Measuring quality of life in the evaluation of health care

This article provides an introduction to the measurement of quality of life or subjective health status. It considers difficulties in defining the concept and the uses of subjective health status measures and describes common approaches to measurement. The criteria used to select a measure for inclusion in a study are summarised and some of the more commonly used measures are described.

Date of acceptance: October 15 1997.

Sophie Staniszewska BSc(Hons), DP, is Research Fellow in Patient Evaluation and Involvement, RCN Institute, London.

TRADITIONALLY, the evaluation of medical treatment has used measures of morbidity and mortality based on clinical, laboratory and radiological tests.

However, over the past few decades there has been increasing emphasis on patient-assessed outcome measures in the evaluation of care (Jenkinson et al 1993). This change in emphasis reflects the broader approach to defining health embodied in the World Health Organization definition (WHO 1948, 1958), which includes physical, psychological and social wellbeing. In order to measure these broader aspects of health, a range of measures has been developed which aim to explore the patient's perception of his or her health status.

DEFINING QUALITY OF LIFE

The term 'quality of life' is difficult to define and it is often used interchangeably with terms such as 'health-related quality of life', 'subjective health status' and 'functional status', which are probably better descriptors of the instrument content (Fitzpatrick et al 1992). Such interchangeable use of terminology relates to the lack of a conceptual definition of quality of life or subjective health status, although most researchers agree it is a multidimensional construct and that there are a variety of approaches to its measurement (O'Boyle 1994).
A number of instruments have been created, in the form of questionnaires, which patients complete themselves or can be completed by an interviewer. These instruments vary in terms of the dimensions, or aspects, of health that they measure and in the items (questions) they use to do this (Wilkin et al 1993).

Although the concept of subjective health status is, obviously, dependent on the patient's own opinion and definitions vary, the content of some of the instruments is similar. Common dimensions include physical function, emotional function, social function, role performance, pain and other symptoms (Fitzpatrick et al 1992).

THE USE OF SUBJECTIVE HEALTH MEASUREMENT

Subjective health status measures can be used in a variety of ways, although some have been designed with a particular purpose in mind. This should be noted when attempting to select a measure for a particular study. Health status measures can be included in clinical trials to provide useful information about the effects of an intervention. In studies about a drug or therapy, inclusion of a subjective health status measure might allow differentiation between two therapies or treatments with marginal differences in survival and morbidity.

Subjective health status measurement can also be used to compare the outcomes between two different treatment modalities, such as using a drug or surgery to treat a condition (Spilker 1990). Although subjective health status measurement has been included in clinical trials, many of the trials which have claimed to measure health status or quality of life have not assessed the concept correctly or have assessed a single or limited aspect of health status (Fitzpatrick et al 1992).

Subjective health status measurement can also be used to understand the impact of a condition on a particular patient group. When the Short Form 36 (SF-36) was used as a measure of outcome in a sample of patients with four clinical conditions, each condition generated a distinctive profile with scores for each of the eight dimensions of the SF-36 (Garratt et al 1993). These profiles allowed researchers to identify the dimensions of health that a particular condition affected. In the case of the SF-36, data on a random sample of the population have been collected to provide a community norm. Data gained from studies can then be compared against these community norms and so can help with the interpretation of results (Jenkinson et al 1993).

Health status measures can also be used routinely in clinical practice to evaluate the health status of patients. It is possible that such measures could be incorporated as part of clinical audit to help monitor the outcomes of care (Jenkinson et al 1993, Ware et al 1993). One measure, the Dartmouth COOP Charts, was developed specifically to help clinicians in primary care to screen, assess and monitor patient function (Nelson et al 1987). It is claimed that they improve doctor-patient communication in a primary care setting by identifying topics which would otherwise go unnoticed (Wilkin et al 1993).

Health status measures have also been used in the evaluation of alternative methods of organising and funding health services. The first major project to undertake such work was the Medical Outcomes Study in the US which aimed to evaluate variations in types of healthcare provision using patient-assessed data (Tarlov et al 1989). There has been increasing emphasis on the inclusion of an economic component in the evaluation of care (Drummond et al 1987). Subjective health status measurement has been included through the application of
cost-utility analysis to assess benefits and inform resource allocation. Cost-utility analysis is a special form of cost-effectiveness analysis in which the measure of effect is quality-adjusted life years gained (QUALYs) (Spiegelhalter et al 1992, Torrance 1986).

It has been suggested that the large-scale collection of health status data could be carried out in order to supplement population mortality statistics, because the latter only provide limited information about the health of a population. Such information could be used to help monitor the health of different regions or countries (Jenkinson et al 1993). It is not clear how helpful such information would be in assessing health needs (Fitzpatrick et al 1992).

METHODS OF MEASUREMENT
It is possible to identify two main types of health status measurement.

Health index measures The first type is called a health index and it is designed to provide a unitary value of subjective health status so all the items in the questionnaire can be summed into a single score. Single indexes can be used in cost-utility analyses, so they are favoured by health economists. Information on treatment cost, change in quality of life as a result of the treatment and the expected length of life after the treatment can be entered into an equation to quantify health gain in terms of QUALYs (Jenkinson et al 1993). However, QUALYs have been criticised for the assumptions underlying their calculation, doubts have been expressed concerning quality of data, ethical objections concerning equity have been raised and the relevance of comparing different numbers of different groups of people with a single index has been questioned (Spiegelhalter et al 1992).

Health profile measures The second type of health status measure is called a health profile and it is designed to provide an indication of the impact of ill health on a range of areas, rather than providing a single score. It is not usually possible to sum health profiles as this is not their primary focus (Wilkin et al 1993). Some of the health profiles that have been developed concentrate on the effects of a particular aspect of ill health, for example, arthritis or pain. Examples of condition-specific scales are the Arthritis Impact Measurement Scale (Meenan et al 1980), measuring the health status of patients with rheumatic diseases, and the McGill Pain Questionnaire, which aims to represent the intensity of pain (Melzack 1975).

A number of generic health status profiles have also been developed. These assess a range of subjective health status dimensions which can be used with different types of patient, and are comprised of different dimensions of health status. A number of these measures have become popular and their psychometric properties have undergone relatively extensive testing. Examples of such measures include the SF-36 (Ware and Sherbourne 1992) and the Nottingham Health Profile (Hunt and McEwan 1980).

HOW TO SELECT A MEASURE
When selecting a health status measure for inclusion in a study, it is important to consider a number of criteria (Cox et al 1992, Fitzpatrick et al 1992, Streiner and Norman 1995). These are:

Multidimensionality The concept of health status is a multidimensional phenomena and this should be reflected in the health status measure (Fitzpatrick et al 1992).
Reliability Subjective health status measures should produce the same results with repeated use under the same conditions (Cox et al 1992). Although it is possible to assess reliability using the test-retest approach, it may be difficult to distinguish a change in health status from any measurement error. Reliability can be assessed by examining the internal consistency of items with the aim of assessing the degree of agreement between items. If the health status measure is administered by interview, then the inter-rater reliability (the extent of agreement between different observers) also needs to be calculated (Fitzpatrick et al 1992).

Validity Validity indicates whether the concept under examination is being measured, although this is difficult to assess because instruments measure a subjective phenomenon. The most basic way to assess validity is by checking whether a measure seems to include the range of relevant topics in a clear and unambiguous way. This should be tested with a wide range of individuals from a variety of backgrounds (Cox et al 1992). In addition, it is also important to consult studies which have examined patients' experiences of particular conditions. These can provide useful evidence concerning patients' perceptions of their condition. While the assessment of face validity is relatively subjective, a more formal approach involves the assessment of construct validity, or the pattern between the subjective health status measure and other more established measures (Fitzpatrick et al 1992).

Thus a hypothesis can be generated and the instrument tested to determine whether it reflects this hypothesis. For example, the construct validity of the SF-36 was checked to ensure that certain groups, such as older people and those with illnesses gained lower scores (indicating worse health) than other groups, such as younger people or those without illness (Jenkinson et al 1993). It is important to remember that when validity has been demonstrated for one particular use, it cannot be assumed to exist for other uses (Fitzpatrick et al 1992).

Sensitivity to change or responsiveness This refers to an instrument's ability to detect changes over time (Cox et al 1992). It is a crucial requirement for a subjective health status measure, particularly within clinical trials, evaluation research or cost-utility analysis. Sensitivity may be affected by the inclusion of items not relevant to a particular group, the assessment of areas or effects that remain relatively static or are not the aim of an intervention, or the measure may suffer from 'floor' or 'ceiling' effects (Fitzpatrick et al 1992).

Floor and ceiling effects refer to the response range and the method of scoring of an instrument. For example, if an instrument is not sensitive to lower levels of ill health, is scored from 0 (good health) to 100 (poor health), and is applied to a random sample of individuals, it would demonstrate a floor effect, as most patients would score 0. If the instrument was scored as 0 (poor health) and 100 (good health) and the same scenario applied, the instrument would suffer from a ceiling effect (Jenkinson et al 1993).

Appropriateness to question or use Selected measures should be appropriate to the research question or study aim. For example, it is important to consider whether an instrument reflects the range of effects that may be important to a particular patient group. One way of overcoming such a potential difficulty is to allow patients to select the dimensions important to them (Fitzpatrick et al 1992).

Practical utility Using subjective health status measures in different types of study should be practical and data should be meaningfully
used. For example, some health status measures are long and so regular use would place a heavy burden on the patient and be difficult for clinicians and decision makers to use. Some of the shorter measures, such as the SF-36, overcome some of these difficulties, although measures which are too brief may omit important aspects of health (Fitzpatrick et al 1992).

COMMON MEASURES OF SUBJECTIVE HEALTH STATUS
A number of generic measures have become commonly used to measure health status, including the Functional Limitations Profile, the Nottingham Health Profile, the SF-36 and the Dartmouth COOP Charts. It is important that any review which attempts to select a measure for a study considers the range of existing measures.

Functional Limitations Profile (FLP) This was developed from the Sickness Impact Profile (SIP), an American measure (Bergner et al 1976), and revised in 1981 (Bergner et al 1981). The FLP is a measure of illness-related behavioural dysfunction as assessed by the individuals' perceptions of the effect of an illness on the usual activities of daily living (Wilkin et al 1993). It is regarded as a gold standard in health status measurement (Jenkinson et al 1993). It contains 136 items in 12 categories; sleep and rest, eating, work, home management, recreation and pastimes, ambulation, mobility, body care and movement, social interaction, alertness behaviour, emotional behaviour, and communication. The FLP can be used for self-completion and takes about 20-30 minutes to complete (Wilkin et al 1993). The length of the FLP has implications for the patient, particularly when it is used with other measures.

Nottingham Health Profile (NHP) This is a most commonly used measure developed in the UK (Hunt and McEwan 1980). Its aim was to overcome some of the limitations of previously existing measures such as excessive length and ambiguous statements (Jenkinson et al 1993). It contains two sections which can be used independently of each other. The first contains 38 items and aims to measure perceptions of health on six dimensions; pain, physical mobility, emotional reactions, sleep, social isolation and energy. The NHP covers the more severe end of ill health, excluding minor health problems. The second section of the NHP asks patients to indicate whether their state of health affects their activities in seven areas of everyday living: paid employment, jobs around the house, social life, sex life, interests and hobbies and holidays. This has been less extensively used (Hunt et al 1985, Jenkinson et al 1993).

Dartmouth COOP Charts This project involved collaboration between primary care clinicians and the Department of Community and Family Medicine in a US hospital. The aim was to design a measure that would be reliable, valid, accurate and also simple and quick to complete. This was partly prompted by the need for practical solutions to assessing functional status (Jenkinson et al 1993).

The COOP Charts were developed for use in clinical practice and ambulatory/primary medical care. They are intended to provide clinicians with an efficient system for screening, assessing, monitoring and maintaining patient function in routine practice (Nelson et al 1987). There are nine charts (8.5 inches by 14 inches) each measuring a particular dimension of subjective health status. Respondents rate each chart according to how well they have felt over the past month or week. Charts cover physical condition, emotional condition, daily work, social activities, pain, change in condition, overall condition, social support and
quality of life. Scoring is very simple with each item rated on a one-to-five point ordinal scale. Each chart takes about 30-45 seconds to complete (Jenkinson et al 1993, Nelson et al 1990, Wilkin et al 1993).

**Short-Form 36** The SF-36 is the result of two large studies in the US. The Health Insurance Experiment (HIE) was undertaken to establish the most effective methods for measuring a broad range of functioning and subjective health status concepts and to compare the outcomes of different methods of delivering care. The Medical Outcomes Study (MOS) then developed and refined measures of health status and tested the feasibility of using self-administered patient questionnaires and generic scales in assessing the outcomes of health (Jenkinson et al 1993, Ware and Sherbourne 1992). The SF-36 contains one multi-item scale measuring eight dimensions; physical functioning, role limitations due to physical health problems, bodily pain, social functioning, general mental health, role limitations due to emotional problems, vitality and general health perceptions. Scores for each dimension of the SF-36 are calculated by summing across items in the same scale and then transforming the raw scores on a 0 to 100 scale (0 for poor health and 100 for good health). The SF-36 represents the health concepts its authors believe are most frequently included in widely-used health surveys (physical, social and role functioning, mental health and general health perceptions), together with pain and vitality (Ware and Sherbourne 1992).

**CONCLUSION**
This article has provided an introduction to the measurement of subjective health status. It has identified some of the difficulties in defining the concept and described some of the more popular generic health status measures. It is important that when selecting a measure for inclusion in a study, all the criteria outlined in this article are considered, to ensure that the most appropriate measure is selected.

**REFERENCES**


World Health Organization (1958) The First Ten Years of the World