Developing valid guidelines: methodological and procedural issues from the North of England Evidence Based Guideline Development Project

Martin Eccles, Zoe Clapp, Jeremy Grimshaw, Philip C Adams, Bernard Higgins, Ian Purves, Ian Russell

There is increasing interest in clinical guidelines within the United Kingdom. With this interest has come an increasing awareness of the methodological issues involved in guideline development, although reports are limited.1 Woof described three main methods of guideline development: informal consensus, formal consensus, and evidence linked.2 In guidelines developed through informal consensus, the method used most often in the United Kingdom, the guideline panel has poorly defined, often implicit, criteria for decision making. Development by formal consensus, used by many consensus development conferences and Delphi groups, provides “greater structure to the analytical process” but still fails to provide “an explicit linkage between recommendations and quality of evidence.” Guideline development linked to evidence requires the explicit link between recommendations and the quality of the supporting evidence.3 4 This allows the user to make an informed choice about whether or not to comply with recommendations. Guidelines are valid if “when followed they lead to the health gains and costs predicted for them.”5 6 We have previously argued that linking guidelines to evidence is one of three developmental prerequisites necessary to maximise guideline validity,6 the other two being systematic review of the evidence and guideline development occurring within an appropriately multidisciplinary group. These methods are unlikely to be used at a local (as opposed to a regional or national) level as necessary skills are unlikely to be available. Local groups should concentrate on identifying valid regional or national guidelines which they can then adapt to their local needs.7 Unfortunately such valid guidelines are few and descriptions of their development in the United Kingdom are non-existent. This article describes the development of explicit, evidence based guidelines for the primary care management of asthma in adults and for stable angina.8 12 Two chronic conditions, predominantly managed in primary care and important because of their associated morbidity and mortality.13 We present our experiences and the methodological issues considered during guideline development, both for those who are involved in this process and for those who wish to know more about the practicalities.

Setting up the guideline development project

To develop guidelines successfully it was necessary to convene four groups: a project team; a project management group; and two guideline development groups. The project team was composed of the principle investigator, a junior research associate employed to work full time on the project, a cardiologist, and a chest physician. The role of the project team was to undertake the day to day running of the project; this involved the identification and interpretation of relevant evidence, the convening and running of the guideline development groups, and the production of the resulting guidelines. The project management group was composed of specialists in guideline methodology and a health information expert; their role was to provide technical advice on a formal or informal basis throughout the project. The guideline development groups were to produce guideline recommendations in the light of the evidence.

Guideline development groups

MEMBERSHIP AND ROLES WITHIN THE GROUPS

The composition of a guideline development group depends on the roles required within the group and the disciplines or backgrounds of the group members who were stakeholders in the processes covered by the guideline. All the parties whose activities would be covered by the guideline or who had other legitimate reasons for having an input into the process have to be identified. This is important to ensure adequate discussion of the evidence (or its absence) when developing the recommendations in the guideline.14 Given that both guidelines were for use in primary care the stakeholders involved (and their eventual numbers in each group) were: general practitioners (five), practice nurses (two), representatives of patients (one), secondary care physicians (one), public health physicians (one), and health economists (one).

From a theoretical basis we identified the roles required within the guideline development groups as those of group member, group leader, specialist clinician, technical assistant, and administrative assistant. All recruited people needed to have two specific characteristics: interest in the project and a positive attitude towards guidelines.
**Developing valid evidence based guidelines**

**Group members**
The role of individual group members from the different specialties were invited to develop recommendations for practice in the light of the evidence or, if there were none, in its absence. They were not, however, expected to represent the whole range of views held within their profession.

**Group leader**
The role of the group leader was both to ensure that the group functioned effectively (the group process) and that it achieved its aims (the group task). Because of the complexity of the task and the anticipated size of the group, this role was divided between two people: one pursued the group task (the development of the guidelines) and the other guarded the group process. The group task was led in both groups by a health services researcher experienced in guideline development (ME). The two group process leaders were general practitioners from outside the project team. They each had considerable experience in leadership of small groups thus fulfilling the role of a group process leader “whose disinterested position is unquestioned by any of the concerned parties but whose expertise in coordinating groups of health professionals is accepted by all.”

**Specialist clinician**
The role of the specialist clinician was to work with the project team to develop the literature review, read and summarise relevant papers, and to lead discussion of the evidence within the group meetings. These roles were fulfilled by the cardiologist and chest physician from the project team. As the skills provided by the small group leader and the specialist were central to the functioning of the group they were paid a fee.

**Technical and administrative assistance**
Technical assistance was required to support the specialist on the literature review and to present this to the group in a form that allowed them to make recommendations. Assistance was also needed to arrange an external review of the draft guidelines, collate reviewers’ comments, feed these back to the panel, and produce the final draft of the guideline. Administrative assistance was required for such tasks as preparing papers for meetings, taking notes, and arranging venues. Both technical and administrative roles were fulfilled by the project team.

**RECRUITMENT**

**Specialist clinicians**
The specialist clinicians were recruited by personal invitation. One was previously known to the principal investigator and the other was identified after an approach to the Department of Respiratory Medicine of a regional teaching hospital. After discussion of the project the invitations were accepted.

**Group leaders**
The group leaders were identified through the Northern Region Postgraduate Institute for Medicine and Dentistry. Both were experienced general practitioners and course organisers on the local vocational training in general practice scheme. As with the specialists their recruitment was by personal invitation and was accepted after discussion of the project.

**Group members**
Group members were identified in several ways. The general practitioners and consultants were identified through personal knowledge or contacts of the project team. Efforts were made to enrol members from outside the former Northern Region although this was only successful in the asthma group, to which we recruited a primary care asthma specialist who had previously sat on a panel for development of a national consensus guideline for asthma. Although twice as many general practitioners were approached as were able to participate, it was relatively easy to identify and recruit general practitioners and indeed all other group members with the exception of practice nurses and patients. Although there were, in several localities, groups of practice nurses and local patients such as cardiac support groups, there was not a regional forum for either of these groups. The practice nurses were eventually identified through personal contacts and the patients through the Community Health Council and a local patient support group.

Potential group members were invited by letter to express initial interest in joining the group. This was followed by a telephone call during which the purpose of the guideline development was explained along with the commitment required. Only those that were interested responded to this invitation; group members thus selected themselves by their interest in the project and positive attitude to guidelines. As meetings were held on a weekday afternoon group members had to have available time to commit themselves to the work. To help overcome the difficulty of taking time off from daily work, group members were offered reimbursement of any expenditure incurred in attending the meetings, such as travel expenses and general practitioner locum costs.

**Scope of the guidelines**
An important decision for the project team, subsequently endorsed by the guideline

---

**Information about the United Kingdom National Health Service (NHS)**

- Northern Region – NHS administrative area in the north of England comprising the counties of Cumbria, Northumberland, Tyne and Wear, Durham, and Cleveland
- Northern Region Postgraduate Institute for Medicine and Dentistry – University department responsible for postgraduate education of doctors and dentists
- Medical Audit Advisory Groups – groups responsible for audit in primary care
development groups, was to define the scope of the guidelines. Hadorn and Baker suggested that, as cost and time are important factors in deciding a guideline's scope, there is a trade off between depth and breadth. The narrower the scope of a guideline the more deeply a group can delve into a given area but this is at the cost of breadth which may leave important questions unanswered. For example, the angina group discussed whether they should restrict themselves to stable angina or include unstable angina and acute myocardial infarction and also whether they should consider the scientific literature on risk factor management. Although they acknowledged that it was important to consider the management of hypertension, hypercholesterolaemia, and obesity it was explicitly agreed that systematic reviews of these areas were beyond the scope of the group; they also chose to restrict themselves to stable angina.

Identification and application of evidence
REVIEW OF EVIDENCE
We have previously argued that the best method of identifying and acting on evidence is by systematic review. This is an efficient scientific technique to identify and summarise evidence on the effectiveness of interventions and to allow the generalisability and consistency of research findings to be assessed and data inconsistencies to be explored. It is achieved by applying scientific principles to reduce random and systematic errors through explicit search strategies to identify studies; explicit inclusion criteria for selecting studies; for review; explicit criteria to assess quality; and rigorous methods of data abstraction and manipulation. Systematic reviews also allow the evidence in different areas of a guideline to be graded on the basis of the research design and its susceptibility to bias.

The identification and assessment of the literature was undertaken by the project team. Initial criteria for papers that were to be considered in the review process were that they had to be published in peer reviewed journals; they should ideally provide evidence of effectiveness from evaluations conducted under typical routine practice (failing this we would accept evidence of efficacy from evaluations conducted under highly controlled circumstances not typical of routine practice); measures of outcome should ideally be patient based but where these were absent other intermediate outcome measures would be used; and if primary cost effectiveness studies were identified then they would be used. Finally, studies using experimental and quasi-experimental designs (randomised controlled trials and both prospective and retrospective non-randomised controlled studies) would be preferred.

The search strategy
Search strategies were defined by the project team and applied through MEDLINE for the 10 years 1985–94. Although this was a practical decision influenced by the volume of papers and the time and resources available it was also thought that any new evidence that had not yet influenced practice was most likely to have been published during the previous 10 years.

As the subjects of the guidelines were broad clinical conditions, a single wide ranging search for each condition was used, rather than a series of focused searches aimed at specific issues. However, in the light of the results of the two main searches a few specific searches were conducted. All searches were limited to studies of human adults written in English. For both areas we conducted medical subject headings (MeSH) and free text searches on the terms meta-analysis, randomised controlled trial, reviews, cohort studies, and case-control studies. For asthma we also sought the terms asthma, peak expiratory flow rate, lung diseases obstructive, forced expiratory volume, and paroxysmal dyspnoea; and for stable angina the terms coronary disease, angina pectoris. Additional specific MEDLINE searches were carried out on the following terms: decision making, theophylline, terbutaline, antihistamine, isosorbide, myocardial infarction, secondary prevention, and buccal. The BIDS electronic database21 was also searched (with the search terms: asthma plus management and angina plus management). References were also identified from two other sources. Firstly, if there was no recent evidence in a clinically important area the specialist clinician was asked to identify, from his personal knowledge, key articles published before 1985. Secondly, the reference lists of non-systematic reviews were checked. We did not attempt to access the "grey literature"; nor did we identify letters in response to original articles.

Initial sifting
The identified papers were first sifted on the basis of title only by the principle investigator, a medically qualified health services researcher; they were then sifted on the basis of title and abstract by the specialist clinician. Both were working to agreed criteria of clinical relevance; as the initial sift was on title alone if there was any doubt about an article’s relevance it was retained for fuller assessment. However, articles about topics irrelevant to the guidelines, such as anaesthetics, were readily identified and removed.

Methodological assessment
The collected papers were then assessed against a set of methodological criteria several of which were pass or fail. As an example one such criterion was that any randomised controlled trial with a group size of less than 20 would be excluded. This criterion was used because we thought that it was reasonable to consider only studies of a minimum size given that we did not propose to undertake a formal meta-analysis. When Kleijnen et al looked at size criteria in a review of clinical trials of homoeopathy they argued that "in trials with limited numbers of participants one cannot be confident that randomisation will equally
Developing valid evidence-based guidelines

divide known and unknown confounders over the experimental and control groups. As well, publication bias may be less likely for experiments with large numbers of participants.\(^{22}\) As a result of this they did not score trials with group sizes less than 25. In a similar review of physiotherapy the same research group did not score trials with group sizes of 50 or less.\(^{23}\) Also, the analysis of such studies would have required statistical correction (Yates's correction or the use of small sample size tests such as Fisher's exact test) and would need to be treated with caution if generalisations were to be made from them.

Table 1 shows the number of papers remaining after the various stages of the sifting process.

**Summarising evidence**

Papers which passed the methodological sift were then read and summarised by the specialist; several discrete areas were reviewed by other invited group members. They were supported in this by members of the project team. The reviewer, with members of the project team, abstracted the evidence from the papers. The evidence was graded according to the study design with a grading system adapted from the Canadian task force classification (table 2).\(^{24}\) The evidence was summarised with qualitative methods. Quantitative (meta-analysis) techniques were not used as we were dealing with studies other than randomised controlled trials.

**Method of constructing guidelines**

**GROUP MEETINGS**

The groups chose to conduct their work over a series of half day meetings, rejecting the option of committing larger blocks of time on fewer occasions. Both groups met on nine occasions. The first half of the first meeting was devoted solely to introductions and discussion of the task. This second topic recurred on several occasions, most often in the early meetings, although aspects of the definition of the group task occurred all the way through. Having agreed on the task the groups were presented with evidence from which they were to make recommendations.

At the early meetings the groups discussed the papers and their interpretation which gave them a clear idea of the review process. As the reviewing process developed they were presented with draft recommendations linked to statements of the evidence and to blocks of summarised papers. The recommendations and the statements were then discussed and suggested changes were edited into the developing guideline by the project team between meetings. In both groups the decision making process was informal consensus; no formal strategies were used.

**DERIVING RECOMMENDATIONS**

Recommendations were graded A to C (table 3). Both groups distinguished between the level of evidence and the strength of the associated recommendation. It was possible to have methodologically sound (level I) evidence about an area of practice that was clinically irrelevant or had such a small effect that it was of little clinical significance and would therefore attract a lower strength of recommendation. More often, a statement of evidence would only cover one part of an area in which a recommendation had to be made or would cover it in a way that was conflicting. Therefore, to produce a comprehensive recommendation the group had to extrapolate from the available evidence. This could lead to a strength C recommendation based on evidence of level I. For example, the strength C recommendation "short acting \(\beta_2\) agonists should be used on an 'as required' basis to relieve symptoms" is supported by the evidence level I statement "there is conflicting evidence on the issue of p.r.n. versus regular dosage (of short acting \(\beta_2\) agonists)." The two methodologically acceptable studies supporting the statement produced contradictory findings due to a combination of patient selection, drug selection, and study design.

**AREAS WITHOUT EVIDENCE**

Both groups had to decide what to do in areas in which recommendations had to be made but evidence was lacking. Limiting recommendations to areas where evidence existed would reduce the scope of the guidelines and limit their value to clinicians and policy makers who need to make decisions in the presence of imperfect knowledge.\(^{25}\) The solution was to produce guidelines that contain a mixture of evidence linked and consensus based recommendations.\(^{1}\) The explicit grading of evidence and recommendations meant that "areas of clinical uncertainty

---

**Table 1** Number of papers remaining after the various stages of the sifting process

<table>
<thead>
<tr>
<th>Stage in the sifting process</th>
<th>Angina</th>
<th>Asthma</th>
</tr>
</thead>
<tbody>
<tr>
<td>All references</td>
<td>38675</td>
<td>27401</td>
</tr>
<tr>
<td>Restricted to human adult studies in the English language</td>
<td>24194</td>
<td>18189</td>
</tr>
<tr>
<td>Restricted by study type</td>
<td>9165</td>
<td>4936</td>
</tr>
<tr>
<td>Sift on title alone</td>
<td>1810</td>
<td>1034</td>
</tr>
<tr>
<td>Sifted on abstract and title</td>
<td>699</td>
<td>513</td>
</tr>
<tr>
<td>Methodological sift</td>
<td>286</td>
<td>150</td>
</tr>
</tbody>
</table>

**Table 2** Level of evidence

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Well designed randomised controlled trials, meta-analyses, or systematic reviews</td>
</tr>
<tr>
<td>II</td>
<td>Well designed non-randomised prospective or retrospective controlled studies</td>
</tr>
<tr>
<td>III</td>
<td>Uncontrolled studies or consensus</td>
</tr>
</tbody>
</table>

**Table 3** Strength of recommendation

<table>
<thead>
<tr>
<th>Strength of recommendation</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Directly based on level I evidence</td>
</tr>
<tr>
<td>B</td>
<td>Directly based on level II evidence or extrapolated recommendation from level I evidence</td>
</tr>
<tr>
<td>C</td>
<td>Directly based on level III evidence or extrapolated recommendation from level I or II evidence</td>
</tr>
</tbody>
</table>
are demarcated . . . and a distinction made between evidence of ineffectiveness . . . and the absence of evidence for effectiveness.”

The two groups chose different strategies to achieve this. The asthma group decided that in the absence of evidence their first recourse would be to a current national consensus guideline.26 If these recommendations were appropriate they would be used. If they had been superseded, or did not provide a relevant recommendation, then the group would reach its own consensus view. In the absence of a widely recognised national guideline the angina group decided that they would make recommendations based on their experience and reach a consensus view on what the recommendation should say. Both groups agreed that if it were not possible to reach a consensus then the relevant area would be explicitly identified as one where the group could not agree. The areas without evidence generated more discussion and disagreement within the group than any of the areas in which there was evidence.

Involvement of patients in guideline development
Although we invited two patients to be group members and they attended meetings regularly they were often non-participating observers of technical discussion to which they could offer no input. Although the input of patients into the process of guideline development is important, this may be better gathered outside the guideline development group, with techniques such as focus groups, rather than having a single patient as a frequently non-participating group member. This is an area where further methodological research is clearly required.

External review of the guideline
We identified the need for three groups of reviewers: guideline experts to review the methods; content area experts to review the interpretation, logic, and clarity; and potential users to appraise the clinical applicability and usefulness of the guidelines. The purpose of the review was to identify errors in the areas that the reviewers were asked to cover; it was not to undertake a further round of consensus interpretation of the contents of the guideline. In choosing the number of reviewers there was a trade off between the breadth of review sought and the subsequent task of dealing with detailed comments; we used up to nine reviewers on each guideline.

Reviewers’ comments were amalgamated and then ordered by page of the guidelines so that they could be easily identified by the project team and the guideline development groups to whom they were sent. All comments and responses were considered and the errors identified were amended in the final draft. An example of an issue raised by reviewers was the concern of the specialists who reviewed the angina guideline about a recommendation suggesting that general practitioners use verapamil with β blockers (the systematic review had produced several randomised controlled trials that had shown the combination to be effective). The reviewers’ comments combined with the cautions within the British National Formulary led to the recommendation being withdrawn. The guidelines were sent to reviewers as both an advanced draft of the complete guideline and as a shorter, four page, “quick reference guide for clinicians.” The potential users of the guidelines (all general practitioners) expressed concern about the usefulness of the quick reference guide on the basis that it was too short to have the legitimacy of a full evidence based document but was too long to be readily usable. In the light of these comments the quick reference guide was dropped and the reviewer’s suggested single sheet summary produced instead.

Timetabling
In the original proposal to the funding body the development of the guidelines was scheduled to take 12 months; in retrospect this was not long enough. The first three months were taken up with the initial organisation of the systematic review, the identification of potential group members, and the necessary time leading up to the first meetings. By the end of 12 months the nine monthly meetings of each group were complete and the draft versions of the guidelines were ready to go out to reviewers. Although reviewers were identified in advance the review process and the subsequent editing and production of the guidelines took a further six months. There are certain fixed periods in the process that could not be shortened. For example, once the decision had been taken to develop the guidelines over nine half day meetings, the process inevitably took nine months because clinicians found it impossible to meet more often. Also, it would have been difficult to have completed the evidence review process in time for more frequent meetings. Therefore, with three months required to set up the project, the phase that was not accurately predicted was that of external review and the subsequent guideline production. Given the other pressures on the reviewers we were very fortunate that they were able to comment within a four week period. The collation of their comments and the responses to them required a further four weeks and these then had to be sent out to the groups for their responses with a further four week turnaround. After further collation over the subsequent two weeks the guidelines were only then ready for the final proof reading and the production phase.

Discussion
We have presented the experiences from, and issues raised by, developing two evidence based guidelines. Guideline development required a range of skills. Identifying and assimilating the evidence required the skills of systematic reviewing to identify relevant papers and then the skills of an experienced and critical clinician to summarise their clinical relevance. The evidence was fed to large
Developing valid evidence based guidelines

multidisciplinary groups that needed skilled leadership, to ensure that the groups both worked effectively and achieved their tasks. The product of the groups' deliberations had to be turned into an evidence based guideline. This range of skills is central to the successful development of any guideline that requires a new review of evidence.

These skills will not be available at a local level; local groups will need valid development of guidelines to occur at a regional or national level and the availability of such guidelines should increase the validity of local development of guidelines. The Royal College of General Practitioners suggested that the first step for local guideline developers is to identify a valid national (or regional) guideline. Therefore, at a national (or regional) level, expertise will be required in conducting systematic reviews, assessing the evidence, and developing valid guidelines. At a local level, expertise will be required in appraising national guidelines (although a national appraisal scheme would reduce the need for this) and then adapting them to meet local resource constraints and barriers to implementation, tasks not considered by this project.

The development of these guidelines required considerable resources. Even having planned the process and assembled a group with appropriate skills we still underestimated the complexity of the whole process. The guidelines were initially costed at £27 000 each and scheduled for development over 12 months; they took 18 months to develop and we would estimate a more realistic cost to be £35 000. It could be argued that the marginal costs of developing valid guidelines are greater than the marginal benefits, relative to guideline development by informal consensus within an expert group. However, small changes in the process of care – for instance, substitution of cheaper treatments – or its outcome – for instance, avoided early deaths – will rapidly recoup the guideline development costs. For example, the costs of developing the stable angina guideline would be recovered by the use of a generic β blocker instead of a commonly used long acting proprietary mononitrare in only 350 patients for one year, a strategy recommended within the guideline and linked to improved outcome.

The issues that we have presented are central to the development of valid guidelines but practical experience of dealing with these issues is very limited in the United Kingdom. We present our experiences as a first step in understanding how these issues affect the development of evidence based guidelines within the National Health Service. Methodological expertise needs to be developed at both national and local levels and resources need to be committed to the process. It is only by investing in such processes that the validity of guidelines can be optimised.

We are grateful to the guideline development group members and to the external reviewers. We thank Liz Wood for providing secretarial support. The North of England evidence based guideline development project was funded by the former Northern Regional Research and Development Division (with contributions from Durham, Northumberland, and South Tyneside Medical Audit Advisory Groups). The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Office Home and Health Department. The views expressed are those of the authors and not the funding bodies.

21 Bath Information and Data Services. Institute for Scientific Information. Bath, University of Bath.