Orthopaedic surgeons are increasingly expected to assess and record the outcome of their practice for reasons of research, audit and clinical governance. Having accepted that assessment is here to stay, how best should we go about it?

An outcome measure is essentially an assessment of change which judges how the patient is now, as compared with a previous occasion, such as before surgery. While the health status after treatment may be seen as the outcome of interest, it will usually be the magnitude of change which effectively represents the effect of treatment. Thus in order that the information gathered is of optimal use it needs to be collected at two intervals in time at least.

Health status can be assessed by a number of methods which in orthopaedics are classified as either objective or subjective. The former includes the measurement of radiological changes, and of strength and range of movement, which involve a clinician making a judgement about the patient. The subjective methods rely on obtaining responses directly from patients about their perceptions of health and illness. Questions relating to patient satisfaction are another consideration and do not necessarily accord with other outcome measures. These findings, when discordant, can often be explained by a failure to achieve the prior expectations of the patient.

Most methods of evaluation of outcome suit some circumstances and not others. It is important to choose the most appropriate measure for a particular task, condition and setting, otherwise the results can be misleading. Most investigators accept that a combination of objective and subjective measures is desirable in order to provide a complete assessment of health-related quality of life and overall outcome. This, however, needs to be balanced against the feasibility and the cost of gathering and processing the information. Lack of planning is likely to lead to poor rates of response and incomplete or ambiguous data which will render the exercise worthless.

An outcome measure must have particular properties. It must be demonstrably reliable, valid and sensitive or responsive to changes in clinical status. The substantiation of such aspects of measurement is not straightforward and methods of so doing have largely been developed by psychologists who specialise in this area, which is termed ‘psychometrics’.

Clinical researchers are well aware of the fallibility of objective assessments and it has become increasingly clear that clinical assessments of key aspects of outcome such as pain, physical function or range of joint movement, are often inaccurate and not reproducible. They may also overly represent the concerns of the clinician, rather than those of the patient. From a practical point of view, persuading patients to attend for review, particularly if they are working, live some distance away or are no longer inconvenienced by pain and disability, may prove to be difficult and requires much organisation.

However, a substantial amount of research carried out over the last decade has demonstrated that reports from patients can be reliable, valid and sensitive to clinical change. The administration of questionnaires is more straightforward to organise than are clinical assessments and, while reminders may be necessary, the rate of response is generally much higher. There are now a number of patient-based outcome measures which are potentially applicable to the analysis of results in orthopaedics, particularly of arthroplasty. The problem for orthopaedic surgeons concerns the appropriate choice of such instruments and two papers in this issue (pp 339-44 and 345-7) illustrate some of the important considerations relevant to the choice of outcome measure for evaluating patients undergoing arthroplasty of the knee.

There are two main types of patient-based outcome measure, namely, generic instruments, such as the SF-36, the Nottingham Health Profile and the Health Assessment Questionnaire (HAQ) which are intended to address a wide range of health problems, and condition-specific instruments which focus on patients’ perceptions in relation to a single condition. The latter include the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the Oxford Hip and Knee Scores. Such
condition-specific instruments are intended to be maximally sensitive to the outcomes in specific groups of patients. Nevertheless, some such groups are more narrowly defined than others and correspondingly, some instruments are more specific than others. For instance, the WOMAC was designed to measure the response to the treatment of osteoarthritis of the hip or knee, while the Oxford Hip and Knee Scores assess outcomes after arthroplasty of the hip or the knee. All these measures produce one or more continuous scores, either by simply adding ratings obtained on individual items on the questionnaire or by the application of an algorithm or weighing formula to grouped responses.

A generic instrument provides a more complete overall picture of physical and psychological health than a condition-specific one. However, such questionnaires tend to be quite long and may attract poor rates of response from particular groups such as the elderly. Many people will balk at the sight of a questionnaire with more than two or three pages and others will rapidly lose interest if it appears to have no relevance to them. It may be preferable to obtain three pages of fully completed questions than six pages which are only partially filled.

An additional concern with generic instruments is that they are sensitive to any changes in health. This may be of interest to the investigator, but if the primary concern is of a more specific nature, such as function after joint arthroplasty, other changes in general health will potentially act as interference (or ‘noise’) which may have the effect of obscuring the particular outcome of interest. Thus, if a patient develops a frozen shoulder 12 months after a very successful hip replacement, a generic instrument could well register the presence of severe pain whereas a hip-specific questionnaire should not.

The likelihood of developing other conditions increases over time and outcomes after surgical treatment are usually long-term. Measurable differences in the relative effectiveness of different joint prostheses may not become evident for eight to ten years. Therefore, if the purpose is to evaluate the outcome of the hip itself, the longer the intended period of follow-up, the more specific the outcome measure should be in order to reduce the effect of ‘noise’ as much as possible.

In a clinical trial of an arthroplasty, patient-based measures of outcome will serve as a useful adjunct to other methods, particularly survival analysis which usually employs revision as the outcome of interest. However, each method may bring different problems of interpretation. Within the context of survival analysis, the use of rates of revision as an indicator of outcome is very crude. The analysis will include information on people who have died or are lost to follow-up but in whom the prosthesis was still in place when they were last seen. Patient-based measures may be more informative and sensitive but are only available for those who are alive, contactable and compliant.

Techniques for combining both methods of evaluation have been described, but this analytical approach is still in its infancy.

In small-scale observational studies, confounding factors may cause problems in the evaluation of outcomes. When they are compared across different populations, with patients attending different hospitals or being operated on by different surgeons, the differences in case-mix give rise to further concern. Here, the inclusion of generic measures makes such comparisons more feasible and will improve cost-effectiveness.

Most condition-specific patient-based measures have been developed primarily to be used in large-scale trials. Since the more highly specific measures have only been in use for a relatively short time, evidence concerning the full extent of their usefulness is only starting to become available. One particular consideration is that scores have only ever been used to make comparisons in aggregate and have not been designed for the assessment and screening of individuals.

None of these instruments has been judged as regards its value in audit or clinical governance, and it is unlikely that evidence produced from outcome evaluation which was not considered of high enough quality would be acceptable when changes in clinical practice and decision-making are at issue.

In choosing an appropriate outcome measure, it must be clear as to the purpose for which the outcomes data are being obtained and the questions must be assessed to ensure that they are appropriate. It must also be remembered that patients will not simply do as instructed, particularly if the reasons are unclear to them. The questionnaire needs to be as clear and self-explanatory as possible since, although a covering letter may carefully explain the reasons for the enquiry, it must not be assumed that the patient will either read or understand it.

Since patients may suffer from more than one medical problem at the time of enquiry, the inclusion of an additional generic questionnaire may allow them to record other problems, while ensuring that the condition-specific assessment is also completed. It is always necessary to undertake pilot studies of the logistics and method of collection of data before embarking on a study so that potential problems can be uncovered early. For those who have little or no experience in the application of patient-based measures of outcome, collaboration with those who have is strongly recommended.

Outcomes assessment is here to stay. The process of collecting data of any value requires careful thought and planning and there will usually be a trade-off between the depth of evaluation and feasibility/cost. All outcome measures need to be reliable, valid and sensitive to change. Condition-specific outcome measures are calibrated for use in particular circumstances and populations and are not appropriate for use in other circumstances. A condition-specific measure, used appropriately, will be more sensitive to change in the outcome of interest than will a generic
measure of health status. Validated patient-based measures increase the feasibility of large-scale outcomes assessment and already exist for a number of orthopaedic conditions, particularly arthroplasty. In general, patient-based outcome measures are designed to be used in clinical trials in which scores are compared in aggregate. They are not meant for assessment and screening at the level of the individual.

References