functions. The responders to our questionnaire and interviews showed much skill, enthusiasm and willingness to improve the selection of audit review criteria in their own conduct of audits. These attributes can be harnessed, valued and resourced by organisations as part of their clinical governance agendas. There were also responders who did not consider it important to use high-quality research evidence in selecting or developing their review criteria. This obstacle of being unwilling to recognise the need for change may need a number of interventions, tailored to individual situations to be overcome.

These findings may inform future policy development relating to the implementation of clinical governance, national service frameworks, clinical guidelines and evidence-based practice. This is of particular relevance to the remit of the national institute for clinical excellence.

## Recommendations for further research

- Trials of interventions designed to improve the selection of review criteria for clinical audit. The ACQ developed in this study could be used as an outcome measure for such trials. One such intervention could be the creation of a library of review criteria that have all the desirable characteristics.
- The development and validation of a simple tool by which review criteria can be assessed. This should be based on the expert consensus view of the desirable characteristics of review criteria, reported in chapter 2 of this report.
- Testing the relative effects on the quality of patient care, of national or regional audits compared with local audits, and their respective quality of review criteria, would inform policy-makers on the implementation of audit.
- Case studies of organisations where selection of review criteria is given appropriate importance and resources, would identify the organisational policies that enable and maintain this.

# Acknowledgements

This study was commissioned by the NHS R&D HTA Programme. The authors are indebted to the HTA referees for their perseverance in reading this report and the quality of their comments. The views expressed in this report are those of the authors, who are responsible for any errors.

The authors are grateful to the experts who contributed to the consensus development, to the audit coordinators who distributed the questionnaires and all those who gave their time for an interview.

# References


1. Pharoah PD, Melzer D. Variation in prescribing of hypnotics, axiolytics and antidepressants between 61 general practices. *British Journal of General Practice* 1995;**45**(Pt 400):595–9.

2. Selby JV, Fireman BH, Lundstrom RJ, Swain BE, Truman AF, Wong CC, *et al.* Variation among hospitals in coronary–angiography practices and outcomes after myocardial infarction in a large health mainte-nance organization. *New England Journal of Medicine* 1996;**335**(Pt 25):1888–96.

3. Tobler HG, Sethi GK, Grover FL, Shroyer AL, Moritz TE, Henderson WG, *et al.* Variations in processes and structures of cardiac surgery practice. *Medical Care* 1995; **33**(Pt 10)Suppl: OS43–58.

4. Department of Health and Human Services. The challenge and potential for assuring quality health care for the 21st century. 1998. Domestic Policy Council, Washington, DC: http:www.ahcpr.gov/qual/21stcena.htm

5. Department of Health. A first class service: quality in the new NHS. London: HMSO; 1998.

6. Department of Health. The new NHS: modern, dependable. London: HMSO; 1997.

7. Fraser RC, Khunti K, Baker R, Lakhani M. Effective audit in general practice: a method for systematically developing audit protocols containing evidence-based review criteria. *British Journal of General Practice* 1997;**47**(Pt 424):743–6.

8. Teasdale S. The future of clinical audit: learning to work together. *British Medical Journal* 1996; **313**:574.

9. Berger A. Action on clinical audit: progress report. *British Medical Journal* 1998;**316**(Pt 7148):1893–4.

10. Davis DA, Taylor-Vaisey A. Translating guidelines into practice – a systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *Canadian Medical Association Journal* 1997;**157**(Pt 4):408–16.

11. Scally G, Donaldson LJ. Clinical governance and the drive for quality improvement in the new NHS in England. *British Medical Journal* 1998; **317**(Pt 7150):61–5.

12. Field M, Lohr KN. Guidelines for clinical practice: From development to use. Washington: Institute of Medicine; 1992.

13. Naylor CD, Guyatt GH. Users' guides to the medical literature. XI. How to use an article about a clinical utilization review. *Journal of the American Medical Association* 1996;**275**(Pt 18):1435–9.

14. Lohr KN. Medicare: a strategy for quality assurance. Vol. I. Washington: National Academy; 1990.

15. Lanska D. A public/private partnership in the quest for quality: development of cerebrovascular disease practice guidelines and review criteria. *American Journal of Medical Quality* 1995;**10**(Pt 2):100–6.

16. James PA, Cowan TM, Graham RP, Majeroni BA, Fox CH, Ja'en CR. Using a clinical practice guideline for the management of heart failure. *Journal of the American Board of Family Practitioners* 1997; **10**(Pt 3):206–12.

17. Hadorn DC, Baker RW, Kamberg CJ, Brooks RH. Phase II of the AHCPR-sponsored heart failure guideline: translating practice recommendations into review criteria. *Joint Commission Journal on Quality Improvement* 1996;**22**(Pt 4):265–76.

18. Cluzeau F, Littlejohns P, Grimshaw J, Hopkins A. Appraising clinical guidelines and development of criteria – a pilot study. *Journal of Interprofessional Care* 1995;**9**:227–9.

19. Ashton CM, Kuykendall DH, Johnson ML, Wun CC, Wray NP, Carr MJ, *et al.* A method of developing and weighting explicit process of care criteria for quality assessment. *Medical Care* 1994; **32**(Pt 8):755–70.

20. Bateman DN, Soutter J, Eccles M, Roberts SJ, Campbell M, Smith JM. Setting standards of prescribing performance in primary care: use of a consensus group of general practitioners and application of standards to practices in the north of England. *British Journal of General Practice* 1996; **46**(Pt 402):20–5.

21. Johnson N. Feasibility of developing and selecting criteria for the assessment of clinical performance. *British Journal of General Practice* 1993;**43**:499–502.

22. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *International Journal of Technology Assessment in Healthcare* 1986;**2**(Pt 1):53–63.

23. Campbell SM, Roland MO, Shekelle PG, Cantrill JA, Buetow SA, Cragg DK. Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in, general practice. *Quality in Health Care* 1999;**8**(Pt 1):6–15.

24. Derry J, Lawrence M, Griew K, Anderson J, Humphreys J, Panher K. Auditing audits: the method of the Oxfordshire Medical Audit Advisory Group. *British Medical Journal* 1991; **303**(Pt 6812):1520–2.

25. Walshe K, Spurgeon P. Clinical audit assessment framework handbook. Series 24. Birmingham: Health Services Management Centre, University of Birmingham; 1997.

26. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al.* Consensus development methods and their use in clinical guideline development. *Health Technology Assessment* 1998; **2**(Pt 3).

27. Turoff M. The policy Delphi. In: Linstone HA, Turoff M, editors. The Delphi method: techniques and applications. London: Addison-Wesley; 1975. p. 84–101.

28. Critcher C, Gladstone B. Utilizing the Delphi technique in policy discussion: a case study of a privatised utility in Britain. *Public Administration* 1998;**76**:431–49.

29. McKnight JN, Edwards N, Pickard L, Underwood J, Voorberg N, Woodcox V. The Delphi approach to strategic planning. *Nursing Management* 1991; **22**(Pt 4):55–7.

30. Mitchell VW. Using Delphi to forecast in new technology industries. *Market Intelligence and Planning* 1992;**10**:4–9.

31. Guba ES, Lincoln YS. Fourth generation evaluation. London: Sage; 1989.

# Appendix 1

## List of expert participants

Affiliations are given as at the time of the study.

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

# Appendix 2

## The audit criteria questionnaire

THE UNIVERSITY OF WARWICK

## Audit Criteria Questionnaire

For this questionnaire, we would like you to think of **one audit only**. It should be one which you have been involved in and which started in the last 12 months. (We are asking everyone to think of a recent audit so that the details may be better remembered, and so they will all be from the same period of NHS changes). If there is more than one audit for you to choose from, please think of the most recent one.

Please give the title or topic of the audit _____

Please tell us the clinical area(s) covered (*eg urology, intensive care nursing, physiotherapy, general medicine, interface with audit in other sectors*)

_____

Would you describe this audit as: (Please tick)

National ☐ Run with other trusts/practices in your area ☐

Regional ☐ Limited to your organisation only ☐

The questions are about the audit criteria used for your audit. "Audit criteria" are aspects of care which are assessed in clinical audits and their presence or absence gives an indication of quality of care.

**How were the audit criteria for your audit selected?** Please tick all which apply.

a. Taken from published audit protocols ☐
Please give the name of the protocol and the publisher
*(eg Royal College of Physicians clinical audit on long term care of elderly people)*

_____

b. Taken from published clinical guidelines ☐
Please give the name of the guideline and the publisher
*(eg Royal College of Radiologists guidelines for referrals for radiology )*

_____

c.Provided by audit support staff ☐
Please give the name (eg *Primary care audit group, hospital clinical governance department)*

_____

d.Developed your own for this audit ☐

e.Taken from a previous audit you were involved in ☐

f. Other method, please describe ☐

It would be interesting to see any written information, or documentation, you may have on the audit criteria. If possible, please provide us with a copy when you return your questionnaire.



### Section 1 The development of the audit criteria

Even if you were not involved in developing the criteria yourself, please tell us what you know about the criteria for your audit, perhaps from information provided with the criteria in the protocol or publication. Please tick the relevant box in response to each question.

| 1.1 Were the audit criteria based upon: | Yes | Partly | No | Don't know |
|---|---|---|---|---|
| a. searching the research literature? | | | | |
| b. consultation with experts? | | | | |
| c. consultation with local experts? | | | | |
| d. consultation with colleagues? | | | | |
| e. consultation with patients or carers? | | | | |
| f. criteria used in previous audits? | | | | |
| Other, please describe | | | | |

If you answered **NO** to question 1.1.a. please continue to question 1.3 overleaf

If you answered **YES** or **PARTLY** to question 1.1.a. – i.e. criteria were based on searching the research literature,

1.2. who carried out the literature search? (Please tick)

a. You or your colleagues ☐

b. Others *(e.g. National Centre for Clinical Audit, or a library information service)* ☐
Please give details: _____

c. Criteria were based on a published systematic review ☐
Please give details of author(s), title and publication date:

_____

| | Less than 1 year old | Between 1 and 3 years old | Over 4 years old | Don't know |
|---|---|---|---|---|
| d. How up to date was the literature review? | | | | |

| | Yes | No |
|---|---|---|
| e. Did you search for additional, more recent literature? | | |

| Was the following information recorded (by you or the authors of the review): | Yes | No |
|---|---|---|
| f. the sources/databases used to identify the literature? | | |
| g. whether the validity of the research was appraised? | | |
| h. the methods used to assess validity | | |

| | Yes | Partly | No | Not applicable |
|---|---|---|---|---|
| 1.3. Is the method of combining evidence-from-literature and expert-opinion made explicit? | | | | |

| | Yes | Partly | No |
|---|---|---|---|
| 1.4. Is the method used to select the audit criteria described in enough detail to be repeated? | | | |

| | | | | Don't know | Not applicable |
|---|---|---|---|---|---|
| 1.5. Were the audit criteria pilot tested for practical feasibility? | | | | | |
| 1.6. Were the audit criteria prioritised on: | | | | | |
| a. impact on health outcome? | | | | | |
| b. quality of supporting evidence? | | | | | |
| c. cost implications? | | | | | |
| d. importance to patients? | | | | | |
| Other priorities, please describe | | | | | |

| | Yes | Partly | No | Don't know | Not applicable |
|---|---|---|---|---|---|
| 1.7. Were the relative values of harms and benefits associated with treatment options considered in selecting criteria? | | | | | |

Please describe any problems you may have had with the development of the audit criteria.

If there were problems – what, if anything, were you able to do about them?

Appendix 2 Audit Criteria Questionnaire (final version)

**Section 2 The attributes of the audit criteria**

| 2.1 Do the criteria | Yes, all criteria do | Some do, some don't | No criteria do | Not applicable |
|---|---|---|---|---|
| a. state the patient populations to which they apply? | | | | |
| b. state the clinical settings to which they apply? | | | | |
| c. give clear definitions of the variables to be measured? | | | | |
| d. use unambiguous terms? | | | | |
| 2.2 Are the criteria linked to improving health outcomes (rather than, say, to reducing costs or increasing throughput)? | | | | |
| 2.3 Do the criteria enable you to differentiate between appropriate and inappropriate care? | | | | |

**Section 3 Usability of audit criteria**

| | Yes | No | Not applicable | Don't know |
|---|---|---|---|---|
| 3.1 Did the criteria have information on: | | | | |
| a. the time period over which data should be collected? | | | | |
| b. the amount of data which should be collected (e.g. sample size)? | | | | |
| c. any special skills needed to collect the data? | | | | |
| d. how the demands of the audit on patients might be minimised? | | | | |
| e. how the demands of the audit on staff might be minimised? | | | | |
| 3.2 Did the criteria have clear instructions for using them? | | | | |
| 3.3 Were patients consulted about the acceptability of these criteria for them? | | | | |
| 3.4 Were all relevant staff consulted about the acceptability of these criteria for them? | | | | |

**Appendix 2 Audit Criteria Questionnaire (final version)**

Please describe any problems you may have had with the attributes or usability of the audit criteria.

If there were problems – what, if anything, were you able to do about them?

Any other comments on aspects which we have not asked about:

Thank you for reaching this far! As a further stage in the research we shall be interviewing some respondents about problems and successes in selecting audit criteria. The interviews will take place later this year (by telephone or face to face) and will last around 30 minutes. We'd be very grateful if you would give us your name and a contact phone number so we can reach you to discuss arranging an interview. Thank you.

Name:_____ Tel:_____

Please return the questionnaire, and any additional documentation about your audit criteria, direct to the MARC project team (Centre for Primary Health Care Studies, University of Warwick, Coventry CV4 7AL) by using the reply paid envelope.

# Appendix 3

## Letter to the PCAG coordinators

*«Title» «Forename» «Surname»*
*«Job_Title»*
*«Organisation»*
*«Address1»*
*«Address2»*
*«Address3»*
*«Town»*
*«County»*
*«Postcode»*

Dear «Title» «Surname»,

A team from the Primary Care Unit, University of Warwick, and the Eli Lilly National Clinical Audit Centre, University of Leicester is studying the methods used to select criteria against which clinical practice can be reviewed. The NHS Research and Development, Health Technology Assessment Programme is funding the project.

We have developed a short questionnaire for people who are conducting audits in a random sample from all NHS primary care audit groups and Trusts in England and Wales. The questions are about the method and process of selecting audit review criteria, including any difficulties experienced and strategies used to overcome them.

Your audit group has been selected for our sample. You represent an important link through which we can identify audit leads. We would appreciate your help in contacting individuals who are conducting audits within your area. There are two alternative ways in which you could help us.

Either

1. Provide us with a list of all practices attached to your audit group, giving us the name and contact details of someone involved in audit within each practice. The research team will write to each practice and ask the named person to complete a short questionnaire (stating that you gave us their name). Please use copies of the attached blue form to list details of practices in your area, or use your own form if that is easier for you, or send us an electronic mailing list.

Or

2. You agree to distribute a questionnaire to each practice on our behalf. We shall provide you with the questionnaires in envelopes with covering letters and reply paid envelopes, with a copy for you to see. If you prefer this option, please complete the enclosed yellow form giving us details of the number of practices involved.

**p.t.o.**



**Appendix 3: Letter to PCAG co-ordinators**

We would be very grateful for your assistance in this project. If you are able to participate, please return either a) a list of all practices in your audit group, or b) a yellow form. A FREEPOST envelope is enclosed for this purpose, or you can fax the details to us on 01203 528375.

Even if you are not able to participate it would be very helpful to us if you could return the yellow form.

We look forward to hearing from you soon.

Many thanks

Project team:
**Hilary Hearnshaw**
**Rachael Harker**
**Gill Grimshaw**
University of Warwick

**Richard Baker**
**Francine Cheater**
University of Leicester

If you have any queries or require further information please contact:
Dr Rachael Harker:
Tel - 01203 572941
Email – R.M.Harker@warwick.ac.uk

# Appendix 4

## Letter to the NHS trust audit coordinators

**Appendix 4: Letter to Trust audit co-ordinators**

*«TITLE» «FORENAME» «SURNAME»*
*«JOB_TITLE»*
*«ORGNAME»*
*«ADDRESS1»*
*«ADDRESS2»*
*«ADDRESS3»*
*«TOWN»*
*«COUNTY»*
*«POSTCODE»*

Dear «TITLE» «SURNAME»,

A team from Universities of Warwick and Leicester is studying the methods used to select criteria against which clinical practice can be reviewed. The NHS Research and Development, Health Technology Assessment Programme is funding the project.

We have developed a short questionnaire for people who are conducting audits in a random sample from all NHS Trusts and primary care audit groups in England and Wales. The questions are about the method and process of selecting audit review criteria, including any difficulties experienced and strategies used to overcome them.

Your trust has been selected for our sample. You represent an important link through which we can contact audit leads. We would appreciate your help in identifying 10 ongoing audits within your Trust and the appropriate people to send our questionnaire to.

By an "ongoing audit" we mean one in which the first data collection has begun, or is complete but the follow up data collection has not finished, **and** it is less than 12 months since the beginning of the first data collection.

Once you have identified audits in your trust that meet this definition there are two alternative ways in which you could help us.

Either

1. Provide us with a complete list of all ongoing audits in your trust so that we can select a random sample of 10 audit leads to send a questionnaire to. For this we will need the name and contact details of the lead person for each audit. Please use copies of the attached green form to list details of ongoing audits in your trust, or use your own form if that is easier for you.

**p.t.o.**

Or

2. You agree to give out questionnaires for us, to the lead people in 10 ongoing audits in your trust. We shall provide you with the questionnaires in envelopes with covering letters and reply paid envelopes, with a copy for you to see. Please note that this alternative may involve you in chasing non-responders for us, since we will not know who has been given the questionnaire. If you prefer this option, please complete the enclosed pink form.

We would be very grateful for your assistance in this project. If you are able to participate, please return either a) a list of audit leads, or b) a pink form. A FREEPOST envelope is enclosed for this purpose, or you can fax the details to us on 01203 528375.

Even if you are not able to participate it would be very helpful to us if you could return the pink form.

We look forward to hearing from you soon.

Many thanks

Project team:
**Hilary Hearnshaw**
**Rachael Harker**
**Gill Grimshaw**
University of Warwick

**Richard Baker**
**Francine Cheater**
University of Leicester

If you have any queries or require further information please contact:
Dr Rachael Harker:
Tel - 01203 572941
Email – R.M.Harker@warwick.ac.uk

# Appendix 5

## The PCAG coordinator response form

**Appendix 5: PCAG co-ordinator response form**

SECTION A:
Please check your personal details and inform us of any errors

«Title» «Forename» «Surname»
«Job_Title»
«Organisation»
«Address1»
«Address2»
«Address3»
«Town»
«County»
«Postcode»
«Tel»

SECTION B:

**I agree** to pass on a questionnaire to all practices in my area

Signature:_____ Date:_____

Number of practices:_____

Thank you for agreeing to help in this way. We will be in touch soon with copies of the questionnaire for you to pass on.

SECTION C:

**If you are not able to take part**, it would be very useful for us to know why. Please can you answer the following section by ticking any boxes that apply.

I did not agree to take part in the research because:

a. I cannot spare the time to contact practices ☐

b. I do not think practices in my area will have time to take part ☐

c. I was not given enough information to allow me to make an informed choice ☐

d. Other (please give details)

_____
_____
_____
_____
_____
_____

# Appendix 6

## The NHS trust coordinator response form

**Appendix 6: Trust co-ordinator response form**

SECTION A:
Please check your personal details and inform us of any errors. Then tell us a little more about the trust you work in.

«TITLE» «FORENAME» «SURNAME»
«JOB_TITLE»
«ORGNAME»
«ADDRESS1»
«ADDRESS2»
«ADDRESS3»
«TOWN»
«COUNTY»
«POSTCODE»

About your Trust:

Type of service (e.g. Acute, Mental Health):_____

Number of staff employed by Trust (Approx.):_____
_____

SECTION B:

**I agree** to pass on a questionnaire to 10 audit leads in my Trust

Signature:_____ Date:_____

Thank you for agreeing to help in this way. We will be in touch soon with copies of the questionnaire for you to pass on.

SECTION C:

**If you are not able to take part**, it would be very useful for us to know why. Please can you answer the following section by ticking any boxes that apply.

I did not agree to take part in the research because:

a. I cannot spare the time to contact audit leads ☐

b. I do not think audit leads in my Trust will have time to take part ☐

c. I was not given enough information to allow me to make an informed choice ☐

d. Other (please give details)

_____
_____
_____
_____

# Appendix 7

## Questionnaire data coding system: Cohen's κ statistics

**DEVELOPMENT PROBLEMS: KAPPA STATISTICS (3 way)**

| Kappa | Code | Title | Examples |
|---|---|---|---|
| 0.91 | D1 | Focussing the audit | Narrowing down criteria sets, refining scope of audit |
| 0.87 | D2 | Different Perspectives | Problems co-ordinating different staff groups, professions, directorates etc |
| 0.81 | D3 | Data Access | Ease of access to required information |
| 0.79 | D4 | Literature access | Lack of access to literature or guidelines to develop criteria from |
| 0.86 | D5 | Sampling | Problems with selecting, or identifying an adequate sample |
| 0.87 | D6 | Standards | Problems identifying acceptable/realistic standards |
| 0.79 | D7 | Clarity | Problems creating clearly defined, unambiguous criteria |
| 0.76 | D8 | Organising the audit | Problems setting up audit due to bureaucracy, co-ordination of different staff, etc |
| 0.93 | D9 | Time | Criteria development took up excessive time |
| 0.98 | D10 | Financial | Cost of developing criteria raised |
| 0.97 | D11 | Appropriate criteria | Problems developing objective criteria that collect the right type of data for audit |
| 0.93 | D12 | Literature availability | Lack of up to date, or non-existence of, evidence on the topic |
| 0.86 | D13 | Local interpretation | Problems interpreting guidelines/protocols to local understanding |

Appendix 7: Questionnaire data coding scheme – Cohen's Kappa statistics

| | | | |
|---|---|---|---|
| 0.86 | D14 | Skill shortage | Lack of insight/training |
| 0.81 | D15 | Motivation | Problems motivating practitioners |
| 0.83 | D16 | No problems/Not applicable | Respondents specifically state that they did not experience problems, or the response does not relate to review criteria development |

**USABILITY PROBLEMS: KAPPA STATISTICS (3 way)**

| Kappa | Code | Title | Examples |
|---|---|---|---|
| 0.98 | U1 | Validity of data collected | inaccurate records, wrong types of data collected |
| 0.87 | U2 | Ambiguity of criteria | unclear criteria open to misinterpretation |
| 0.94 | U3 | Sampling problems | small or unrepresentative samples, audit may have included inappropriate cases |
| 0.93 | U4 | Time demands | time needed to audit, conflict with other duties |
| 0.78 | U5 | Organisational | delays caused by bureaucracy, co-ordination of different staff, clinical directorates etc |
| 0.80 | U6 | Compliance with audit | practitioners fail to provide necessary information, reluctance to participate in |

| | | | |
|---|---|---|---|
| | | | audit |
| 0.83 | U7 | Data availability | accessibility of required information, problems recording data needed for audit, incomplete or un-located records. |
| 0.89 | U8 | Financial | problems funding the audit |
| 0.71 | U9 | Inappropriate criteria | Criteria felt to be too strict/rigid, failed to reflect reality. Audit tool difficult to apply to patient groups. |
| 0.71 | U10 | Setting standards | Difficulties setting an appropriate standard |
| 0.91 | U11 | No problems/Not applicable | Respondents specifically state that they did not experience problems, or the response does not relate to review criteria development |

Appendix 7: Questionnaire data coding scheme – Cohen's Kappa statistics

**OTHER COMMENTS: KAPPA STATISTICS (3 way)**

| Kappa | Code | Title | Examples |
|---|---|---|---|
| 0.86 | 0C1 | Attitude to audit | Audit viewed as effective method of quality improvement improvement, led to changes in practice that are seen as beneficial |
| 0.80 | OC2 | Negative outlook | Audit process seen as waste of time and resources, failed to effect change, failed to complete |
| 0.96 | OC3 | Protected time | Need for protected time to audit noted |
| 0.91 | OC4 | Funding | Importance of providing financial support for audit activities noted |
| 0.72 | OC5 | Clarity | Need to provide clear protocols, focus audit appropriately noted |
| 0.95 | OC6 | Standards | Importance of identifying appropriate standard levels of performance noted |
| 0.95 | OC7 | Skills training | Mention of need for training in literature searching, audit methods etc |
| 0.95 | OC8 | Literature base | Importance of access to and availability of relevant literature noted |
| 0.94 | OC9 | Motivation | Importance of motivating staff involved in audit noted |
| 1.0 | OC10 | Co-ordination of staff | Importance of involving all relevant staff, integrating different perspectives |
| 0.83 | OC11 | Patient factor | Issues of changing patient behaviour/attitudes noted |

**Appendix 7: Questionnaire data coding scheme – Cohen's Kappa statistics**

| | | | |
|---|---|---|---|
| 0.87 | OC12 | Data availability | Issues surrounding access to required data noted |
| 1.0 | OC13 | Practical advantages | Aids to data collection noted |
| 0.96 | OC14 | Adapting criteria | Adaptation, further development of, criteria for local use |
| 0.85 | OC15 | Comment on questionnaire | Comments on time taken, difficulty of questionnaire |
| 0.99 | OC16 | Solutions suggested | Comments backed up by suggested strategies for improvement. |
| 0.77 | OC17 | Other | |

# Appendix 8

## Sample interview schedule

**Respondents' Interview Schedule**

Respondent: **X**
Remind respondent of reason for interview:

You kindly completed one of our questionnaires on the review criteria you used in an audit of _____ The questionnaire looked at how review criteria were selected and developed, as well as how easy or difficult they were to use in an audit. The purpose of this interview is to discuss in more detail the way you selected and developed the review criteria for this audit, as well as whether you felt the criteria were appropriate.

You may remember we sent you this feedback sheet which summarised the overall results of the questionnaire study. If you look at the scores for the development of the review criteria, you can see that they range from 0 to 0.92 with the average score being around 0.42.

Development questions largely dealt with whether criteria were explicitly evidence-based. Higher scores indicated greater reliance on evidence.

Your responses gave you a score of **0.76** for the development of the criteria. This is higher/lower than the average score. Do you think that's a fair assessment for the criteria you used?

**Individual questions**

1a. Criteria based on searching research literature = 1

1.2d Up to date = 1

1.2e Did you search for more recent literature = 1

1.2f Sources to identify literature recorded = 1

1.2g Validity appraised = 1

1.2h Methods of assessing validity recorded = 1

1.3 Synthesis of evidence and expert made explicit = 0.5

1.4 Repeatable = 0.5

Appendix 8: Sample interview schedule

1.5 Were criteria pilot tested = 1

1.6a Prioritised on impact on health outcome = 1

1.6b Prioritised on quality of evidence = 0.5

1.6c Prioritised according to cost implications = 0.5

1.6e Prioritised according to patient importance = 1

1.7 Tradeoff considered = 0.5

Problems experienced in developing criteria:
Trying to reach consensus agreement

Solutions outlined:
Discussion and negotiation, settling for minority agreement

*Can you mention anything (else) that made it difficult for you to select evidence-based criteria? [Prompt: For instance, some people in our study mentioned lack of reliable evidence, limited time to carry out a literature search]*

If mentioned – Do you think anything could have been done to overcome these obstacles?
Were any of these strategies available to you?

*b. Can you think of anything else that would help you select evidence-based criteria? [Prompt: For instance, some people in our study mentioned help with literature searching, protected time for audit]*

If mentioned – Are any of these supports available to you?

# Methodology Group

**Members**

**Methodology Programme Director**
**Professor Richard Lilford**
Director of Research and Development
NHS Executive – West Midlands, Birmingham

**Chair**
**Professor Martin Buxton**
Director, Health Economics Research Group
Brunel University, Uxbridge

Professor Douglas Altman
Professor of Statistics in Medicine
University of Oxford

Dr David Armstrong
Reader in Sociology as Applied to Medicine
King's College, London

Professor Nicholas Black
Professor of Health Services Research
London School of Hygiene & Tropical Medicine

Professor Ann Bowling
Professor of Health Services Research
University College London Medical School

Professor David Chadwick
Professor of Neurology
The Walton Centre for Neurology & Neurosurgery
Liverpool

Dr Mike Clarke
Associate Director (Research)
UK Cochrane Centre, Oxford

Professor Paul Dieppe
Director, MRC Health Services Research Centre
University of Bristol

Professor Michael Drummond
Director, Centre for Health Economics
University of York

Dr Vikki Entwistle
Senior Research Fellow, Health Services Research Unit
University of Aberdeen

Professor Ewan B Ferlie
Professor of Public Services Management
Imperial College, London

Professor Ray Fitzpatrick
Professor of Public Health & Primary Care
University of Oxford

Dr Naomi Fulop
Deputy Director, Service Delivery & Organisation Programme
London School of Hygiene & Tropical Medicine

Mrs Jenny Griffin
Head, Policy Research Programme
Department of Health
London

Professor Jeremy Grimshaw
Programme Director
Health Services Research Unit
University of Aberdeen

Professor Stephen Harrison
Professor of Social Policy
University of Manchester

Mr John Henderson
Economic Advisor
Department of Health, London

Professor Theresa Marteau
Director, Psychology & Genetics Research Group
Guy's, King's & St Thomas's School of Medicine, London

Dr Henry McQuay
Clinical Reader in Pain Relief
University of Oxford

Dr Nick Payne
Consultant Senior Lecturer in Public Health Medicine
ScHARR
University of Sheffield

Professor Joy Townsend
Director, Centre for Research in Primary & Community Care
University of Hertfordshire

Professor Kent Woods
Director, NHS HTA Programme, & Professor of Therapeutics
University of Leicester



*continued*

# HTA Commissioning Board

**Members**

**Programme Director**
**Professor Kent Woods**
Director, NHS HTA Programme, &
Professor of Therapeutics
University of Leicester

**Chair**
**Professor Shah Ebrahim**
Professor of Epidemiology of Ageing
University of Bristol

**Deputy Chair**
**Professor Jon Nicholl**
Director, Medical Care Research Unit
University of Sheffield

Professor Douglas Altman
Director, ICRF Medical Statistics Group
University of Oxford

Professor John Bond
Director, Centre for Health Services Research
University of Newcastle-upon-Tyne

Ms Christine Clark
Freelance Medical Writer
Bury, Lancs

Professor Martin Eccles
Professor of Clinical Effectiveness
University of Newcastle-upon-Tyne

Dr Andrew Farmer
General Practitioner &
NHS R&D Clinical Scientist
Institute of Health Sciences
University of Oxford

Professor Adrian Grant
Director, Health Services Research Unit
University of Aberdeen

Dr Alastair Gray
Director, Health Economics Research Centre
Institute of Health Sciences
University of Oxford

Professor Mark Haggard
Director, MRC Institute of Hearing Research
University of Nottingham

Professor Jenny Hewison
Senior Lecturer
School of Psychology
University of Leeds

Professor Alison Kitson
Director, Royal College of Nursing Institute, London

Dr Donna Lamping
Head, Health Services Research Unit
London School of Hygiene & Tropical Medicine

Professor David Neal
Professor of Surgery
University of Newcastle-upon-Tyne

Professor Gillian Parker
Nuffield Professor of Community Care
University of Leicester

Dr Tim Peters
Reader in Medical Statistics
University of Bristol

Professor Martin Severs
Professor in Elderly Health Care
University of Portsmouth

Dr Sarah Stewart-Brown
Director, Health Services Research Unit
University of Oxford

Professor Ala Szczepura
Director, Centre for Health Services Studies
University of Warwick

Dr Gillian Vivian
Consultant in Nuclear Medicine & Radiology
Royal Cornwall Hospitals Trust
Truro

Professor Graham Watt
Department of General Practice
University of Glasgow

Dr Jeremy Wyatt
Senior Fellow
Health Knowledge Management Centre
University College London



Current and past membership details of all HTA 'committees' are available from the HTA website (see inside front cover for details)

## Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

***We look forward to hearing from you.***

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278