# MEASURING HEALTH STATUS AND QUALITY OF LIFE

**Crispin Jenkinson**

## Introduction

This text introduces the concept of health status measurement, and outlines the requirments of measures in terms of reliability, validity and sensitivity to change.

## The purpose of medical care

The primary aim of medical care is to improve or maintain the overall functional capacity and general health of patients. It has historically concentrated on the diagnosis and treatment of physiological and anatomical conditions (Wasson et al, 1992) and for the most part this approach has tended to overlook global functioning, well being, and quality of life. Traditionally, evaluation of medical treatment has relied upon measures of morbidity and mortality, whilst medical practitioners have based judgements for intervention on traditional clinical, radiological and laboratory measures (Albrecht, 1994). This is anomalous given that clinically assessed outcomes of treatment do not always reflect those of patients (Blazer and Houpt, 1979; Jenkinson, 1994a). However, over the past few decades there has been a gradual shift away from this approach, and increasingly there is incorporation of patient based data into the evaluation of care (Geigle and Jones, 1990; Jenkinson, 1995).

The recognition of the patient's point of view as central to the monitoring and evaluation of medical care has brought with it numerous approaches to the measurement of subjective well being. The purpose of such evaluation is to provide more accurate assessments of individuals' or populations' health and the benefits and harms that may result from medical care (Fitzpatrick et al, 1992a). The ideal outcome of treatment is a return to the normal or usual quality of life for a given age and medical condition (Ware 1993; Silver, 1990). To evaluate the outcome of treatments subjective health measures can be utilised. However there are a wide variety of applications of health status measures, and the requirements of measures differ across these applications. Before considering the nature of subjective health measures it is worth considering the variety of applications in which data gained directly and systematically from the patients perspective could be of value.

## Applications of health status measures

Subjective accounts of functioning and well-being can be used in a variety of ways in the evaluation of health and medical care. Health status measures have been advocated as

appropriate tools for the screening of patients needing particular care or attention (Fitzpatrick, 1994). For example, health status measures were more accurate than traditional measures of health state in predicting long term morbidity and mortality in rheumatoid arthritis (Leigh and Fries, 1991). However, the data made available from such questionnaires should never be the sole grounds on which treatment decisions should be based. It has been suggested that before standardised health measures are routinely incorporated into clinical practice for individual patient assessment, and the evaluation of treatment options, then score confidence intervals must be fully documented (McHorney et al, 1994). The less reliable an instrument (i.e. the greater the level of measurement error) the wider the confidence intervals around any individual score. For example the short form 36 health survey questionnaire (SF-36), which has been the subject of considerable validation (Brazier et al, 1992; Jenkinson et al, 1993; Jenkinson et al, 1996a; Ware and Sherbourne, 1992; Ware et al, 1993) has been found to manifest wide confidence intervals. Wide confidence intervals call into question the validity of using brief multi-item scales for individual patient assessment. However, it has been suggested that on an individual basis health status data can act as an adjunct to the standard clinical interview, and may be useful for informing medical practitioners of the well being of individual patients in their care. This was one of the possible applications suggested by the designers of the Nottingham Health Profile (NHP) (Hunt et al, 1986), although no studies have documented its use in this manner. However, the Dartmouth COOP charts were designed with this purpose in mind (Nelson et al, 1990; Nelson et al, 1996; Wasson et al, 1992). Studies suggest that both patients and clinicians believe the use of the charts has led to improved interaction, and better treatment (Kraus, 1991) .

At the level of group analysis perhaps the most obvious use for standardised health measurement profiles is as outcome measures in randomised controlled trials. Whilst, the use of such measures in randomised control trials has been relatively limited there use in this arena of outcomes research is growing (Spilker, 1996). One potential problem with the use of such measures in trials relates the difficulties in determining meaningful differences on health assessment measures. This problem has probably been one reason for the relatively slow uptake of subjective health outcomes as primary end point measures in clinical trials and the relative paucity of trials including such measures has in turn has been suggested as one reason why many clinicians have been unwilling to utilise such measures in clinical practice (Bergner et al, 1992). In many instances clinical trials that have claimed to utilise quality of life instruments have done so with measures that are often limited in the range of dimensions covered, and have not been psychometrically validated (Aaronson, 1989). For results to be meaningful in such studies then it is imperative that psychometrically validated measures covering appropriate domains must be used. Such measures now exist and are increasingly being utilised in audit, and routine evaluation of health care (Wasson et al, 1992). Routine

systems to collect outcomes have been successfully demonstrated in England (Bardsley and Coles, 1992) and America (Lansky et al, 1992). Such systems have proved acceptable to clinicians, although widespread utilisation of 'outcomes management' systems has been slow to get off the ground. In part this is due to a lack of a agreement on what standardised measures should be used, and concern as to what, if any, effect such measurement will have upon clinical practice (Wasson et al, 1992).

Perhaps the most emotive use for health status measures is in the arena of cost containment and prioritisation. When utilised in cost utility studies measures are required from which a single figure can be derived, which can then be used to rank order treatments, or indeed patients. The most famous attempt that has as yet been made to derive a set of priorities on the basis of a cost benefit analysis was the Oregon experiment (Oregon Health Services Commission, 1991). It utilised the quality of well-being scale (Kaplan et al, 1987) and produced results that were so counterintuitive that informal procedures were used to reorder the resulting list. The value of utility measures in prioritisation is discussed more fully in this book in the chapter by Katherine Watson (Chapter Eight).

Health status measures also permit for the monitoring of populations health, or sub-samples within the population (Ware, 1992). Furthermore, comparisons of the health status of different countries can also be undertaken (Orley and Kuyken, 1994). Thus, there is currently interest in developing measures that can be used across cultures. This is the thrust of the work being undertaken by, for example, the WHOQOL Group (Szabo, 1996) and the IQOLA Group (Aaronson et al, 1992). The development of such instruments is not without its difficulties. It is certainly not enough to simply translate an instrument from one language to another. Careful checks are required to ensure that the meaning of questions remains the same. This can mean that it is actually necessary to ask somewhat different questions in different cultures to ensure that the same underlying concept is being tapped (Bullinger et al, 1995). Even more problematic is the possibility that issues of importance in one culture in relation to health are unimportant elsewhere (Hunt, 1995). However, if these problems can be overcome the potential exists of not only comparing the quality of life of different countries, which seems an undertaking of limited value, but also undertaking large multi-centre cross-cultural trials that incorporate self perceived health as a major outcome measure.

In the evaluation of medical care health assessment questionnaires can be used for a variety of purposes. It is important to realise, however, that different types of evaluation require different methods of assessment. A questionnaire such as the Sickness Impact Profile (SIP), of which the Anglicised version is the Functional Limitations Profile (FLP) (Bergner et al, 1981; Patrick and Peach, 1989), contains 136 questions and is thus not appropriate for routine

monitoring, or as an adjunct to the clinical interview, as it is simply too long and takes time for patients to complete and score. Similarly, cost benefit analyses require single index figures to be gained from health assessment questionnaires, and the use of multi-dimension questionnaires such as the SF-36 cannot be used for this purpose, although work is underway to attempt to gain a single index utility based figure from the profile of scores this measure produces (Brazier et al, 1994). When considering undertaking some form of evaluation of the quality of life of patients careful and informed choice of instruments is essential.

**Requirements of measures**

It would be naive to assume that designing a health assessment measure, or indeed any, questionnaire is an easy task (Oppenheim, 1992). A number of issues must be considered when designing a questionnaire. Instruments must be reliable, valid and sensitive to change.

***Reliability***

Questionnaires must be reliable over time. Thus, they should produce the same, or very similar results, on two or more administrations to the same respondents, provided, of course, there is good reason to believe that the health status of the patients has not changed. The difficulty with such a method of validating a questionnaire is that it often uncertain as to whether results that may indicate a questionnaire is unreliable are in fact no more than a product of real change in health status. Due to the potential difficulties in gaining an accurate picture of reliability in this way, many researchers adopt the Cronbach's alpha statistic (Cronbach, 1951), to determine internal reliability. Internal reliability refers to the extent to which items on a scale are tapping a single underlying construct, and therefore there is a high level of inter-item correlation. Assuming that such high levels of inter-item correlation are not a product of chance it is commonplace to assume that high a alpha statistic indicates the questionnaire is tapping an underlying construct and hence is reliable. There is, however, disagreement as to whether such a method can be viewed as appropriate for assuming a questionnaire is reliable over time (Ruta et al, 1993; Sheldon 1993).

***Validity***

Essentially there are four aspects to validity. Face validity, content validity, criterion validity and construct validity.

*Face validity* refers to whether items on a questionnaire superficially appear to make sense, and can be easily understood. This may seem a simple enough test for a questionnaire to pass, but there are ambiguities on some of the most respected and well utilised measures. For example the FLP requests respondents to complete the questionnaire with reference to today. They are thus asked to affirm or disaffirm items on the basis of how they are feeling today.

The basis of this judgement should, further, be related to their health. Let us take the example outlined in the FLP itself. It concerns the ability to drive. The statement given is `I am not driving my car'. Thus, if a respondent cannot drive a car today, and this is due to a health complaint then they should affirm the question 'I am not driving my car'. If they are not driving because they never learnt to do so, then they must answer this question in the negative. Thus, respondents are asked to make two judgements for each response. It could be argued that in such a long questionnaire (136 items) respondents may well forget or ignore the initial rubric. However, even if this were not the case, some questions don't make any sense on the basis of the rubric. For example, the item 'I have attempted suicide'. Respondents must tick 'Yes' or `No' to this item. Further, they must not tick 'Yes' if they have attempted suicide today, but did so because their spouse has been killed in a car accident (this is, afterall, not a problem with *their* health). Maybe it would be legitimate to tick `Yes' if the respondent reasoned that their mental health had been adversely affected by a relative's death, and they had attempted it today (just before filling in the questionnaire, in fact). There are more of these confused requests on the FLP. For example, respondents are clearly told to answer questions on the basis of today, and to only affirm questions which reflect some problem caused by health. This seems a tall order for some of the items, for example the item `I sleep or doze most of the time, day and night' sounds very much like a question relating to a broader time period than just the activities engaged in today. Questions such as these must make researchers sit back and take stock of how such questions are interpreted (or re-interpreted) by respondents if results from such instruments are to be of any meaningful use whatsoever. Further some items on the FLP can be influenced by place of administration. It has been suggested that individuals are more likely to affirm certain statements when in hospital than elsewhere. For example, items such as 'I stay in bed most of the day' are more likely to be affirmed in hospital than, for example, at home, though the need to stay in hospital may be induced by hospital requirements rather than by state of health per se (Jenkinson et al, 1993b; Ziebland et al, 1992). The FLP is certainly not the only questionnaire at which such criticisms can be aimed. The SF-36, which has gained increasingly popularity and use, contains this item:

The rubric reads;

> `During the past 4 weeks, have you had any of the following problems with your work or other regular activities as a result of any emotional difficulties (such as feeling depressed and anxious)?'

There are a number of items, of which one of them is;

> Didn't do work or other activities as carefully as usual (answer Yes or No).

Thus, respondents are informed in the rubric that the following items are a list of problems. The item would make perfect sense if it was phrased as `I worked less carefully than usual'. However, in it's present form it is difficult to know whether a `Yes' or `No' affirms the items content. A respondent could tick `Yes' assuming that this affirms the items content. Alternatively, to tick `No' in order to affirm the item would be grammatically more appropriate (and a double negative).

*Content validity* refers to choice of, and relative importance given to, items on a questionnaire. In a matter as fundamental as the selection of items a number of approaches are available to the potential designer. Broadly speaking, items can be developed by the researcher, or from studies of lay or patients populations, or any combination of these.

Both the NHP and SIP were developed on the basis of surveys of health perceptions of non-medically trained populations, with items weighted by a psychometric scaling technique. Hunt et al, the designers of the NHP, claimed that the scoring and weighting for seriousness of items on many health assessment questionnaires often reflect the values of the physician and not those of the lay person. As such they claimed that items tapping *subjective* health status should be generated from studies of lay people (Hunt, et al, 1986).

The NHP is a short easily administered questionnaire designed to overcome the potential criticism of many pre-existing instruments that both the domains and the questions contained in them are more a reflection of the assessments of clinicians and academic researchers than of lay people. To overcome this problem Hunt and her colleagues undertook a great deal of research upon lay people in order to ascertain what they believed to be the most salient dimensions of health that could be affected by illness. Six distinct dimensions emerged, Pain, Social Isolation, Energy, Sleep Disturbance, Mobility, and Emotional Reactions. Lay people were then asked to generate items that could be incorporated into these dimensions. Large numbers of statements were gained. A small number were then selected and weighted for inclusion in the questionnaire. To undertake this process, Hunt et al utilised a method similar to that which had been used by Bergner and her colleagues in the development of the SIP (Bergner et al, 1976, 1981).

There are 38 questions on the first section of the NHP (designed to assess subjective health state), and each item on the questionnaire carries a specific weight, ascribed to it by the developers, by an attitude scaling technique developed by Thurstone early this century (Thurstone 1928). Respondents can affirm all or none, or indeed any number, of the statements, as the developers claim they all tap an underlying attribute on any given dimension. It has been suggested that it is misleading to use a scaling technique such as

Thurstone's method to attempt to scale statements that are, or could be viewed as, factual (Edwards, 1957). The NHP contains factual statements, or ones that certainly could be viewed in this light (e.g. 'I'm unable to walk at all'). It is because of this that the NHP contains illogical groups of (factual) statements. It is possible, for example, to gain higher scores (indicating worse health) for less severe symptoms on the mobility dimension of the NHP. Some of the statements contained in the mobility section of the NHP logically preclude subjects responding to other items. For example an affirmation of the statement 'I'm unable to walk at all' (with a weight value of 21.30) technically precludes positive responses to some other aspects of mobility. For example, if a respondent affirms the statement that they are unable to walk, they should not, logically, be able to affirm the statements 'I can only walk about indoors' (weight 11.54), and 'I have trouble getting up and down stairs and steps' (weight 10.79), which make a total score of 22.33. Thus the score of a respondent with walking difficulties may exceed that of someone is unable to walk at all. Such an outcome can make the results gained from a questionnaire such as the NHP difficult to interpret (Jenkinson, 1991; Jenkinson, 1994b).

*Criterion validity* refers to the ability of an instrument to correspond with other measures. Whilst subjective health assessment questionnaires are constructed with the intention of measuring subjective perceptions a large number of items for such questionnaires have been designed by clinicians and researchers themselves. There exists the potential criticism, therefore, that such questionnaires may reflect more the interests of clinical judgement than those of patients themselves. For example, the Stanford Arthritis Center Health Assessment Questionnaire (Fries et al, 1980, 1982) and the Arthritis Impact Measurement Scales (Meenan et al, 1980) were developed in this way. To then attempt to provide information on the validity of the questionnaire by using existing clinical measures is to fall into the trap twice. Thus a clinician who designs a questionnaire and validates its scoring properties on the basis of existing medical and clinical measures could stand accused of not paying sufficient attention to the very phenomena they wish to measure, namely *subjective* (non-clinical) assessment. For example, in developing the AIMS, Meenhan et al (1982) argue that;

> 'the most commonly used measures, such as joint count and ESR, address disease activity only. The ARA Functional Classification, and Katz's Activities of Daily Living Scale, focus onfunctional abilities. These approaches fall far short of conceptualising or measuring health in the WHO sense of a physical, psychological and social state. Despite their long-standing use and widespread acceptance, disease activity and functional measures of outcome also have significant shortcomings as measurement tools. They have been accepted and disseminated primarily because they appear to be objective, and very little work has been done on documenting their

> measurement properties. The work which has been done, in fact, suggests that they are far from perfect", (Meenan, 1982, p785).

Thus, here Meenhan seems to be claiming that existing measurement tools for rheumatoid arthritis are far from perfect. This prompts him and his colleagues to develop a more refined questionnaire that will cover areas currently not tapped by existing medical assessments. He then uses clinical data to support the construct validity of the new instrument;

> "In the discriminant analysis, the clinical and health status measures were very similar in their ability to discriminate among the groups and between the treatment and no treatment. This provides further evidence that the health status measures performed as well as standard measures in this trial." (Meenan et al, 1984, p1351).

In this example we see that on the one hand measures such as clinical assessments are viewed as inaccurate and warranting further investigation, whilst on the other these measures can be used to bolster the case for the measurement properties of the AIMS. In the absence of a gold standard such practice has become commonplace. However, when items have been selected by clinicians rather than from surveys of lay people or patient groups, such results provide only limited support for the construct validity of the instrument. Put another way, the fact that the items of the questionnaire were chosen by Meenhan and his colleagues (the dexterity and pain items were developed by Meenhan, and other items were adapted from Katz's Index of Independence of Activities of Daily Living (Katz and Akpom, 1976; Katz et al, 1963, 1970) the RAND instruments developed by Ware and his colleagues (Ware et al, 1980) and the Index of Well Being (Kaplan et al, 1976)) and can be found to associate with existing clinical variables is perhaps to suggest that this instrument taps the dimensions of interest to clinicians. Indeed an updated AIMS questionnaire, the AIMS2, contains somewhat different items on the basis that not all appropriate dimensions of interest to patients were covered (Meenan et al, 1992). This would provide some support for the claim that questionnaires should be developed, at least in part, on surveys of lay people or appropriate patient groups. Certainly, this principle has been used in the development of other generic (Hunt et al, 1986) as well as disease specific measures (for example, Guyatt et al, 1987a; Peto et al, 1995).

*Construct validity* refers to the ability of an instrument to confirm expected hypotheses. Thus one would expect those who are ill, who are in lower social classes, and/or who make more frequent visits to their GP to gain scores indicating worse health than those who are well, in higher social classes and rarely visit their GP. Preliminary validation of questionnaires

involves ensuring questionnaires can discriminate between such groups (Brazier et al, 1992; Hunt et al, 1985, 1986; Jenkinson, 1993a, 1996a; Ware et al, 1993)

***Sensitivity to Change***

Sensitivity to change or 'responsiveness' is an important requirement of health status measures when utilised to evaluate the impact of medical interventions (Guyatt et al, 1987b. In general most attention in the development of health status questionnaires has been aimed at examination of the reliability and construct validity of measures (Fitzpatrick et al, 1993). However, recent work suggests that different measures can provide different pictures of change (Fitzpatrick et al, 1992b). This is in large part due to item content, which in part reflects the way in which items were selected (e.g. from patient interviews, or physician judgements) and the primary purpose of the instrument (e.g. the NHP was designed primarily as a population survey tool, whilst many disease specific measures are designed with outcome evaluation of treatment as their primary objective). Measures can reflect different conceptualisations of illness, health and disability (Ziebland et al, 1993). Many instruments often contain similarly labelled dimensions these are not necessarily tapping the same aspects of the attribute. Thus for example, the social dimension on the Arthritis Impact Measurement Scales asks respondents about social interactions over a longer time span than the FLP. The FLP is therefore more sensitive to recent small changes in social interactions than the Arthritis Impact Measurement Scales.

**Overview of measures**

Broadly speaking there have emerged two general approaches to the measurement of health status. The first is an attempt to develop instruments that provide a single global score of well being. These are designed in such a way as to permit all items on a questionnaire to be summed into a single health index. The other method is the development of questionnaires designed to measure a number of dimensions of health status.

*Single Index Measures of Health Status*

Single index measures of health status are designed to provide a single scale of health states. Perhaps the most famous example of such a measure is that described by Rosser (1988). This measure, designed initially to place in perspective the magnitude of change achieved in clinical trials, consists of two dimensions, disability and distress, in the form of a matrix. There are eight levels of disability and four levels of distress. For each combination of distress and disability the Rosser Index provides a single figure. The figures in the matrix were developed by Rosser on the basis of project where seventy subjects, including doctors,

nurses, psychiatric patients and healthy volunteers were asked to rank illness states and estimate relative severity. Whilst this scale gains a single index figure of health state, and, when used in routine clinical practice it takes only a few seconds for those familiar with its use to complete, it has to be borne in mind that the original weighting exercise, which produced the matrix, was undertaken on a very small sample. The valuations, therefore, are unlikely to reflect those of the population as a whole.

Whilst the Rosser Index was essentially developed for completion by physicians and staff, and not patients, attempts have been made to develop self completion single index measures. An attempt to gain a single index value of health state from the perspective of the patient is the Quality of Well Being Scale (QWB). The complex method of developing this questionnaire has been described fully elsewhere (Kaplan and Anderson, 1987) . The intention of this index is to combine mortality, morbidity and the benefits and side effects of treatment into a single global score. Such a global score can permit for the comparison of health states and treatments. Its value in comparing disease states is, however, dependent on gaining reliable prognoses. Without this latter information it is not possible to calculate potential 'well years' accruing from treatments. Another limitation of this questionnaire is it's length. It can take up to 15 minutes to complete, and the developers suggest it is administered by an interviewer, as the self completion version resulted in unreliable data (Anderson et al, 1986). As such the QWB does not lend itself to easy use in clinical settings, or for routine evaluation of care.

Attempts have been made to gain a questionnaire that is both short, easy to complete and a reliable indicator of health state. This has been a venture that has had few successes, though the Health Measurement Questionnaire (HMQ), which was derived from the Rosser Index, and the EuroQol have both had their advocates. The HMQ is a relatively brief, easy to complete questionnaire that elicits information on dimensions of mobility, capacity for self care, constraints on usual activities, social relationships and perceived stress. A single index figure is derived from responses to these domains. More information on this questionnaire is provided in Kind and Gudex (1991). A more widely used measure is the EuroQol EQ-5D (EuroQol Group 1990; Kind, 1996; Rosser and Sintonen, 1993). The EuroQol EQ-5D was developed by a multidisciplinary group of researchers from five European countries (EuroQol Group, 1990). There are five questions covering the areas of mobility, self-care, usual activity, pain/discomfort and anxiety/depression. Each question has three response categories; level 1 - 'no problems', level 2 - 'some problems' and level 3 - 'inability or extreme problems'. Overall health state can ostensibly be calculated from responses to these items. For example the response set '11111' indicates no problems with any of the five areas, and subsequently perfect overall health. There are in total 243 possible health states (i.e. $3^5$), and weighted

values have been assigned to each of these on the basis of national and international surveys (van Agt, et al, 1994). A single overall score can also be gained from the EuroQol thermometer, on which respondents mark their overall perceived health from 'Worst imaginable health State' to 'Best imaginable health state'. The development of the EuroQol is covered in detail in this text by Katherine Watson in chapter eight.

All of the above single index measures are based upon questionnaires which include fixed format items. However, a number of researchers have begun to analyse the possibility of asking patients to individually nominate areas of their life which have been adversely affected by health state, and to then assess the extent of this impact. The results from each of the items selected is then aggregated to form a single index figure. A variety of methodologies to this approach exist, but in essence they all permit each individual to select and weight their own chosen areas (McGee et al, 1991; Ruta et al, 1994). Such a procedure has the advantage of not imposing pre-existing definitions of health state upon respondents (Ruta and Garratt, 1994; Ruta et al, 1994). Research in this area has been undertaken in a number of groups including patients undergoing orthopaedic surgery, HIV positive patients, arthritis patients and patients reporting low back pain (Hickey et al, 1996; McGee et al, 1991; O'Boyle et al 1992; Ruta and Garratt, 1994; Ruta et al, 1994; Tugwell et al, 1990). Such methods are, like many research projects attempting to gain single index figures of health state, still in their infancy and hence not widely applied. A number of issues need to be addressed, such as whether respondents should select new dimensions each time they complete the questionnaire in longitudinal studies, whether aggregating potentially unrelated dimensions is an appropriate methodology and whether patients should select dimensions from a list (which perhaps does away with the whole philosophy of this approach) or simply select from any areas they think important. Such issues are at present receiving attention from a number of researchers, and whilst the generalised applicability of this new technique seems a long way off, it is an interesting and potentially worthwhile new approach to the whole field of subjective health measurement.

*Health Status Profiles*

Health status profiles are measures that tap a number of dimensions of functioning and well-being. Many instruments that have been developed are illness specific or are aimed at tapping a specific aspect of ill-health (such as pain or depression). However, the search for short comprehensive health status measures, which are able to detect differences between illness groups, and which are sensitive to changes over time, has produced remarkably few regularly utilised, and psychometrically validated, instruments. For example the McMaster Health Index Questionnaire (Chambers, 1988, 1993) has been used infrequently, evidence for the psychometric reliability and validity of the Functional Status Questionnaire (Jette et al, 1986) is very limited, the Duke-17 (Parkerson et al, 1991) has been rarely used and the Duke UNC Profile (Parkerson et al, 1981) has been criticised on psychometric grounds (Wilkin et al, 1992, 1993). The most frequently reported generic health measures have been the Sickness Impact Profile, (Bergner et al, 1976, 1981) the Functional Limitations Profile (Patrick and Peach, 1989), the Nottingham Health Profile (Hunt et al, 1985, 1986), and, more recently, the COOP Charts (Nelson et al, 1996; Wasson et al, 1992), and the Short-Form 36 (SF-36) (Brazier et al, 1992; Jenkinson 1996a, 1996b; Ware and Sherbourne, 1992; Ware et al, 1993, 1994) and Short-Form 12 (Ware et al, 1995a; Ware et al, 1995b; Ware et al, 1996). These measures cover a wide variety of dimensions of health status and are not primarily designed to give a single index of health status but to provide a profile of scores. However, for all these measures methods of data reduction have been suggested (see Figure 1 for attributes of these questionnaire).

**Discussion**

Single index figures of health status appeal to those who wish to compare different treatments and interventions. However whilst such single index figures give the impression of comparability between illness states and treatments they may do so unfairly. For example the EuroQol (EuroQol Group, 1990; Kind, 1996) questionnaire does not contain a dimension evaluating sleep disturbance, and a treatment aimed primarily at improving this dimension of health may not appear to have been efficacious if assessed by this measure.

Single index figures gained from patient generated measures, such as those of Ruta and Garratt (1994) and O'Boyle et al (1992) may overcome the above criticism. Essentially the dimensions chosen by patients are seen as paramount, and so if a patient is primarily concerned about the impact of illness on their sleep patterns, this will be incorporated in the measure. However, difficulties arise here. At initial interview a patient may claim their quality of life in five areas is affected. At follow up, these areas may have improved, so if the patient completes the questionnaire using the same dimensions chosen at time one an

improvement in health status will be apparent. However, side effects of drug treatment may have influenced other aspects of the respondent's life, and the patient's overall quality of life may not have improved at all. When using such a measure it is therefore appropriate to also include a generic instrument so as to ensure as wide as possible coverage of health related dimensions.

Generic measures, such as the FLP/SIP, the SF-36 Health Survey Questionnaire and the NHP, indicate clearly which dimensions of health status are being measured, but the dimensions included may not be appropriate in the assessment of every intervention. For example, the FLP, despite having 12 dimensions, lacks a specific category measuring pain. Results from generic measures can, of course, be compared with data from other populations and illness groups. For example, normative data can be used to compare the health status of a particular patient group with that of the general population (Ware, 1993). However, it is still important that disease specific measures are used alongside such generic measures, as disease specific measures are, by their very nature, likely to tap particular aspects of ill health that are unique to particular illnesses.

Ceiling and floor effects must be considered. The NHP has been criticised because it detects only the severe end of ill health, and thus most respondents score zero on many, if not all, of the six dimensions of the questionnaire (Kind and Carr-Hill, 1987). The items on the questionnaire were chosen to represent severe health states, and so individuals who have mild to moderate illness may not be detected with this instrument. In a study of change over time, respondents with minor ailments may improve, but if their initial score on dimensions of the NHP was zero, such improvement may not be detected (floor effect). Similarly respondents may score as maximally ill on a health measurement questionnaire. However, the extent of their illness state may still not be fully reflected in the questionnaire. Such severely ill respondents would fall beyond the measurement range. Thus, while these patients may improve over time, it is still possible they may continue to score as maximally ill on the questionnaire (ceiling effect). Such floor and ceiling effects are more likely to be found on instruments with small numbers of items (Bindman et al, 1990).

Related to floor and ceiling effects is another important aspect of health status measures: sensitivity to change or 'responsiveness'. For health status measures to be useful in evaluating the impact of medical interventions they must be 'sensitive'. It is thus imperative, when selecting a measure, to determine the exact nature of the questions asked and the time scales utilised. For example, a questionnaire such as the NHP, designed to tap the extreme end of ill health, is unlikely to be sensitive to small changes in health status among patients with minor illnesses.

Furthermore, in longitudinal studies it is important that the mode of administration of questionnaires is kept consistent. For example, due to the nature of some of the items in the FLP, respondents may gain higher scores in hospital than as out-patients or when at home, and such scores may not actually reflect health state. Items such as 'I stay in bed more' are more likely to be affirmed in hospital, and may not accurately reflect the impact of the illness per se on a person's life (Jenkinson et al, 1993b)

**Conclusion**

It is important to note that subjective health measurement questionnaires are not designed to be used as substitutes for traditional measures of clinical endpoints. On the contrary, they are intended to compliment existing measures and to provide a fuller picture of health state than can be gained by medical measures alone. However, to be useful such measures must be carefully chosen. Health status measures can provide a useful adjunct to the data traditionally obtained from mortality and morbidity statistics, or from traditional clinical and laboratory assessments, but careful consideration must be given to the choice of measures. At present it seems reasonable to assume that health status measures may permit scientific questions to be answered fully in the context of clinical trials, and, in time, they may find their way into routine use. However, the results obtained from such measures must be made intuitive and meaningful to clinicians, as well as to researchers, and adequate care must be taken to ensure appropriate measures tapping relevant domains are being utilised. Subjective health status measurement could provide much needed data on the impact of clinical interventions on the day to day lives of patients; done without due care of the pitfalls, however, such data could be irrelevant, or misleading.

**FUNCTIONAL LIMITATIONS PROFILE (FLP)/SICKNESS IMPACT PROFILE (SIP)**

**No of items:** 136

**No of dimensions:** 12

**Dimensions:**
Ambulation
Body care and movement
Mobility
Household management
Recreation and pastimes
Social interaction
Emotion
Alertness
Sleep and rest
Eating
Communication
Work

**Other:**
- The FLP is the Anglicised version of the SIP.
- An overall single index score can be derived from the FLP/SIP, as can a psycho-social dimension score and a physical dimension score.
- Note scoring rules differ slightly for the FLP and SIP.

**NOTTINGHAM HEALTH PROFILE (NHP)**

**No of items:** 38

**No of dimensions:** 6

**Dimensions:**
Energy
Pain
Emotional reactions
Sleep
Social isolation
Physical mobility

**Other:**
- The original NHP contained a second section but the developers no longer recommend its use (Hunt and McKenna, 1991)
- A single index (the NHP distress index) can be created from a sub-set of the items (see McKenna et al, 1993).

**SHORT FORM 36 (SF-36) HEALTH SURVEY**

**No of items:** 36

**No of dimensions:** 8

**Dimensions:**
Physical functioning
Role limitations due to physical problems
Role limitations due to emotional problems
Social functioning
Mental health
Energy
Pain
Health Perception

**Other:**
- Two summary scores can be derived from the SF-36: the physical component summary (PCS) and mental component summary (MCS). For further information see Ware et al, 1994; Jenkinson et al, 1996a.

**SHORT FORM 12 (SF-12) HEALTH SURVEY**

**No of items:** 12

**No of dimensions:** recommended for the derivation of the two summary scores although the original eight tapped in the SF-36 can be obtained (this is not recommended)

**Dimensions:**
Designed to provide the PCS and MCS (see SF-36 column), but can provide eight dimensions of Sf-36

**Other:**
- The developers do not recommend the SF-12 for use where the eight dimensions are required, but in instances when only the PCS and MCS scores are required (see SF-36 column). The SF-12 was designed to provide these scores yet in a shorter form instrument.

**COOP CHARTS**

**No of items:** 9

**No of dimensions:** 9

**Dimensions:**
Physical fitness
Feelings
Daily activities
Social activities
Change in health
Overall health
Social support
Quality of life
Pain

**Other:**
- There is only one item pre dimension.
- The charts were intended for use in the clinical interview.
- A version for children has been developed (see Nelson et al, 1996)

**Figure 1:** Properties of some commonly used generic health status measures

## REFERENCES


Aaronson N. (1989) Quality of life assessment in clinical trials: methodologic issues. *Controlled Clin Trials*, 10: 195-208S.

Aaronson, N.K., Acquadro, C., Alonso, J., Apolone, G., Bucquet, D., Bullinger, M., Bungay, K., Fukuhara, S., Gandek, B., Keller, S., Razavi., Sanson-Fisher, R., Sullivan, M., Wood-Dauphinee, S., Wagner, A., and Ware, J.E. (1992) International Quality of Life Assessment (IQOLA) Project. *Quality of Life Research*, 1: 349-51.

Albrecht, G. (1994) Subjective health assessment, in Jenkinson, C. (ed) *Measuring Health and Medical Outcomes.* London: UCL Press.

Anderson J.P., Bush J.W., and Berry C.C. (1986) Classifying function for health outcome and quality of life evaluation. *Medical Care,* 24: 54-69.

Bardsley, M. and Coles, J. (1992) Practical experiences in auditing patient outcomes. *Quality in Health Care*; 1: 124-30.

Bergner, M., Bobbitt, R.A., Kressel, S., Pollard, W.E., Gilson, B.S. and Morris J.R. (1976) The Sickness Impact Profile: conceptual formulation and methodological development of a health status measure. *International Journal of Health Services*, 6: 393-415.

Bergner, M., Bobbitt, R.A., Carter, W.B. and Gilson B.S. (1981) The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. *Medical Care*, 18, 787-805.

Bergner, M., Barry, M.J., Bowman, M.A., Doyle, A., Guess, H.A. and Nutting, PA. (1992) Where do we go from here? Opportunities for applying health status assessment measures in clinical settings. *Medical Care,* 30 (Supplement): MS219-MS230.

Bindman, A.B., Keane, D. and Lurie N. (1990) Measuring health changes among severely ill Patients: the floor phenomenon. *Medical Care,* 28: 1142-52.

Blazer, D. and Houpt, J. (1979) Perception of poor health in the healthy older adult. *Journal of the American Geriatrics Society,* 27: 330-34.

Brazier, J.E., Harper R., Jones, N.M.B., O'Cathain, A., Thomas, K.J., Usherwood, T. and Westlake, L. (1992) Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British Medical Journal,* 305: 160-4.

Brazier, J.E., Usherwood, T., Harper, R., Jones, N. and Thomas, K. (1994) Deriving a single index measure from the Short Form 36 health survey (abstract). *Journal of Epidemiology and Community Medicine.*

Bullinger, M. (1995) in Guggenmoos-Holzmann, I., Bloomfield, K., Brenner, H. and Flick, U. (eds) *Quality of Life and Health: Concepts, Methods and Applications.* Berlin: Blackwell-Wissenschafts.

Chambers, L. (1988) The McMaster Health Index Questionnaire - an update, in Walker, S.R. and Rosser, R. M. (eds) *Quality of Life: Assessment and Application*. Lancaster: MTP.

Chambers, L. (1993) The McMaster Health Index Questionnaire - an update, in Walker, S.R. and Rosser, R. M. (eds) *Quality of Life Assessment: Key Issues in the 1990's*. London: Kluwer.

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrica*, 16: 297-334.

Edwards, A. (1957) *Techniques of Attitude Scale Construction*. Englewood Cliffs, New Jersey: Prentice Hall.

EuroQol Group (1990) EuroQol - A new facility for the measurement of health related quality of life. *Health Policy*, 16: 199-208.

Fitzpatrick, R. (1994) Applications of health status measures, in Jenkinson, C. (ed) *Measuring Health and Medical Outcomes.* London: UCL Press.

Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D. and Cox D. (1992a) Quality of life measures in health care. I: applications and issues in assessment. *British Medical Journal*, 305: 1074-77.

Fitzpatrick, R., Ziebland, S., Jenkinson, C., Mowat A. and Mowat A. (1992) The importance of sensitivity to change as a criterion for selection of health status measures. *Quality in Health Care,* 1; 89-93

Fitzpatrick, R., Ziebland, S., Jenkinson, C., Mowat, A. and Mowat, A. (1993) A comparison of the sensitivity to change of several health status measures in rheumatoid arthritis. *Journal of Rheumatology,* 20: 429-436.

Fries, J.F., Spitz, P.W., Kraines, R.G. and Holman H.R. (1980) Measurement of Patient Outcome in Arthritis. *Arthritis and Rheumatism,* 23: 137-145.

Fries, J.F., Spitz, P.W. and Young, D.Y. (1982) The Dimensions of Health Outcomes: the Health Assessment Questionnaire, Disability and Pain Scales. *Journal of Rheumatology,* 9: 789-93.

Geigle, R. and Jones, S.B. (1990) Outcomes measurement: a report from the front. *Inquiry*; 27: 7-13.

Guyatt, G.H., Berman, L.B., Townsend, M., Pugsley, S.O. and Chambers, L. (1987a) A measure of quality of life for clinical trials in chronic lung disease. *Thorax*, 42; 773-8.

Guyatt, G., Walter, S. and Norman, G. (1987) Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases*, 40: 171-8.

Hickey, A.M., Bury, G., O'Boyle, C., Bradley, F., O'Kelly, D. and Shannon, W. (1996) A new short form quality of life measure (SEIQoL-DW): application in a cohort of individuals with HIV/AIDS. *British Medical Journal*, 313; 29-33.

Hunt, S. (1995) Cross-cultural comparability of quality of life measures, in Guggenmoos-Holzmann, I., Bloomfield, K., Brenner, H. and Flick, U. (eds) *Quality of Life and Health: Concepts, Methods and Applications.* Berlin: Blackwell-Wissenschafts.

Hunt, S., McEwen, J. and McKenna, S. (1985) Measuring health status: a new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners*, 35: 185-88.

Hunt, S., McEwan, P. and McKenna, S. (1986) *Measuring Health Status.* London: Croom Helm.

Hunt, S. and McKenna, S. (1991) *The Nottingham Health Profile User's Manual, Revised Edition.* Manchester: Galen Research and Consultancy.

Jenkinson, C. (1991) Why are we weighting? A critical analysis of the use of item weights in a health status measure. *Social Science and Medicine,* 32: 1413-16.

Jenkinson, C. (1994a) Measuring health and medical outcomes: an overview, in Jenkinson, C. (ed) *Measuring Health and Medical Outcomes.* London: UCL Press.

Jenkinson, C. (1994b) Weighting for ill health: the Nottingham health profile, in Jenkinson, C. (ed) *Measuring Health and Medical Outcomes.* London: UCL Press.

Jenkinson, C. (1995) Evaluating the efficacy of medical treatment: possibilities and limitations. *Social Science and Medicine,* 41: 1395-1403.

Jenkinson, C., Coulter, A. and Wright, L. (1993a) Short Form 36 (SF 36) health survey questionnaire. Normative data for adults of working age. *British Medical Journal,* 306: 1437-40.

Jenkinson, C., Layte, R., Wright, L. and Coulter, A. (1996a) *The UK SF-36: An Analysis and Interpretation Manual.* Oxford: Health Services Research Unit, University of Oxford.

Jenkinson, C., Layte, R., Wright, L. and Coulter, A. (1996b) Evidence for the sensitivity of the SF-36 health status measure to inequalities in health. *Journal of Epidemiology and Community Health*, 50: 377-80.

Jenkinson, C., Ziebland, S., Fitzpatrick, R., Mowat, A. and Mowat, A. (1993b) Hospitalisation and its influence upon results from health status questionnaires. *International Journal of Health Sciences*, 4: 13-18.

Jette, A.M., Davies, A.R., Cleary, P.D., Calteins, D.R., Rubenstein, L.V., Fink, A., Kosekoff, J., Young, R.T., Brook, R.H. and Delbonco, T.L. (1986) The Functional Status Questionnaire: reliability and validity when used in primary care. *Journal of General and Internal Medicine*, 1: 143-9.

Kaplan, R.M., Bush, J.W. and Berry, C.C. (1976) Health status: types of validity and the Index of Well-Being. *Health Services Research,* 11: 478-507.

Kaplan, R.M. and Anderson, J.P. (1987) The quality of well-being scale: Rationale for a single quality of life index, in Walker, S.R., and Rosser, R. (eds) *Quality of Life: Assessment and Application*. Lancaster: MTP/Kluwer.

Katz, S. and Akpom, C.A. (1976) Index of ADL. *Medical Care,* 14: 116-18.

Katz, S., Ford, A.B., Moskowitz, R.W., Thompson, H.M. and Svec, KH. (1963) Studies of illness in the aged. The Index of ADL: a standardised measure of biological and psychosocial function. *Journal of the American Medical Association,* 185: 914-919.

Katz, S., Downs, T.D., Cash, H.R. and Grotz R.C. (1970) Progress in development of the Index of ADL. *Gerontologist,* 10: 20-30.

Kind, P. (1996) The EuroQol Instrument: an index of health related quality of life, in Spilker, B. (ed) *Quality of life and Pharmacoeconomics in Clinical Trials, Second Edition.* Philadelphia: Lippincott-Raven.

Kind, P. and Carr-Hill, R. (1987) The Nottingham Health Profile: a useful tool for epidemiologists? *Social Science and Medicine,* 25: 905-10.

Kind, P. and Gudex, C. (1991) *The HMQ: Measuring Health Status in the Community. Centre for Health Economics Discssion Paper, number 93.* York: University of York, Centre for Health Economics.

Kraus, N. (1991) *The InterStudy Quality Edge, Volume 1, Number 1*. Excelsior, Minneapolis: InterStudy.

Lansky, D., Butler, J.B.V. and Frederick W.T. (1992) Using health status measures in the hospital setting: from acute care to 'outcomes management'. *Medical Care,* 30 (Supplement): MS57-MS73

Leigh, P. and Fries, J. (1991) Mortality predictors among 263 patients with rheumatoid arthritis. *Journal of Rheumatology,* 18: 1298-306.

McGee, H.M., O'Boyle, C.A., Hickey, A., O'Malley K. and Joyce, C.R.B. (1991) Assessing the quality of life of the individual: the SEIQoL with a healthy and gastroeneterology unit population. *Psychological Medicine,* 21: 749-59.

McKenna, S., Hunt, S., and Tennant, A. (1993) The development of a patient-completed index of distress from the Nottingham health profile: a new measure for use in cost-utility studies. *British Journal of Medical Economics,* 6: 13-24.

McHorney CA, Ware JE, Lu JF. The MOS 36-Item short-form health survey (SF-36): III. (1994) Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care,* 32: 40-66.

Meenan, R.F., Gertman, P.M. and Mason J.H. (1980) Measuring health status in arthritis: the Arthritis Impact Measurement Scales. *Arthritis and Rheumatism,* 23: 146-52.

Meenhan, R.F., Gertman, P.M., Mason, J.H. and Dunaif, R. (1982) The Arthritis Impact Measurement Scales: further investigations of a health status measure. *Arthritis and Rheumatism,* 25: 1048-53.

Meenhan, R.F., Anderson, J.J., Kazis, L.E., Egger, M.J., Altz-Smith, M., Samuelson, C.O., Willkens, R.F., Solsky, M.A., Hayes, S.P., Blocka, K.L., Weinstein, A., Guttadauria, M., Kaplan, S.B. and Klippel, J. (1984) Outcome assessment in clinical trials: evidence for the sensitivity of a health status measure. *Arthritis and Rheumatism,* 27: 1344-1352.

Meenan, R.F., Mason, J.H., Anderson, J.J., Guccione, A.A. and Kazis, L.E. (1992) AIMS2: the content and properties of a revised and expanded Arthritis Impact Measurement Scales health status questionnaire. *Arthritis and Rheumitism*, 35: 1-10.

Nelson, E.C., Landgraf, J.M., Hays, R.D., Wasson, J.H. and Kirk, J.W. (1990) The Functional Status of Patients: How can it be Measured in Physicians' Offices? *Medical Care* 1990; 28: 1111-26.

Nelson, E.C., Wasson, J.H., Johnson, D.J. and Hays, R.D. (1996) Dartmouth COOP functional assessment charts: brief measures for clinical practice, in Spilker, B. (ed) *Quality of life and Pharmacoeconomics in Clinical Trials, Second Edition.* Philadelphia: Lippincott-Raven.

O'Boyle, C., McGee, H., Hickey, A., O'Malley, K., Joyce, C.R.B. (1992) Individual quality of life in patients undergoing hip replacment. *Lancet* , 339: 1088-1091.

Oppenheim, A.N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement.* London: Pinter.

Oregon Health Services Commission. (1991) *Prioritization of Health Services*. Salem: Oregon Health Commission.

Orley, J. and Kuyken, W. (eds) (1994) *Quality of Life Assessment: International Perspectives.* Berlin: Springer-Verlag.

Parkerson, G.R., Broadhead, W.E. and Chiu-Kit, J. (1991) The Duke Health Profile: a 17-item measure of health and dysfunction. *Medical Care* 28; 1056-1069.

Parkerson, G.R., Gehlbach, S.H., Wagner, E.H., James, S.A. and Clapp, N.E. (1981) The Duke-UNC Health Profile: an adult health status instrument for primary care. *Medical Care*, 19; 806-28.

Patrick, D. and Peach, H. (1989) *Disablement in the Community*. Oxford: Oxford University Press.

Peto, V., Jenkinson, C., Fitzpatrick, R, Greenhall, R. (1995) The development of a short measure of functioning and well-being for patients with Parkinson's disease. *Quality of Life Research*, 4; 241-48.

Rosser, R.M. (1988) A health index and output measure, in Stewart, S.R. and Rosser, R.M. (eds) *Quality of Life: Assessment and Application*. Lancaster: MTP.

Rosser, R.M. and Sintonen, H. (1993) The EuroQol quality of life project, in Stewart, S.R. and Rosser, R.M. (eds) *Quality of Life Assessment: Key Issues in the 1990's*. London: Kluwer.

Ruta, D. and Garratt, A. (1994) Health status to quality of life measurement in Jenkinson, C. (ed) *Measuring Health and Medical Outcomes.* London: UCL Press.

Ruta, D., Garratt, A., Abdalla, M., Buckingham, K. and Russell, I. (1993) The SF-36 health survey questionnaire: a valid measure of health status. *British Medical Journal*; 307: 448-9.

Ruta, D., Garratt, A., Leng, M., Russell, I. and Macdonald, L. (1994) A new approach to the measurement of quality of life: the Patient Generated Index (PGI). *Medical Care*, 32: 1109-23.

Spilker B. (Ed) (1996) *Quality of Life and Pharmacoeconomics in Clinical Trials*. New York: Lippincott-Raven.

Szabo, S., on behalf of the World Health Organisation Quality of Life (WHOQOL) Group. (1996) The World Health Organisation Quality of Life (WHOQOL) assessment instrument in Spilker, B. (ed) *Quality of life and Pharmacoeconomics in Clinical Trials, Second Edition.* Philadelphia: Lippincott-Raven.

Thurstone, L. (1928) Attitudes can be measured. *American Journal of Sociology*, 33: 529-54.

Tugwell, C., Bombardier, C., Buchanan, W. Goldsmith, C., Grace, E., Bennett, K., Williams, J., Egger, M., Alarcon, G.S., Guttadauria, M., Yarboro, C., Polisson, R.P., Szydlo, L., Luggen, M.E., Billingsley, L.M., Ward, J.R. and Marks, C. (1990) Methotrexate in rheumatoid arthritis: impact on quality of life assessed by traditional standard item and individualized patient preference health status questionnaires. *Archives of Internal Medicine*, 150: 59-62.

van Agt, H.M.E., Essink-Bot, M., Krabbe, P.F.M. and Bonsel, G.J. (1994) Test-retest reliability of health state valuations collecting using the EuroQol questionnaire. *Social Science and Medicine*, 39: 1537-44.

Wasson, J., Keller, A., Rubenstein, L., Hays, R., Nelson, E. and Johnson D. and the Dartmouth Primary Care COOP Project. (1992) Benefits and Obstacles of Health Status Assessment in Ambulatory Settings: The Clinician's Point of View. *Medical Care,* 30 (Supplement): MS42-MS49.

Ware, J.E. (1992) Measures for a new era of health assessment, in Stewart, A.L. and Ware, J.E. (eds) *Measuring Functioning and Well-Being*. London: Duke University Press.

Ware J. (1993) Measuring patients' views: the optimum outcome measure. SF 36: a valid, reliable assessment of health from the patient's point of view. *British Medical Journal,* 306: 1429-1430.

Ware, J.E., Brook, R.H., Stewart, A.L. and Davies-Avery, A. (1980) *Conceptualisation and Measurement of Health for Adults in the Health Insurance Study: Volume I, Model of Health and Methodology*. Santa Monica, California: The RAND Corporation.

Ware, J.E., Kosinski, M. and Keller, S.D. (1994) *SF-36 Physical and Mental Health Summary Scales: A User's Manual.* Boston, Massachusetts: The Health Institute, New England Medical Center.

Ware, J.E., Kosinski, M. and Keller, S.D. (1995a) *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales, Second Edition.* Boston, Massachusetts: The Health Institute, New England Medical Center.

Ware, J.E., Kosinski, M. and Keller, S.D. (1995b) A 12-item short-form health survey. Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34: 220-33.

Ware, J.E., Kosinski, M. and Keller, S.D. (1996) SF-12: an even shorter health survey. *Medical Outcomes Trust Bulletin*, 4: 2.

Ware, J.E. and Sherbourne C.D. (1992) The MOS 36-Item Short-Form Health Survey 1: conceptual framework and item selection. *Medical Care*, 30: 473-83.

Ware, J.E., Snow, K.K., Kosinski, M. and Gandek, B. (1993) *The SF36 Health Survey Manual and Interpretation Guide*. Boston, Massachusetts: The Health Institute, New England Medical Center.

Wasson, J., Keller, A., Rubenstein, L., Hays, R., Nelson, E., Johnson, D. and the Dartmouth Primary Care COOP Project. (1992) Benefits and obstacles of health status assessment in ambulatory settings: the clinician's point of view. *Medical Care*, 30: (Supplement) MS42-MS49.

Wilkin, D., Hallam, L. and Doggett, M. (1992) *Measures of Need and Outcome for Primary Health Care*. Oxford: Oxford University Press.

Wilkin, D., Hallam, L. and Doggett, M. (1993) *Measures of Need and Outcome for Primary Health Care, revised edition*. Oxford: Oxford University Press.

Sheldon, T. (1993) Reliability of the SF-36 remains uncertain. *British Medical Journal*, 307: 125-6.

Silver, G.A. Paul Anthony Lembcke, MD, MPH: A pioneer in medical care evaluation. *American Journal of Psychiatry*, 80: 342-48.

Winslow, R. (1992) Questionnaire probes patients quality of life. *The Wall Street Journal*, July 7, B1 & B4.

Ziebland, S., Fitzpatrick, R. and Jenkinson C. (1992) Assessing Short Term Outcome. *Quality in Health Care*, 1: 141-142.

Ziebland, S., Fitzpatrick, R. and Jenkinson C. (1993) Tacit models of disability in health assessment questionnaires. *Social Science and Medicine*, 37; 69-75.