

6 Forming guideline recommendations

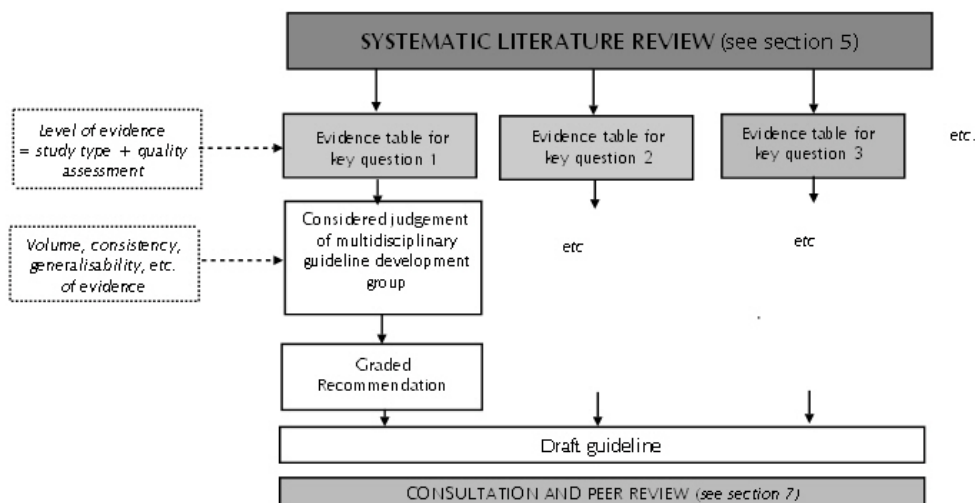
6.1 SYNTHESISING THE EVIDENCE

Guideline recommendations are graded to differentiate between those based on strong evidence and those based on weak evidence. This judgement is made on the basis of an (objective) assessment of the design and quality of each study (*as discussed in section 5*) and a (perhaps more subjective) judgement on the consistency, clinical relevance and external validity of the whole body of evidence.¹ The aim is to produce a recommendation that is evidence-based, but which is relevant to the way in which health care is delivered in Scotland and is therefore *implementable*.

It is important to emphasise that the grading does not relate to the *importance* of the recommendation, but to the strength of the supporting evidence and, in particular, to the predictive power of the study designs from which that data was obtained. Thus, the grading assigned to a recommendation indicates to users the likelihood that, if that recommendation is implemented, the predicted outcome will be achieved.

The process for synthesising the evidence base to form graded guideline recommendations is illustrated in figure 6.1.

Figure 6.1
FORMATION OF GUIDELINE RECOMMENDATIONS



Evidence tables are compiled by SIGN Executive staff based on the quality assessments of individual studies provided by guideline development group members. The tables summarise all the validated studies identified from the systematic literature review relating to each key question. They are presented in a standard format to make it easier to compare results across studies, and will present separately the evidence for each outcome measure used in the published studies. These evidence tables form an essential part of the guideline development record and ensure that the basis of the guideline development group's recommendations is transparent. An example evidence table is shown in Annex E.

6.2 CONSIDERED JUDGEMENT

It is rare for the evidence to show clearly and unambiguously what course of action should be recommended for any given question. Consequently, it is not always clear to those who

were not involved in the decision making process how guideline developers were able to arrive at their recommendations, given the evidence they had to base them on. In order to address this problem, SIGN has introduced the concept of **considered judgement**.

Under the heading of considered judgement, guideline development groups summarise their view of the total body of evidence covered by each evidence table. This summary view is expected to cover the following aspects:

- Quantity, quality, and consistency of evidence
- Generalisability of study findings
- Directness of application to the target population for the guideline.
- Clinical impact (i.e. the extent of the impact on the target patient population, and the resources needed to treat them.)
- Implementability (i.e. how practical it would be for the NHS in Scotland to implement the recommendation.)

Guideline development groups are provided with a pro forma in which to record the main points from their considered judgement. An example of this form and the associated notes for users is included in Annex D. Once they have considered these issues, the group are asked to summarise their view of the evidence and assign a **level of evidence** to it, before going on to derive a graded recommendation.

6.3 LEVELS OF EVIDENCE AND GRADES OF RECOMMENDATION

SIGN formerly used the levels of evidence developed by the US Agency for Health Care Policy and Research (AHCPR, now the US Agency for Health Research and Quality, AHRQ).² As a number of limitations were becoming apparent in that system, a review was carried out and new levels of evidence and associated grades of recommendation were developed. Following extensive consultation and international peer review, the new grading system was introduced in Autumn 2000.^{3, 4} The current grading system is shown in Figure 6.2.

The assignment of a level of evidence should involve all those on a particular guideline development group or subgroup involved with reviewing the evidence in relation to each specific question. The allocation of the associated grade of recommendation should involve participation of all members of the guideline development group. Where the guideline development group are unable to agree a unanimous recommendation, the difference of opinion should be formally recorded and the reasons for dissent noted.

The revised grading system is intended to place greater weight on the *quality* of the evidence supporting each recommendation, and to emphasise that the *body of evidence* should be considered as a whole, and not rely on a single study to support each recommendation. It is also intended to allow more weight to be given to recommendations supported by good quality observational studies where RCTs are not available for practical or ethical reasons. Through the considered judgement process guideline developers are also able to downgrade a recommendation where they think the evidence is not generalisable, not directly applicable to the target population, or for other reasons is perceived as being weaker than a simple evaluation of the methodology would suggest.

Experience has shown that, as expected, the new grading system produces more Grade B recommendations, and fewer Grade A. It has also become clear that establishing the link between evidence and recommendation is a neglected aspect of the development process and there may be a need for some kind of formal consensus method to be introduced.

Section 6: Forming guideline recommendations

Figure 6.2

SIGN GRADING SYSTEM

LEVELS OF EVIDENCE	
1 ⁺⁺	High quality meta analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1 ⁺	Well conducted meta analyses, systematic reviews of RCTs, or RCTs with a low risk of bias
1 ⁻	Meta analyses, systematic reviews of RCTs, or RCTs with a high risk of bias
2 ⁺⁺	High quality systematic reviews of case-control or cohort or studies High quality case-control or cohort studies with a very low risk of confounding, bias, or chance and a high probability that the relationship is causal
2 ⁺	Well conducted case control or cohort studies with a low risk of confounding, bias, or chance and a moderate probability that the relationship is causal
2 ⁻	Case control or cohort studies with a high risk of confounding, bias, or chance and a significant risk that the relationship is not causal
3	Non-analytic studies, e.g. case reports, case series
4	Expert opinion

GRADES OF RECOMMENDATION

- A** At least one meta analysis, systematic review, or RCT rated as 1⁺⁺, and directly applicable to the target population; *or*
A systematic review of RCTs or a body of evidence consisting principally of studies rated as 1⁺, directly applicable to the target population, and demonstrating overall consistency of results
- B** A body of evidence including studies rated as 2⁺⁺, directly applicable to the target population, and demonstrating overall consistency of results; *or*
Extrapolated evidence from studies rated as 1⁺⁺ or 1⁺
- C** A body of evidence including studies rated as 2⁺, directly applicable to the target population and demonstrating overall consistency of results; *or*
Extrapolated evidence from studies rated as 2⁺⁺
- D** Evidence level 3 or 4; *or*
Extrapolated evidence from studies rated as 2⁺

GOOD PRACTICE POINTS

- Recommended best practice based on the clinical experience of the guideline development group*

* see overleaf

On occasion, guideline development groups find that there is an important practical point that they wish to emphasise but for which there is not, nor is there likely to be, any research evidence. This will typically be where some aspect of treatment is regarded as such sound clinical practice that nobody is likely to question it. These are marked in the guideline as Good Practice Points, and are indicated . It must be emphasised that these are not an alternative to evidence-based recommendations, and should only be used where there is no alternative means of highlighting the issue.

6.4 RESOURCE IMPLICATIONS

SIGN recognises that, in an NHS with limited resources and ever-increasing costs, the ability to cost individual items of care and weigh these against some quantification of patient benefit is important.⁵ However, the science of economic analysis of health care is at a relatively early stage and many published studies do not meet the required methodological standard to be incorporated as part of the evidence base for a guideline. A number of approaches to the incorporation of resource issues into clinical guidelines are under development^{6,7} but at this stage none are regarded as sufficiently well proven or appropriate for use in the SIGN methodology. Where there is published economic evidence this has to be identified and evaluated in a consistent manner. Health economic databases have been added to the coverage of literature searches (see section 5) and studies identified from those searches should be evaluated using SIGN Checklist 6, which is based on the criteria for evaluation of economic studies published in the British Medical Journal.⁸ This evidence can be considered alongside clinical evidence at the considered judgement stage (see section 6.2).

In other areas, the appropriate action may be inclusion in the guideline of a commentary on the main economic issues that should be considered in relation to the subject of the guideline. Examples of this approach can be found in the SIGN guidelines on hip fracture or cardiac rehabilitation. A final option is the provision of basic information that will allow guideline users to work out the cost implications for their own service, and examples of this approach can be seen in the guidelines on osteoporosis¹ or epithelial ovarian cancer⁹.

6.5 CURRENT AREAS FOR DEVELOPMENT

The SIGN Methodology Development Group was established to consider new developments in guideline methodology, and to attempt to answer specific questions on methodological issues. They are currently looking at the following questions:

Key recommendations: The grading system identifies those recommendations with the strongest evidence base. These are *not necessarily* the recommendations with the greatest clinical impact. This could be addressed by asking the guideline development group to use their collective expertise to identify the recommendations that would have the greatest clinical impact, and highlight these in the guideline. How can this be done while guarding against bias in the development process?

Qualitative studies as evidence: Qualitative methods are increasingly being used to inform practice in some aspects of medical care. At present, there is no mechanism for incorporating such studies in the evidence base. Some progress has been made on methods of identifying qualitative studies, and in evaluating their methodological quality. The use of qualitative evidence to identify issues of concern to patients, and to help identify key questions to be addressed in the guideline is becoming an established part of SIGN methodology. Incorporation of qualitative evidence alongside quantitative evidence in supporting guideline recommendations remains an issue for further investigation.

Section 6: Forming guideline recommendations

Revision of the grading system: The grading system described in Section 6.3 is an improvement on the previous system, but still has weaknesses that need to be addressed. SIGN has been participating in the international GRADE project aimed at developing a methodologically sound system that can be applied across countries and cultures. There is no intention at present to abandon the existing system, but rather to use lessons from GRADE and elsewhere to bring about incremental changes that will help the grading system evolve to meet the increasingly complex needs of guideline developers.

References to section 6

1. Guyatt G, Sackett D, Sinclair J, Hayward R, Cook D, Cook R. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. JAMA 1995;274:1800-4.
 2. Clinical Practice Guideline No.1: acute pain management: operative or medical procedures and trauma. Rockville (MD): US Department of Health and Human Services. Agency for Health Care Policy and Research; 1993. AHCPR Publication No. 92-0023. [cited 1993].
 3. Scottish Intercollegiate Guidelines Network. Methodology Review Group. Report on the review of the method of grading guideline recommendations. Edinburgh: SIGN; 1999.
 4. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ 1991;323:334-6.
 5. Drummond M, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. 2nd ed. Oxford: Oxford University Press; 1997.
 6. Mason J, Eccles M, Freemantle N, Drummond M. NICEly does it: economic analysis within evidence -based clinical practice guidelines. (CHE Discussion Paper No.164). Available from <http://www.york.ac.uk/inst/che/DP164.pdf>: [Accessed. 6 February. 2004.]
 7. Eccles M, Mason J. How to develop cost-conscious guidelines. Health Technology Assessment 2001;5(16).
 8. Drummond M, Jefferson T. Guidelines for authors & peer reviewers of economic submissions to the BMJ. BMJ 1996;313:275-83.
-