A finger is the first choice of site, and a non-disposable clip type probe is adequate for most purposes, particularly when single readings are to be taken or little movement is expected. Self adhesive probes are most useful for long term monitoring and when motion artefact is expected. Toes may be used instead of fingers, but poor perfusion is more likely. The lobe or pinna of the ear is the second choice of site with a clip probe. Fixation is often more difficult than with the digits, and care must be taken that the clip does not exert sufficient pressure to impair perfusion. It is usually easier to achieve adequate perfusion in the ear than in the digits.

The bridge of the nose, the lips, and the tongue have been advocated when perfusion is poor, but there is no evidence that they are any better than conventional sites. Reflectance probes, where emitter and detector are adjacent to each other, have been developed but are not recommended as they are susceptible to inaccuracy and artefact.

CHOICE OF INSTRUMENT

When choosing an instrument, consider the likely areas of use. If spot measurements are intended—for example, in general practice or in general wards—the emphasis should be on size, battery capability, and robustness. In contrast, if continuous monitoring of sick patients is expected then a good range of probes, provision of a waveform display, and comprehensive alarms will be important. The recording of data will require an analogue or serial connection or an internal memory. The instruments store either the average or the lowest value within an epoch that may range from one second to one minute. The latter is suitable if an overall impression of oxygenation is required but not when episodic hypoxaemia is being sought—such as

with sleep apnoea, where the periodicity of the apnoeas may be less than a minute.

- 1 Comroe JH, Bothelo S. Unreliability of cyanosis in recognition of arterial anoxemia. *Am J Med Sci* 1947;214:1-6.
- 2 Hay WW, Brockway JM, Eyzaguirre M. Neonatal pulse oximetry: accuracy and reliability. *Pediatrics* 1989;83:717-22.

Bibliography

- Severinghaus JW. Nomenclature of oxygen saturation. *Adv Exp Med Biol* 1994;345:921-3.
- Council on Scientific Affairs, American Medical Association. The use of pulse oximetry during conscious sedation. *JAMA* 1993;270:1463-8.
- Mardirossian G, Schneider RE. Limitations of pulse oximetry. *Anesthesia Progress* 1993;39:194-6.
- Young IH, Le Souef PN. Clinical oximetry. *Med J Aust* 1993;159:60-2.
- Williams AJ, Stein M. Screening for obstructive sleep apnea using pulse oximetry. *West J Med* 1992;157:175-6.
- Mendelson Y. Pulse oximetry: theory and applications for noninvasive monitoring. *Clin Chem* 1992;38:1601-7.
- Severinghaus JW, Kelleher JF. Recent developments in pulse oximetry. *Anesthesiology* 1992;76:1018-38.
- Dawson KP. The asthma clinical score and oxygen saturation. *Aust Clin Rev* 1991;11:20-1.
- Ralston AC, Webb RK, Runciman WB. Potential errors in pulse oximetry. I: pulse oximeter evaluation. *Anaesthesia* 1991;46:202-6.
- Webb RK, Ralston AC, Runciman WB. Potential errors in pulse oximetry. II: effects of changes in saturation and signal quality. *Anaesthesia* 1991;46:207-12.
- Ralston AC, Webb RK, Runciman WB. Potential errors in pulse oximetry. III: effects of interferences, dyes, dysaemoglobins and other pigments. *Anaesthesia* 1991;46:291-5.
- Pierson DJ. Pulse oximetry versus arterial blood gas specimens in long-term oxygen therapy. *Lung* 1990;168(suppl):782-8.
- Tobin MJ. Respiratory monitoring. *JAMA* 1990;264:244-51.
- Schnapp LM, Cohen NH. Pulse oximetry. Uses and abuses. *Chest* 1990;98:1244-50.
- Moyle JTS. *Pulse oximetry, principles and practice series*. London: BMJ Publishing Group, 1994.
- Clayton DG, Webb RK, Ralston AC, Duthie D, Runciman WB. A comparison of the performance of 20 pulse oximeters under conditions of poor perfusion. *Anaesthesia* 1991;46:3-10.
- Alexander CM, Teller LE, Gross JB. Principles of pulse oximetry: theoretical and practical considerations. *Anesth Analg* 1989;68:368-76.
- Tremper KK, Barker SJ. Pulse oximetry. *Anesthesiology* 1989;70:98-108.
- Wraith A. A review of pulse oximetry. *Society for Advancement of Anaesthesia in Dentistry Digest* 1993;10(2):22-9.

(Accepted 2 May 1995)

Development of review criteria: linking guidelines and assessment of quality

Richard Baker, Robin C Fraser

Review criteria are designed to enable clinicians and others to assess care. However, there is no established method for developing criteria, and they are often confused with guidelines. Criteria should comprise measurable activities that are appropriate for the setting in which they are to be used. They should also be based on research evidence and prioritised according to the strength of that evidence and effect on outcome. Good criteria can be used to aid implementation of guidelines by providing a standard against which to monitor performance and enabling clinical audit.

In the continuing debate about the most effective methods for assessing high quality care reference is often made to "guidelines" and "review criteria." Although the purpose of guidelines is to assist in making clinical decisions and criteria are used in the assessment of care¹ (box 1), these crucial distinctions are not always clearly made, leading to confusion about their development and application in clinical practice. Furthermore, much less attention has been given to methods for developing and using review criteria compared with guidelines. In improving care, sound measures for the assessment of quality are as necessary as "systematically developed statements to assist practitioner and patient decisions."¹ The

aims of this paper are, firstly, to make explicit the desirable attributes of criteria and, secondly, to propose a framework for linking them with the process of development of guidelines.

The respective roles of guidelines and criteria can be clarified by the following example. The guidelines of the British Hypertension Society state that "great emphasis should be placed on encouraging patients to stop smoking as the coexistence of smoking as an additional risk factor in hypertensive patients confers a much increased risk of subsequent cardiovascular events."² To convert this guideline into review criteria it would need to become "the records show that at least annually (a) there has been an assessment of smoking habit and (b) appropriate advice has been given to smokers." The criteria make clear what information is required to assess clinician compliance, how the information is to be obtained, and the time period in which smoking habit should be assessed. It illustrates how criteria used for assessment need to be more detailed and specific than guideline statements used to assist decision making.

The importance of establishing a method for developing criteria lies in part in the role they can have in the implementation of guidelines. Guidelines have been produced by clinicians (that is, doctors, nurses, and other professionals directly providing clinical care)

Eli Lilly National Clinical Audit Centre, Department of General Practice, University of Leicester, Leicester LE5 4PW
Richard Baker, director

Department of General Practice, University of Leicester, Leicester
Robin C Fraser, professor of general practice

Correspondence to: Dr Baker.

BMJ 1995;311:370-3

Box 1—Definitions of guidelines, criteria, and protocols for clinical practice

Clinical practice guidelines	Systematically developed statements to assist decisions for practitioner and patient about appropriate health care for specific clinical circumstances*
Review criteria	Systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services, and outcomes*
Protocol	A comprehensive set of criteria for a single clinical condition or aspects of organisation
Standard	The percentage of events that should comply with the criterion

*From Institute of Medicine.¹

for many years, but the current interest in their potential value is due to several new factors. Firstly, only recently has a method been suggested for the development of systematic guidelines based on evidence, which may offer one approach to incorporating the findings of research into routine clinical practice.^{3,4} Secondly, evidence has now become available to show that guidelines can encourage improvements in performance when coupled with effective implementation strategies.^{5,6} Thirdly, the NHS Executive has recognised the potential part clinical guidelines could play in improving clinical effectiveness if incorporated by purchasers and providers in contracts for care.⁷ Despite these developments, however, benefit from using guidelines does not automatically follow.⁸⁻¹⁰ Changes in the performance of clinicians depend not only on the guidelines themselves but also on the methods used to encourage their acceptance and use.^{5,6}

Criteria can be used in two ways to augment the implementation of guidelines. They can be used by purchasers and providers to monitor compliance with the recommendations of guidelines, with one or more of a range of methods being chosen when necessary to improve compliance. Criteria are also an essential component of clinical audit, which is itself one method that can be used by clinicians to help them comply with guidelines. In audit, information about performance is collected for comparison with explicit criteria,¹ and feedback of the findings to participants is then used either alone or in combination with other strategies to encourage appropriate changes in performance.¹¹⁻¹³

Attributes of criteria for monitoring and clinical audit

Some desirable attributes of medical review criteria have been proposed by the American Institute of Medicine's committee on clinical practice guidelines.¹ It should be remembered, however, that these medical review criteria are intended for use in health services research¹⁴ or in peer review programmes to assess the appropriateness of care provided by clinicians for reimbursement by organisations such as insurance companies. The process of developing such criteria for the external assessment of care is often detailed and prolonged.^{15,16} In Britain criteria are used mainly in the context of clinical audit by clinicians themselves to identify aspects of care that could be improved. The use of criteria by purchasers to monitor performance is still uncommon. Moreover, because they are used for different purposes the properties of criteria used in the American context are different from those used in Britain, whether by purchasers, providers, or clinicians.

Grimshaw and Russell have proposed a useful set of requirements for review criteria³ derived from Kessner *et al*.⁷ and Irvine.¹⁸ They recommend that criteria should be easy to define; relate to morbidity amenable to improvement by medical care; have a sound scientific basis; yield sufficient data for statistical

analysis; and span the range of morbidity, skills required, and resources specified by the guideline; also the effects of non-medical factors on performance should be understood. These proposals form a valuable starting point, but more consideration still needs to be given to which requirements are most important and whether they are all applicable. For example, in some types of monitoring of performance such as significant event audit or occurrence screening the need for statistical analysis would not arise.

In our opinion, the most important requirement of criteria—as with guidelines—is that they should be based on evidence from research (box 2). Indeed, the approach recommended for the development of systematic evidence linked guidelines can be directly applied to criteria. Evidence from research should be evaluated by using reputable methods^{3,19} to differentiate criteria for which there is strong supporting evidence from those for which evidence is less clear or completely lacking. To increase the confidence of practitioners in the practical value of the criteria and their impact on the outcome of care, information should be provided about how the review of the research literature was undertaken, including the database used and how research reports were evaluated before being accepted. Although criteria can be derived from reputable guidelines, few current guidelines are of sufficient scientific validity to recommend such a course of action routinely.⁷

Furthermore, as the strength of the supporting evidence from research for different criteria will vary there is a need to make clear which criteria are most or least important in influencing quality of care—that is, criteria need to be prioritised. We would recommend that this should be judged primarily on the strength of the evidence from research and, secondly, on the potential of a particular criterion for influencing the outcome of care. Some elements of care which have only a marginal impact on outcome may have strong evidence from research and vice versa. In any event, selected criteria should have a substantial impact on outcome backed by satisfactory evidence from research. A practical advantage of prioritisation is that it enables participants in audit to focus their energies selectively. This will ensure that audit is efficient with effort devoted only to those criteria that lead to maximum improvements in outcome.

Several classification systems for prioritising criteria have been suggested. For example, Donabedian proposed two types, categorical or contingent, depending on the strength of evidence from research.²⁰ Complex procedures for weighting criteria for external assessment have been developed.¹⁶ To meet the needs of clinical audit for a system which is easy to understand and practical in general use the Eli Lilly National Clinical Audit Centre has devised the classification of “must do,” “should do,” and “could do” criteria.²¹ The “must do” criteria are those for which there is solid evidence from research of substantial impact on outcome. “Should do” criteria are those for which there is good evidence, but the impact on outcome is less substantial or the evidence is less strong. “Could do” criteria are those elements of care for which evidence of impact is inconclusive. If it cannot be shown that an

Box 2—Attributes of criteria for assessment of quality

- Based on research evidence
- Prioritised according to strength of research evidence and influence on outcome
- Measurable—clear and precise
- Appropriate to the clinical setting

element of care is either beneficial or harmful in clinical or economic terms there is little to be gained by undertaking an audit to ensure that element of care is provided. Such criteria should not, therefore, be routinely included in assessments of the quality of care.

Criteria for audit or monitoring should also be measurable. Each criterion should specify clearly and precisely what is to be measured so that it is possible to determine compliance. For example, a criterion stating that "a patient with hypertension should avoid excessive alcohol intake" would be difficult to measure accurately as the term "excessive alcohol intake" is not precisely defined and the patient's habits cannot easily be directly observed. To be measurable the criterion might be worded "the records show that annually the patient has been advised to limit weekly alcohol intake to less than 21 units if male and less than 14 units if female."²²

While guidelines may allow some flexibility on the part of the clinician to take into account unusual circumstances, criteria have greater rigidity. In the assessment of care the response to most criteria will be one of three possible answers—yes (complied), no (not complied), or not known (missing information). The use of a standard—that is, the percentage of cases which should comply with the criterion (box 1)—is the mechanism by which allowances are made for the variability of patients and clinical settings. The standard should be set at a realistic level which can still trigger systematic efforts to improve care. For many criteria the ultimate objective may be a standard of 100%, but in the short term a lower standard may be chosen which is then gradually increased in a series of steps.

The criteria must also be appropriate to the setting in which care is provided as clinical probabilities are related to context. For example, investigation of patients who are diagnosed as hypertensive in the setting of general practice would probably not identify many with secondary hypertension. Patients referred to hospital with uncontrolled hypertension, however, might warrant investigation for secondary causes. Thus, criteria used in an audit of hospital managed hypertension would be—and should be—different from those in general practice. Furthermore, by ensuring that criteria are specific for the setting participants in audit are more likely to accept them as relevant and practicable in their day to day work.

A framework for linking guidelines and criteria

To evaluate the quality of care given to patients with a specific condition such as hypertension criteria are required which cover every important aspect of care. For example, they should include specific diagnostic criteria that must be satisfied before patients are labelled as hypertensive and management criteria for the modification of risk factors and the control of blood pressure. In particular, all criteria in the "must do" category must be included. A complete set of criteria constitutes a protocol (box 1). Audit protocols also need to include specific advice on setting standards and the practical steps to be taken to conduct an audit in such a way as to be confident of the reliability and validity of the findings. Therefore, the procedures for collection and analysis of data must also be described.

An increasing number of scientifically valid guidelines will probably become available in the next few years. Various approaches will be used to encourage their adoption by clinicians, including monitoring of contracts, audit, and the feedback of information about performance. It has been proposed that clinical guidelines developed by expert agencies could be adapted locally for use in audit.²² This approach can be justified only if the locally adapted guidelines comply

with the evidence from research, but no studies have yet been undertaken to evaluate the quality of guidelines that have been adapted for local use.²³ Likewise, criteria and protocols could be developed locally by doctors or other clinicians undertaking audit, but the validity of the protocols produced may be questionable. The evaluation of evidence from research requires both skill and time that may not be readily available in every locality. Guidelines could be used to identify criteria, but if adequate evidence linked guidelines are not available the task of developing protocols will be even more complex. Even if guidelines are available, they have not usually been developed or presented to facilitate the identification of criteria. For example, current guidelines do not usually indicate which elements of care should be prioritised into categories such as "must do," "should do," or "could do."

Ownership is commonly seen as an important influence on whether the criteria or guidelines are used and lead to improved care.^{22, 24} It is argued that ownership is achieved only if the criteria are developed by those who use them. In a recent review of rigorous evaluations of methods for the implementation of guidelines, however, the balance of evidence did not indicate that ownership was of critical importance.⁶ Furthermore, local development need not be the only way to ensure that clinicians accept and comply with the criteria. Ownership is facilitated by allowing clinicians the freedom to choose whether performance is judged against "must do" criteria or "must do" plus "should do" and to select the most appropriate standards for their circumstances.

The concentration of resources for development of protocols regionally or nationally would be more efficient and liable to lead to the use of more appropriate and rigorous methods. As national or regional guidelines linked to evidence are produced related protocols should be developed. The developers of a guideline might themselves devise a relevant protocol, but it is important that the specific circumstances of potential users of the protocol are taken into account. Most guidelines are not developed with a single clinical setting in mind, such as general practice or cardiology outpatient departments, and in most cases clinicians from the relevant setting must take the lead in defining the wording of criteria in the light of the guidelines and the research evidence. Locally, health commissions, audit groups, and other agencies concerned in the assessment of care can select a protocol which has been rigorously developed by an expert group for an aspect of care in which they are interested. Alternatively, if a set of adequate and approved guidelines have been issued they can then choose the criteria to be used, taking into account their prioritisation and the requirements of local clinicians. Local standards may also be set, and a systematic method of collecting and analysing information about compliance with the criteria can be used.

We hope that our clarification of the meaning and practical application of terms used will encourage the linkage of guidelines with criteria, thus providing a more effective approach to implementing the findings of research in everyday practice. Clinical practice based on evidence is more likely to be achieved if guidance to assist decision making and assessments of care are both used as related parts of an integrated system. The combination of an increasing number of scientifically valid guidelines, equally sound regional or national criteria, and an explicit mechanism used to ensure their appropriate use, such as that described in this paper, should facilitate the rational monitoring and audit of care.

1 Field M, Lohr KN, eds. *Institute of Medicine Guidelines for Clinical Practice. From development to use.* Washington, DC: National Academy Press, 1992.

- 2 Sever P, Beevers G, Bulpitt C, Lever A, Ramsay L, Reid J, *et al.* Management guidelines in essential hypertension: report of the second working party of the British Hypertension Society. *BMJ* 1993;306:983-7.
- 3 Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines. I. Developing scientifically valid guidelines. *Quality in Health Care* 1993;2: 243-8.
- 4 Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines. II. Ensuring guidelines change medical practice. *Quality in Health Care* 1994;3:45-52.
- 5 Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;342:1317-22.
- 6 Grimshaw J, Freemantle N, Grimshaw J, Wallace S, Russell I, Hurwitz B, *et al.* Developing and implementing clinical practice guidelines. *Quality in Health Care* 1995;4:55-64.
- 7 NHS Management Executive. *Improving clinical effectiveness*. Leeds: Department of Health, 1993. (EL(93)115.)
- 8 Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med* 1989;321:1306-11.
- 9 Jones RH, Lydeard S, Dunleavy J. Problems with implementing guidelines: a randomised controlled trial of consensus management of dyspepsia. *Quality in Health Care* 1993;2:217-21.
- 10 Neville RG, Clark RC, Hoskins G, Smith B. National asthma attack audit 1991-2. *BMJ* 1993;306:559-62.
- 11 Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. *JAMA* 1992;268:1111-7.
- 12 Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *BMJ* 1991;303:398-402.
- 13 Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J. Opinion leaders versus audit and feedback to implement practice guidelines. *JAMA* 1991;265:2202-7.
- 14 Brook RH, Park RE, Chassin MR, Solomon DH, Keesey J, Koscoff J. Predicting the appropriate use of carotid endarterectomy, upper gastrointestinal endoscopy, and carotid angiography. *N Engl J Med* 1990;323: 1173-7.
- 15 Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ* 1994;309:730-3.
- 16 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. *Med Care* 1994;32:755-70.
- 17 Kessner DM, Kalk CE, Singer AA. Assessing health care quality—the case for tracers. *N Engl J Med* 1973;288:189-94.
- 18 Irvine D. Managing for quality in general practice. London: King's Fund Centre, 1990. (Medical audit series 2.)
- 19 Oxman AD. Checklists for review articles. *BMJ* 1994;309:648-51.
- 20 Donabedian A. The quality of medical care. Methods for assessing and monitoring the quality of care for research and for quality assurance. *Science* 1978;200:856-63.
- 21 Baker R, Khunti K, Lakhani M. *Monitoring diabetes*. Leicester: Eli Lilly National Clinical Audit Centre. University of Leicester, 1993.
- 22 Haines A, Jones R. Implementing findings of research. *BMJ* 1994;308:1488-92.
- 23 Baker R. Audit and standards in new general practice. *BMJ* 1991;303:32-4.
- 24 North of England Study of Standards and Performance in General Practice. Medical audit in general practice: effects on doctors' clinical behaviour and the health of patients with common childhood conditions. *BMJ* 1992;304: 1480-8.

(Accepted 1 May 1995)

Commentary: The perils of checklist medicine

Iona Heath

This paper has a logic and clarity that does not seem to belong in the messy world where I and my patients strive, often against daunting socioeconomic odds, to be "co-producers of health."¹ The honourable intentions of the authors are evident, but the notion of defining and using criteria with which purchasers can monitor compliance with agreed guidelines implies levels of coercion and control which are potentially destructive of the fragile good that is the doctor-patient relationship.² Guidelines are constructed from evidence from research derived from studies of populations and are predicated on the notion of a composite patient which may have little immediate relevance to the troubled person who presents in the consulting room.³ In hospitals, diseases stay and patients come and go; in general practice, patients stay and diseases come and go. Guidelines depersonalise individual patients and turn them into diseases even in general practice.

The authors give a single example of what a review criterion will look like in reality. As far as I understand it, the evidence from research is that excessive intake of alcohol increases the patient's risk of raised blood pressure. This becomes translated into a review criterion with the suggested wording "the records show that annually the patient has been advised to limit weekly alcohol intake to less than 21 units if male and less than 14 units if female." I have objections at two levels. Firstly, there is evidence of the role of excessive alcohol in the aetiology of hypertension, but I am not aware of evidence to support the contention that advising patients annually to limit their alcohol consumption improves the control of raised blood pressure, let alone that a note in the records to this effect has any impact. Surely, this is to confuse process with outcome. Secondly, patients may resent the imposition of the doctor's agenda for the consultation at the expense of their own. The doctor will feel compelled to give this advice repeatedly and increasingly mechanistically whether the patient is a longstanding alcoholic or a renowned teetotaler. A year is a short time in general practice, and patients who never drink will not take it kindly when their doctor reminds them yet again of the perils of excess. The patient will be concerned about both the memory and the sensitivity

of the doctor. At the other end of the range, an alcoholic patient often faces social and economic problems which render the risks of hypertension unimportant. Will he or she be helped by being repeatedly informed of the recommended limits?

Patients, perhaps even more than doctors, make many perverse decisions. They persist in smoking and drinking to excess despite being aware of the risks to their future health; they refuse to comply with medication and technology, such as hearing aids, of proved efficacy. This seems to suggest that health, narrowly defined as the absence of disease, is not the pre-eminent aspiration of human endeavour.⁴ Doctors, managers, and politicians need to recognise that self determination, dignity, and autonomy are important components of a wider definition of health and that personal and socioeconomic pressures can make longevity less desirable.

Guidelines and review criteria can contribute to the pursuit of quality in clinical care, but if politicians and purchasers seek to implement them without acknowledging the complexities of real clinical settings they may do more harm than good and consume huge resources in the process.⁵

- 1 Hart JT. *Feasible socialism: the National Health Service, past, present and future*. London: Socialist Health Association, 1994.
- 2 Willis J. *The paradox of progress*. Oxford: Radcliffe Medical Press, 1995.
- 3 Royal College of General Practitioners. *The development and implementation of clinical guidelines*. Exeter: Royal College of General Practitioners, 1995. (Report from general practice 26.)
- 4 Toon P. *What is good general practice?* Exeter: Royal College of General Practitioners, 1994. (Occasional paper 65.)
- 5 McKee M, Clarke A. Guidelines, enthusiasms, uncertainty, and the limits to purchasing. *BMJ* 1995;310:101-4.

Correction

Ethical debate: Financial ties as part of informed consent to postmarketing research

The contributions to this debate by R E Ferner and Malcolm VandenBurg and Ian Dews (24 June, pp 1660-3) contained an error in the references. The second author of the guidelines on safety assessment of marketed medicine (references 2 and 1 respectively) should be the Royal College of General Practitioners not the Royal College of Physicians.

Caversham Group
Practice, Kentish Town
Health Centre, London
NW5 2AJ
Iona Heath, general
practitioner