

Developing clinical guidelines

The methods of guideline development should ensure that treating patients according to guidelines will achieve the outcomes that are desired. This article presents a combination of the literature about guideline development and the results of our combined experience in guideline development in North America and Britain. It considers the 5 steps in the initial development of an evidence-based guideline.



Paul G Shekelle
West Los Angeles Veterans Affairs Medical Center (111G)
11301 Wilshire Blvd
Los Angeles, CA 90073

Steven H Woolf
Health Services Research and Development Service
Department of Family Practice
Virginia Commonwealth University
Fairfax, VA 22033

Martin Eccles
Center For Health Services Research
University of Newcastle upon Tyne
Newcastle upon Tyne NE2 4AA
United Kingdom

Jeremy Grimshaw
Health Services Research Unit
University of Aberdeen
Aberdeen AB9 2ZD
United Kingdom

Correspondence to:
Dr Shekelle
shekelle@rand.org

Identifying and refining the subject area of a guideline

Prioritizing topics

Guidelines can be developed for a wide range of subjects. Clinical areas can be concerned with conditions (abnormal uterine bleeding, coronary artery disease) or procedures (hysterectomy, coronary artery bypass surgery). Given the large number of potential areas, some priority setting is needed to select an area for guideline development. Potential areas can emerge from an assessment of the major causes of morbidity and mortality for a given population, uncertainty about the appropriateness of health care processes or evidence that they are effective in improving patient outcomes, or the need to conserve resources in providing care.

Refining the subject area

The topic for guideline development will usually need to be refined before the evidence can be assessed in order to answer exact questions. The usual way of refining the topic is through dialogue among clinicians, patients, and the potential users or evaluators of the guideline.

Discussions about the scope of the guideline will also take place within the guideline development panel.

If the topic is not refined, the clinical condition or question may be too broad in scope. For example, a guideline on the management of diabetes could cover primary, secondary, and tertiary care elements of management and also multiple aspects of management, such as screening, diagnosis, dietary management, drug treatment, risk factor management, or indications for referral to a specialist. Though all of these legitimately could be dealt with in a guideline, the task of developing such a guideline would be considerable; therefore, a group to develop guidelines needs to be clear which areas are and are not within the scope of their activities. It is possible to develop guidelines that are both broad in scope and evidence-based; but to do so usually requires considerable time and money, both of which are frequently underestimated by inexperienced developers of evidence-based clinical practice guidelines.

One method of defining the clinical question of interest and also identifying the processes for which evidence needs to be collected and assessed is the construction of models or causal pathways.¹ A causal pathway is a diagram that illustrates the links between interventions of interest and the intermediate, surrogate, and health outcomes that the interventions are thought to influence. In designing the pathway, guideline developers make explicit the premises on which their assumptions of effectiveness are based and the outcomes (benefits and harms) that they consider important. This identifies the specific questions that must be answered by the evidence to justify conclusions of effectiveness and highlights gaps in the evidence for which future research is needed.

Summary Points

- Identifying and refining the subject area is the first step in developing a guideline
- Convening and running guideline development groups is the next step
- On the basis of systematic reviews, the group assesses the evidence about the clinical question or condition
- This evidence is then translated into a recommendation within a clinical practice guideline
- The last step in guideline development is external review of the guideline

Running guideline development groups

Setting up a guideline development project

To successfully develop a guideline, it may be necessary to convene more than 1 group. A project or management team could undertake the day-to-day running of the work, such as the identification, synthesis, and interpretation of relevant evidence; the convening and running of the guideline development groups; and the production of the resulting guidelines. An additional guideline development group would produce recommendations in light of the evidence or of its absence.

Group membership and roles

Group members—Identifying stakeholders involves identifying all of the groups whose activities would be covered by the guideline or who have other legitimate reasons for having input in the process. This is important to ensure adequate discussion of the evidence (or its absence) when developing the recommendations in the guideline. When presented with the same evidence, a single specialty group will reach different conclusions than a multidisciplinary group—the specialty group will be systematically biased in favor of performing procedures in which it has a vested interest.^{2,3} For example, a group of vascular surgeons favored the use of carotid endarterectomy more than a mixed group of surgeons and medical specialists.⁴ Individuals' biases may be better balanced in multidisciplinary groups, and such balance may produce more valid guidelines. Ideally, the group should have at least 6 but no more than 12 to 15 members; too few members limit adequate discussion and too many members make effective functioning of the group difficult. Under certain circumstances (for example, in guidelines for broad clinical areas), it may be necessary to trade off full representation against the requirement of having a functional group.

Roles—Roles required within guideline development groups are those of group member, group leader, specialist resource, technical support, and administrative support. Group members are invited to participate as individuals working in their field; their role is to develop recommendations for practice based on the available evidence and their knowledge of the practicalities of clinical practice.

Skills needed for guideline development

- Literature searching and retrieval
- Epidemiology
- Biostatistics
- Health services research
- Clinical experts
- Group process experts
- Writing and editing

The role of the group leader is to ensure that the group both functions effectively (the group process) and achieves its aims (the group task). The process is best moderated by someone who is familiar with (though not necessarily an expert in) the management of the clinical condition and the scientific literature yet is not an advocate. He or she stimulates discussion and allows the group to identify where true agreement exists but does not inject his or her own opinions in the process. This requires someone with both clinical skills and group process skills. Using formal group processes rather than informal ones in group meetings produces different and possibly better outcomes.⁵⁻⁷

Identifying and assessing the evidence

Identifying and assessing the evidence is best done by performing a systematic review. The purpose of a systematic review is to collect all available evidence, assess its potential applicability to the clinical question under consideration, inspect the evidence for susceptibility to bias, and extract and summarize the findings.

What sort of evidence?

Identifying the clinical questions of interest will help set the boundaries for admissible evidence (types of study design, year of publication, etc.). For example, questions of the efficacy of interventions usually mean that randomized, controlled trials should be sought, while questions of risk usually mean that prospective cohort studies should be sought.

Where to look for evidence?

The first step in gathering the evidence is to see if a suitable, recent systematic review has already been published. Searching the *Cochrane Library* will also identify relevant Cochrane review groups, which should also be contacted to see if a review is in progress.

If a current systematic review is not available, a computer search of Medline and EMBASE is the usual starting point, using search strategies tailored to appropriate types of studies (though such strategies have been validated only for randomized, controlled trials.⁸ For example, randomized, controlled trials provide the best evidence to answer questions about the effectiveness of treatments, whereas prospective cohort studies generally provide the best evidence for questions about risk. The Cochrane Controlled Trials Register (part of the *Cochrane Library*) contains references to over 218,000 clinical trials that have been identified through database and hand searching; it should be examined early on in any review process. Checking the references in articles will show additional relevant articles not identified by the computer search, and having experts in the field examine the list of articles helps ensure that there are no obvious omissions. Additional

search strategies, including searches for articles published in languages other than English,⁹⁻¹¹ computer searches of specialized databases, hand searching of relevant journals, and searching for unpublished material will often yield additional studies, but the resources needed for such activities are considerable. The cost-effectiveness of various search strategies has not been established. It is best to match the scope of the search strategy to the available resources.

Assessing studies for relevance

Once studies have been identified, they are assessed for relevance to the clinical questions of interest and for bias.¹² Screening for relevance is often possible from the abstract;¹³ it narrows the set of studies to those needing a more detailed assessment. Using explicit rather than implicit criteria should improve the reliability of the process.

Summarizing evidence

Data are extracted from the relevant studies on the benefits, the harms, and (where applicable) the costs of the interventions being considered. These data are usually presented in a form that allows the designs and results of studies to be compared. Where appropriate, meta-analysis can be used to summarize the results of multiple studies.

Categorizing evidence

Summarized evidence is categorized to reflect its susceptibility to bias. This is a shorthand method of conveying specific aspects of the evidence to a reader of the guideline. A number of such "strength of evidence" classification schemes exist, but empirical evidence exists only for schemes that categorize effectiveness studies by study design.^{14, 15} The box shows a simple scheme for classifying the evidence that supports statements in practice guidelines and the strength

of the recommendations. Guideline developers should use a limited number of explicit criteria, incorporating criteria for which there is explicit evidence.

Factors contributing to the process of deriving recommendations

- The nature of the evidence (for example, its susceptibility to bias)
- The applicability of the evidence to the population of interest (its generalizability)
- Costs
- Knowledge of the health care system
- Beliefs and values of the panel

Translating evidence into a clinical practice guideline

The evidence, once gathered, needs to be interpreted (see box). Since conclusive evidence exists for relatively few health care procedures, deriving recommendations solely from such examples would lead to developing a guideline of limited scope or applicability.¹⁶ This could be sufficient if, for example, the guideline is to recommend the most strongly supported treatments for a given illness. More commonly, however, the evidence needs to be interpreted into a clinical, public health, policy, or payment context. Therefore, within the guideline development process, a decision should be made about how opinion will be both used and gathered.

Using and gathering opinion

Opinion will be used to interpret evidence and also to derive recommendations in the absence of evidence. When evidence is being interpreted, opinion is needed to assess issues such as the generalizability of the evidence—for example, to what degree evidence from small, randomized clinical trials or controlled observational studies may be generalized or to what degree results from a study in one population can be extrapolated to the population of interest in the guideline (for example, extrapolating a study from a tertiary, academic medical center to the community population of interest to potential users of the guideline).

Recommendations based solely on clinical judgment and experience are likely to be more susceptible to bias and self-interest. Therefore, after deciding what role expert opinion is to play, the next step is deciding how to collect and assess expert opinion. There is currently no optimal method for this, but the process needs to be made as explicit as possible.

Resource implications and feasibility

In addition to scientific evidence and the opinions of expert clinicians, practice guidelines must often take

Classification Schemes

Category of evidence:

- IA evidence for meta-analysis of randomized, controlled trials
- IB evidence from at least one randomized, controlled trial
- IIA evidence from at least one controlled study without randomization
- IIB evidence from at least one other type of quasi-experimental study
- III evidence from nonexperimental descriptive studies, such as comparative studies, correlation studies, and case-control studies
- IV evidence from expert committee reports or opinions or clinical experience of respected authorities, or both

Strength of recommendation:

- A directly based on category I evidence
- B directly based on category II evidence or extrapolated recommendation from category I evidence
- C directly based on category III evidence or extrapolated recommendation from category I or II evidence
- D directly based on category IV evidence or extrapolated recommendation from category I, II, or III evidence

into account the resource implications and feasibility of interventions.

Judgments about whether the costs of tests or treatments are reasonable depend on how cost effectiveness is defined and calculated, on the perspective taken (for example, clinicians often view cost implications differently than payers or society at large), and on the resource constraints of the health care system (for example, cash-limited public systems versus private insurance-based systems). Feasibility issues worth considering include the time, skills, staff, and equipment necessary for the provider to carry out the recommendations and the ability of patients and systems of care to implement them.

Grading recommendations

It is common to grade each recommendation in the guideline. Such information provides the user with an indication of the guideline development group's confidence that following the guideline will produce the desired health outcome. "Strength of recommendation" classification schemes (such as the one in the box) range from simple to complex: no one scheme has been shown to be superior. Given the factors that contribute to a recommendation, strong evidence does not always produce a strong recommendation, and the classification should allow for this. The classification is probably best done by the group panel, using a democratic voting process after group discussion of the strength of the evidence.

Reviewing and updating guidelines

Guidelines should receive external review to ensure content validity, clarity, and applicability. External reviewers should cover three areas: people with expertise in clinical content who can review the guideline to verify the completeness of the literature review and ensure clinical sensibility; experts in systematic reviews, guideline development, or both, who can review the method by which the guideline was developed; and potential users of the guideline, who can judge its usefulness.

The guideline can be updated as soon as each piece of relevant new evidence is published, but it is better to specify a date for updating the systematic review that underpins the guideline.

Conclusions

New advances in understanding the science of systematic reviews, the workings of groups of experts, and the relation between guideline development and implementation are all likely in the next 3 to 5 years.

We believe that 3 principles will remain basic to the development of valid and usable guidelines:

- The development of guidelines requires sufficient resources in terms of financial support and people with a wide range of skills, including expert clinicians, health services researchers, and group process leaders;
- A systematic review of the evidence should be at the heart of every guideline; and
- The group assembled to translate the evidence into a guideline should be multidisciplinary.

References

1. Woolf SH. An organized analytic framework for practice guideline development: using the analytic logic as a guide for reviewing evidence, developing recommendations, and explaining the rationale. In: McCormick KA, Moore SR, Siegel RA, eds. *Methodology perspectives*. Washington, DC: US Department of Health and Human Services, Agency for Health Care Policy and Research, 1994:105-13. (AHCPR publication number 95-0009.)
2. Kahan JP, Park RE, Leape LL, Bernstein SJ, Hilborne LH, Parker L, et al. Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Med Care* 1996; 34:512-23.
3. Coulter I, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels—a comparison of a multi- and single disciplinary panel. *Health Serv Res* 1995; 30: 577-91.
4. Leape LL, Park RE, Kahan JP, Brook RH. Group judgments of appropriateness: the effect of panel composition. *Quality Assur Health Care* 1992; 4:151-9.
5. Kosecoff JH, Kanouse DE, Rogers WH, McCloskey L, Winslow CM, Brook RH. Effects of the National Institutes of Health consensus development program on physician practice. *JAMA* 1987; 258:2708-13.
6. Shekelle PG, Schriger DL. Evaluating the use of the appropriateness method in the Agency for Health Care Policy and Research clinical practice guideline development process. *Health Serv Res* 1996; 31:453-68.
7. Shekelle PG, Kravitz RL, Beart J, Morger M, Wang M, Lee M. Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific or specific guidelines on physician decision making. *Health Serv Res* (in press).
8. Cochrane Review Handbook. In: Cochrane Collaboration. *Cochrane Library*. Issue 4. Oxford: Update Software, 1998.
9. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309:1286.
10. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995; 48:159-63.
11. Egger M, Zellweger-Zohner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; 350:326-9.
12. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-12.
13. Moher D, Jones A, Cook D, Jadad AR, Moher M, Tugwell P, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609-13.
14. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. *Stat Med* 1989; 8:441-54.
15. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II: Surgical. *Stat Med* 1989; 8:455-66.
16. Shekelle P. Assessing the predictive validity of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. *Int J Technol Assess Health Care* 1998; 14:707-27.

This paper originally appeared in the *BMJ* 1999; 318:593-6.