



UNIVERSITY
OF YORK

**CENTRE FOR HEALTH ECONOMICS
HEALTH ECONOMICS CONSORTIUM**

The Design and Construction of Quality of Life Measures

by

PAUL KIND

DISCUSSION PAPER 43

The Design and Construction of Quality of Life Measures

by

Paul Kind

The Author

Paul Kind is a Research Fellow in the Centre for Health Economics at the University of York.

Acknowledgements

The author would like to acknowledge Sal Cuthbert and Vanessa Windass for typing of various drafts of this paper.

Further Copies

Further copies of this document are available (at price £3.00 to cover the costs of publication, postage and packing) from:

The Secretary
Centre for Health Economics
University of York
Heslington
York, YO1 5DD.

Please make cheques payable to the University of York. Details of other Discussion Papers can be obtained from the same address, or telephone York (0904) 430000, ext. 5751/2.

The Centre for Health Economics is a Designated Research Centre of the Economic and Social Research Council and the Department of Health and Social Security.

Contents

	<u>Page</u>
Abstract	
Strategies for Measuring QoL	3
Measurement Techniques	7
QoL Measures	12
Specific measures	12
Generic measures	18
Summary	30
Bibliography	33

Abstract

There is growing acceptance of the need for improved outcome measures in monitoring and evaluating the impact of health care. Past emphasis on indicators of mortality and morbidity is being replaced with a greater awareness of the possibilities for measuring quality of life (QoL). This discussion paper describes some of the methodological issues which lie behind the design and construction of such measures and reviews some of the principal examples. The derivation and application of disease-specific and generic measures is documented, and the bibliography provides ready access to the main sources of reference material in this field.

The Design and Construction of Quality of Life Measures

The present high level of public interest in the difficulties which beset the NHS has exposed many of the underlying problems which face clinicians, health service managers and Government alike. As a nation we can only provide limited resources for our health service. Given this limitation we require appropriate information that enables decision-makers to distribute and commit those resources to maximum effect. Unfortunately little or no information which could aid this process is produced routinely since much of the data which is collected is geared to performance and throughput, rather than outcome and output.

The benefits of health care have, in the past, been predominantly expressed in terms of reductions in mortality and/or morbidity. Changing patterns of illness, the advent of new medical technologies and the increased range of treatment options, have eroded the usefulness of these traditional indicators. There is evident need for new forms of measure, capable of reflecting different health experiences in the population, a need acknowledged in the Black Report (paragraph 1.14, p. 10). Such measures, it was considered, could be used to shape decisions in the "allocation of resources for health and organisation of the health care system". Whether applied at health authority level or in respect of the individual patient, some degree of monitoring and evaluation of health care is essential. This should take account of the full range of costs and benefits - those of the patient and their family, as well as those of providing health care. Since treatment may impact upon both the quality and quantity of a patient's life it is clearly necessary to collect data relating to both these elements. Health care interventions are often seen as beneficial and therefore somehow desirable in themselves. The problem confronting those who make decisions about priorities in the health service

is one of quantifying the differential benefit to be derived from alternative forms of treatment or health care programmes. Measuring the quality of life contributes naturally to this process.

Despite increased awareness of the need to measure QoL¹ - in evaluating alternative modes of treatment and assessing the full range of benefits - no single generally accepted method of achieving such measurement has so far emerged. The need for QoL measures is often an incidental requirement in research projects which focus on other matters. The additional research resources required to develop new measures can be prohibitive and it is easy to see why researchers (clinicians and economists) have often elected to use or adapt existing instruments,² rather than dilute their limited research effort by investing in the development of new measures. Those who do embark upon research in this field face a unique cluster of methodological problems.

This discussion paper has been written to provide an account of those methodological issues which influence the design, construction and use of QoL measures. The first section describes some of the general approaches adopted in the measurement of QoL. The concept of "measurement" is important in its own right and great care has to be exercised in assessing the measurement characteristics of any QoL instrument, hence the second section deals with the scaling techniques that have been used in constructing weighted QoL measures. Finally, the third section describes

1 Quality of life.

2 Throughout this paper reference is made to instruments, as a general term which covers the means by which QoL measurements are made - irrespective of the format, design, mode of administration or type of rating procedure used within them.

the principal examples of QoL measures, in terms of their derivation and usage. The bibliography which is appended contains many of the more commonly cited references and should enable the reader to gain ready access to most of the source material which they would wish to consult before embarking upon his/her own QoL measurement enterprise.

Strategies for Measuring QoL

Approaches to the problem of measuring QoL fall into 2 general categories. Firstly, the relative paucity of instruments specifically designed for measuring quality of life has forced researchers to adopt a system of shadow or proxy measurement. Rather than attempting to measure QoL directly, an indicator which seems to be closely linked to QoL is selected, and observations made on this proxy indicator are used as QoL measures. Since quality of life is seldom defined, the legitimacy of such proxy measurement is sometimes hard to see. Variables such as employment status, days lost through sickness, frequency of hospitalisation and counts on symptom checklists, have been used in this way (see Figure 1). Indicators of this type have the undoubted advantage of being relatively simple to observe and data collection consequently poses no special problems. However, the linkage between these variables and QoL is often a matter of conjecture. Whilst the researcher or clinician may intuitively feel that there is a positive correlation between them, there is precious little hard evidence. Neither is there evidence about how they are perceived by providers or consumers of health care. Improvement in a single proxy index is often treated as a positive indication of enhanced quality of life. This might indeed be true for some indicators. Relief of symptoms, say, with consequent increased mobility and overall functioning

Figure 1 : Indicators used as Single Proxy Measures of Quality of Life

Employment status	Leisure activity
Days of disability	Dependence on others
Anxiety	Dependence on drugs
Depression	Fatigue
Mood	Hospitalisation
Sexual function	Problems with sleep
Relief of symptoms	Intellectual function

may improve quality of life, but it is also conceivable that no impact is made on quality of life at all.

For some patient groups, such as those with coronary conditions, there can be a fundamental shift in their attitudes as a direct consequence of their illness. Patients who seem to have regained full physical capacity may no longer regard a return to work as a desirable option. Changes in the type of employment that come about as a result of illness will also be missed. Employment status will equally be inappropriate as a QoL measure for groups of patients who face poor prospects of ever being able to return to work at all - either because of local unemployment levels, or where chronic sickness makes work impossible.

In an attempt to compensate for the inherent weaknesses in the single proxy indicator approach some researchers employ a battery of indicators which reflect many different areas of activity and functioning. These batteries often include well-established instruments drawn from psychiatry and psychology (for example, the General Health Questionnaire, Wechsler Memory Scale, Life Events Scale). This has created a more broadly-based approach to measuring quality of life in which there is scope for observing a wide range of QoL components. Where this technique is used, however, it may be difficult to set improvements measured on one indicator against deterioration as measured on another, and to produce an assessment of overall benefit. It would be difficult to evaluate a treatment, for example, which improved the physical condition of a patient but at the same time had a deleterious effect on their psychological state. Measurements which have been made using a collection of different instruments may be reduced to a single index score provided that some form of weighting system can be evolved which permits results on one dimension to be traded-off against those on other dimensions. The aggregation of results may be based

on the individual value judgements of clinicians or researchers, or more formally by seeking the preferences of patients and others connected with their health care.

The second approach to the measurement of quality of life utilises one of the limited number of instruments which have been developed for measuring health status or quality of life. These formal measures are characterised by a more structured concept of QoL, and an explicit recognition of many of the methodological problems connected with the task of constructing QoL instruments. Whether intended for use within a particular disease setting, or as generic measures, almost all formal instruments are composed of 2 complementary elements - a system of describing health states, and a means of weighting those states. These twin systems of description and valuation characterise most of the current generation of instruments. The process of constructing the descriptive system, and subsequently deriving valuations for it, raises a series of important questions - who or what is the source, how is the information collected, how is it processed and represented? The descriptive content in some cases originates with the researcher, being specified as part of a preconceived model. Alternatively, a number of different reference groups might be canvassed - patients with first-hand experience of illness, doctors and other professionals with 'objective' insights, Government Ministers with ultimate responsibility for health care provision. Using multiple sources of information forces the researcher to introduce value judgements of their own, even at the relatively simple level of describing health states. The refinement of the raw data requires a structure, however unobtrusive. If data from different sources yield significantly different descriptive classifications, then how are these to be reconciled - indeed is it meaningful to extend data refinement and reduction beyond

this point? These questions are by no means restricted to the process of describing health states. Valuations or weights have been generated using a variety of procedures. At the present time no single procedure or reference source has proved to have a substantial advantage over alternatives. It would be surprising, if administratively rather convenient, to find that a single set of valuations for health states could represent the preferences of all sections of society. Ultimately, sensitivity analysis might reveal that although different valuations existed, the results of applying any single set did not materially affect conclusions. Time will tell.

Measurement Techniques

In the physical sciences the term "measurement" conveys an impression of precision, of generally accepted procedures carried out under controlled laboratory conditions, and of results expressed in standard units of measure. This scenario contrast markedly with the efforts of social scientists concerned with the measurement of quality of life where not only is the phenomenon under investigation defined in many different ways, but there are varying opinions as to how it might be represented or quantified - even whether the measurement of quality of life is morally justified at all.

Underlying all forms of measurement is the basic operation of description. In order to be able to count objects with shared characteristics their identity has first to be described in some way. The set of patients treated in hospital can only be established once the qualifications of membership of that set have been described; thereafter the size of that population can be represented as a quantitative measure.

Measurement has been variously defined as a means of assigning numbers to objects according to some rule (Stevens, 1946). The assignment rules seem very precisely defined in the physical sciences where "measurement" has a connotation which implies exactitude, reproducibility and known degrees of error. For the social scientist there are problems surrounding the semantics used to describe objects or events, leave alone in attempting to place numbers on them. The problems which emerge in measuring QoL are then two-fold. Firstly, the descriptive foundation of QoL measurement is not universally agreed, and consequently there is scope for dispute about just what a given instrument is tapping. Secondly, there is no single definitive method of weighting an instrument so as to enhance its measurement potential. Different procedures appear, not unexpectedly, to produce different results. Much basic work still remain to be carried out to understand the relationship between the many valuations methods, and to seek the means of reconciling them.

The measurement characteristics of QoL instruments are directly influenced by its design, but are rarely given close scrutiny. In fact they should be fundamental in considering their projected use. A blood pressure or temperature reading is expressed terms of numbers which have true arithmetic information. A patient's weight of say 180 kg can properly be said to be twice that of another who weighs 90 kg. Where numbers are used as labels then this type of relationship cannot and should not be implied. The severity of a patient's condition might be expressed in terms of a "none/mild/moderate/severe" classification, which in turn could be given labels "1/2/3/4". In this case "severe" is not twice "mild", indeed the numbers could equally well be replaced by letters, or by other convenient symbols. No arithmetic advantage is gained by using numeric labels save for specific circumstances where prudent statistical analysis is intended.

A large number of different scaling procedures are available for researchers who wish to construct a weighting/valuation system. These have been most authoritatively described by Torgerson (1958), but are also detailed in many psychological textbooks (e.g. Dawes, 1972; Edwards, 1957).

All the the methods outlined below have been used in deriving valuations for health states. Magnitude estimation, in its basic form, is designed to elicit valuations directly from respondents. A reference health state or condition is selected, either by the researcher or by the respondents themselves. Other states are then judged in terms of, say, their perceived severity/seriousness compared with this standard. Repeated testing of a single rater, or more likely single testing of many raters, produces data which are usually represented by the median or geometric mean, since it is not uncommon to find that valuations elicited using this method are widely dispersed. Magnitude estimation procedures yield ratio judgements which directly represent subjects' preferences, and do not rely upon complex post-processing of data to generate scale values. The procedures have been extensively used in psychophysical experimentation, as well as in scaling of moral judgements, for example, the seriousness of crime (Sellin and Wolfgang, 1964). The techniques have been most thoroughly reviewed by Stevens (1966).

The so-called 'standard gamble' described by von Neumann and Morgenstern (1947) is a second means of deriving a quantitative expression of individual preferences. Raters are given two alternative choices, each with related benefits but with different probability of a successful outcome. In one alternative they might be offered the prospect of remaining in a specific health state, in the second alternative they are offered a risky choice in which there is a possibility of return to good

health and a complementary probability of immediate death. These probabilities are varied until the rater is unable to choose between the gamble and the certainty of remaining in the initial health state. Time trade-off (TTO) is a derivative of the standard gamble technique, developed by Torrance and his co-workers specifically for use in constructing health state utilities (Torrance, 1976). The subject in this variant is similarly presented with two alternatives, say

either being in an intermediate health state (i) for a time (t), followed by death

or being healthy for a shorter period of time (x), followed by death.

The duration of x is varied until the subject is indifferent between the two alternatives. The value of the individual subject's preferences for health state i, is given by the ratio of x/t.

The use of graphical, or visual analogue scales has become widespread in a deliberate move to avoid some of the more sweeping assumptions which lie behind techniques that require responses to be made in numeric terms. Quality of life can be recorded using a 10 cm line with defined end-points, for example 'best imaginable state' and 'worst imaginable state'. A level recorded between these points is then given a numeric value by measuring its distance along the analogue scale. Data generated using graphical rating scales has been interpreted as continuous, numeric data but Torgerson lists this technique as a form of categorical scaling.

Category rating methods force respondents to classify objects in terms of a limited number of ordered and consecutive intervals. The typical rating scale will usually contain a description of some or all of the

categories. These may consist of short phrases or numbers. In its most commonly found form, "scores" are assigned to each category according to the rank of the selected interval. Mean category scores can be calculated by aggregating data across subjects or by having a single subject make repeated judgements. Other variants of the basic category rating procedure include rank ordering in which the number of categories is effectively equivalent to the number of objects being reviewed.

Paired comparisons methods require subjects to make judgements about pairs of states, essentially answering the question "is state A worse than state B?" No estimate is made of the magnitude of the relationship. Judgements about all pairs of states are required for the original model and this typically necessitates $n \times (n-1)/2$ judgements, although modification to the procedure can circumvent this limitation where large numbers of states are involved. The analysis of paired comparisons data usually precludes the possibility of examining responses from individual subjects but measures of internal consistency are easily calculated and can be used to assess the quality of the subjects' performance and the extent of any agreement amongst them.

Scaling procedures can be divided broadly-speaking, into two groups. There are those procedures which generate information that can be interpreted as a direct quantitative expression of an individual's preferences (for example magnitude estimation). The second group of procedures give rise to data that are essentially qualitative. Preferences are expressed in terms of ordinal responses that have subsequently to be processed in one of many representational models, (for example, Thurstone's categorical judgement model). The efficiency with which any scaling procedure is able to capture and represent personal preferences for health

states is largely unknown, since no standard values have been, or are likely to be, promulgated, and comparisons of alternative procedures have raised as many questions as they have answered. What is clear however, is that different scaling methods produce different results (Kaplan et al 1979, Llewellyn-Thomas, 1982, Kind & Rosser, 1988).

Some knowledge of the techniques used to construct scale values for formal quality of life measures is essential since the means by which these values have been obtained is an important consideration in the selection of an appropriate evaluative instrument. If the scale values do not have a sound methodological basis then scores based on them may have little or no quantitative status.

QoL Measures

Specific measures

Various classification of diseases or illness groups are used routinely (e.g. New York Heart Association Classification; Harris, 1970). They perform a useful function for the doctor seeking to monitor/stage his patient through the progression of their illness. By and large, however, such classifications are made up of an ordinal series of levels usually described in terms of the signs and symptoms relevant to the condition. Disease-specific measures of health status and quality of life have been developed for some specialties (most notably in rheumatology, oncology and gerontology) and these enable more detailed information to be recorded. The examples reviewed here are merely a sample.

The Arthritis Impact Measuring Scale (AIMS) was developed as an instrument capable of assessing the physical, emotional and social well-being of patients with various rheumatic conditions. The instrument

consists of 45 questions drawn from 9 component dimensions - Mobility, Physical Activity, Dexterity, Household Activity, Activities of Daily Living, Anxiety, Depression, Social Activity and Pain - which are rated by the patient using a self-administered questionnaire. The first 6 of these scales were taken in modified form from instruments developed by other research groups (Bush, San Diego; Ware, RAND). Sexual activity was dropped as an item on the Social Activity scale because of very low response rates encountered in pilot studies.

The Functional Living Index-Cancer (FLIC) was developed to meet the need for an evaluative instrument which was capable of detecting changes in patients across a range of dimensions, not just physical well-being (Schipper, 1984), and it was designed specifically for use with cancer patients. The questionnaire was designed by an 11-member panel which included patients, spouses of patients, physicians, nurses and a clergyman. An unspecified number of patient interviews were also conducted to establish important aspects of daily functioning - as seen by patients themselves. The panel considered the review and interview material before producing their initial 92-item version of the Index. Items which were too specific, or unclear, were eliminated following field-testing. Subsequent validation exercises were conducted and with further refinement lead to a final 22-item questionnaire.

The FLIC's scoring system combines graphical and categorical scaling procedures. A scoring line, as used in visual analogue scales, is divided into 7 categories and the patient records their response by marking the point along the line which s/he feels best corresponds to their current state. Responses are then scored by taking the nearest category boundary to the visual analogue score. Each question is scored differently, and the

overall value of the Index is produced by aggregating scores for each question. The least dysfunctional state on the FLIC is scored zero, unlike many other instruments which assign this score to death. The scoring system is based on ordinal categories, and the ratings for questions appear to be aggregated across distinct dimensions. Scores on the FLIC Index, however, are reported to be significantly correlated with scores recorded on the Karnofsky Performance Status Index as well as with measures of psychosocial function.

The QL-Index (Spitzer, 1981) was specifically designed for use by physicians in measuring the quality of life of cancer patients. The content of the Index was derived following consultations with specially convened advisory panels that included patients and their relatives, doctors, nurses and healthy individuals. The items in the final form of the Index were selected from 'plausibly distinct groupings' resulting from statistical analysis of the panels' responses. A scoring system of 0, 1 or 2, adopted from the Apgar Index (1953), was arbitrarily designated for each category (Figure 2). It has some potential for use with other groups of severely ill patients, but has demonstrable limitations in discriminating adequately among relatively well individuals. Although this measure is described as an index, it is in fact a profile consisting of a series of categorical scores on separate dimensions. No means of weighting the individual dimensions has so far emerged to rectify this defect. The strength of the QL-Index lies in the way in which the descriptive content was generated. Its usefulness is limited however, by the weakness of its rudimentary scoring system.

The quality of life of patients with breast cancer has been examined using an instrument derived from the Sickness Impact Profile (Selby, 1984). The 12 dimensions in the SIP in some instances cover more than one domain.

Figure 2 : The Spitzer QL-Index

A score of 2, 1 or 0 is given according to the physician's assessment of the patient during the past week.

Activity

- 2 has been working or studying full time, or nearly so, in usual occupation; or managing own household; or participating in unpaid or voluntary activities, whether retired or not
- 1 has been working or studying in usual occupation or managing own household or participating in unpaid or voluntary activities; but requiring major assistance or a significant reduction in hours worked or a sheltered situation or was on sick leave
- 0 has not been working or studying in any capacity and not managed own household.

Daily living

- 2 has been self-reliant in eating, washing, toileting and dressing; using public transport or driving own car
- 1 has been requiring assistance (another person or special equipment) for daily activities and transport but performing light tasks
- 0 has not been managing personal care nor light tasks and/or not leaving own home or institution at all

Health

- 2 has been appearing to feel well or reporting feeling 'great' most of the time
- 1 has been lacking energy or not feeling entirely 'up to par' more than just occasionally
- 0 has been feeling very ill or 'lousy', seeming weak and washed out most of the time

Support

- 2 the patient has been having good relationships with others and receiving strong support from at least one family member and/or friend
- 1 support received or perceived has been limited from family and friends and/or by the patient's condition
- 0 support from family and friends occurred infrequently or only when absolutely necessary or patient was unconscious

Outlook

- 2 has usually been appearing calm and positive in outlook, accepting and in control of personal circumstances
- 1 has sometimes been troubled because not fully in control of personal circumstances or has been having periods of obvious anxiety or depression
- 0 has been seriously confused or very frightened or consistently anxious and depressed or unconscious.

These have been disaggregated and together with a number of disease-specific dimensions form a set of 31 (see Figure 3). Each dimension is represented by a visual analogue scale, bounded by simple descriptive phrases. The respondent is asked to mark the position on the line which best indicates their current state. The distance along each scale is measured to give a score for each dimension. Comparison of item scores with those obtained using the full SIP indicated strong positive correlation, as might have been anticipated. Physicians' assessments of the patient on the same scales correlated highly with the ratings made by the patients themselves.

The Barthel Index (Mahoney and Barthel, 1965) was developed as a simple measure of independence in chronically ill patients. The Index covers many activities of daily living and is made up of 10 categories.

Feeding	Walking on level surfaces
Moving from wheelchair	Getting up/down stairs
Personal toilet	Dressing
Getting on/off toilet	Control of bowels
Bathing self	Control of bladder

Each category has a "score" of between 0 and 15. The wholly independent patient accrues a maximum of 100 points but such a patient 'may not be able to cook, keep house, and meet the public'. Despite the arbitrary scoring system and the unsophisticated descriptive content of the categories, this Index continued to be used, for example in a study of patients with acute stroke (Wade 1983).

Figure 3 : Dimensions in the Selby QoL / Cancer profile

General Health Items

Work	Social life
Increased eating	Housework
Writing	Reduced eating
Anger	Physical activity
Reduced sleep	Family relations
Concentration	Anxiety
Self-care	Mobility
Depression	Speech
Increased sleep	Recreation

Disease-related items

Dysuria	Sore mouth
Attractiveness	Breathing
Pain	Fatigue
Information	Diarrhoea
Constipation	Nausea
Hair loss	Vomiting
Appearance	

Although specific indicators are, by definition, designed for use in a given disease setting, they are often applied in a wider context. Perhaps the most commonly encountered instrument of this type is the Karnofsky Performance Status Index (Figure 4). First used in 1948 as part of an investigation of the palliative treatment of patients with respiratory cancer, it consists of an 11-point scale with arbitrary scores ranging from 0 (dead) to 10 (normal). It is worth noting that Karnofsky also recorded observations on subjective benefits to patients in his trials unlike many of those who have subsequently used his Index. Despite problems with reliability (Hutchinson, 1979), it has been estimated that the overwhelming majority of studies of cancer which seek to measure quality of life use the Karnofsky Index. Its status as a measure of quality of life rests upon its longevity and its widespread usage, rather than on any demonstrable methodological soundness.

Generic Measures

A relatively large number of disease-specific instruments are available for use in classifying or describing levels of severity, or dysfunction. Most of these instruments are essentially ordinal measures capable of distinguishing between states but with no information about their relative values. In contrast to this essentially descriptive approach, the group of generic instruments which are outlined next have been developed, in the most part, with the intention of producing a cardinal measure capable of generating numeric information. They were designed, too, for use in the widest possible range of health settings, and are independent of specific disease connotation. The descriptive systems are based on two or more distinct dimensions/domains, which are in turn broken down into separate categories. Combinations of categories from

Figure 4 : Karnofsky Performance Status Index

Definition	%	Criteria
Able to carry on normal activity and to work No special care needed	100	Normal; no complaints; no evidence of disease
	90	Able to carry on normal activity; minor signs or symptoms of disease
	80	Normal activity with effort; some signs or symptoms of disease
Unable to work. Able to live at home, care for most personal needs	70	Cares for self. Unable to carry on normal activity or to do active work. A varying amount of assistance is needed
	60	Requires occasional assistance, but is able to care for most of his needs
	50	Requires considerable assistance and frequent medical care
Unable to care for self Requires equivalent of institutional or hospital care Disease may be progressing rapidly	40	Disabled; requires special care and assistance
	30	Severely disabled; hospitalisation is indicated although death not imminent
	20	Very sick; hospitalisation necessary; active supportive treatment necessary
	10	Moribund; fatal processes progressing rapidly
	0	Dead

different dimensions can be used to define unique health states for which valuations are generated. All but one of the examples cited here can be reduced to such a single index score, where data from separate dimensions cannot be. Not surprisingly there are relatively few measures of this type, the principal examples being developed from earlier work on the measurement of health status.

The Sickness Impact Profile (Bergner et al, 1976) was designed as a behaviourally based measure of dysfunction, and consists of 136 statements covering 12 areas of activity. Statements of the type shown in Figure 5 were collected originally from patients, individuals concerned with patient care, healthy subjects and health care professionals. This initial set of statements was sorted into separate categories. The severity of each statement was rated on an 11-point scale and the ratings converted into item scores. Seven of the categories have been used to define major physical and psychosocial dimensions within the Profile. Respondents answer 'Yes' or 'No' to each statement and the corresponding item scores are used to construct a total for each of the 12 categories, an aggregate score for the 2 principal dimensions, or a global sum for the questionnaire as a whole.

The validity and reliability of the Profile have been examined in some depth (Bergner et al, 1976) and the instrument has been used in a number of settings, including a study of the effects of early cardiac rehabilitation on quality of life (Ott, 1983). A randomised control trial of patients following myocardial infarction, examined the subjective benefits derived from two alternative forms of exercise program. Patients were selected during their hospitalisation and were followed up at 3 and 6 months, post discharge. Significant differences between experimental and control groups

Figure 5 : Example Statements from the Sickness Impact Profile

I do not walk at all (Ambulation)

I am staying in bed most of the time (Mobility)

I do not have control of my bowels (Body Care)

I am sleeping or dozing most of the time (Social Interactions)

I communicate mostly by gestures (Communication)

I have attempted suicide (Emotional Behaviour)

I sometimes behave as if I were confused or disorientated in place or time (Alertness Behaviour)

I am not doing heavy work around the house (Home management)

I am going out for entertainment less (Recreation and Past-times)

I am eating special or different food (Eating)

I sleep or nap during the day (Sleep and Rest)

I am not working at all (Work)

were recorded on the psychosocial and physical dimensions, as well as on the component categories.

A study of patients with closed head injury (Klonoff, 1986) used the SIP monitor changes in QoL over a period of up to 4 years post-injury. Residual impairment on a number of dimensions was reported with the greatest dysfunction being detected for mental alertness and work.

The SIP was one of a cluster of measures used in a study of nearly 600 patients with recently diagnosed primary malignancies (Greenwald, 1987). A battery of 5 instruments was used to measure quality of life and consisted of the SIP, Profile of Mood States (POMS), McGill Pain Questionnaire (MPQ), together with a graphical rating scale of pain sensation and an 8-point category scale of satisfaction with health care. The statistical analysis in this study was almost exclusively concerned with factor analytic methods. Four principal factors were identified corresponding to SIP, POMS, MPQ and satisfaction with health care. The SIP factor alone was able to account for nearly a third of the variance in the data.

There have been efforts to translate the Profile from its American setting and if these are successful, as indicated in a recent paper (Patrick et al, 1985), then this would represent a significant development in QoL measurement. Thus far there has been only patchy evidence of the ability to translate other measures from one culture/society to another.

The origins and design of the Nottingham Health Profile (Hunt, 1980) owe much to the influence of the SIP and to earlier work which examined patients recovering from hip replacements (McDowell et al, 1978). Originally designed for use in population surveys the Profile was based on

descriptive material collected by interviewing patients with acute and chronic conditions. A large number of statements were recorded covering social, psychological and physical function. These statements were refined by a variety of means so as to produce a set of 38 statements covering 6 categories (Figure 6). Groups of some 200 subjects used paired comparisons methods to construct a set of weights for each category and these were standardised so that each category total has a score of 100. Very little of these empirical data have appeared in press and only the data for the sleep category have been more fully reported. Since the sleep data clearly violate the goodness-of-fit statistics for the Thurstone paired comparisons model (Kind, 1982), the published weights can have no legitimate basis. In the absence of any contrary evidence, the reliability of the weights in all categories remains unsubstantiated. This has not, however, prevented age/sex norms being published (Hunt et al., 1986). The NHP was designed so as to exclude the possibility of aggregating the 6 category scores and this places a limit on the Profile's usefulness as a quantitative measure.

Much of Torrance's early work on health status measurement involved the use of scenarios or vignettes. More recently, however, (Torrance, 1982) published a classification system based on 4 dimensions - physical function, role function, socio-emotional function and health problems. The system (Figure 7) defines a large number of health states ($6 \times 5 \times 4 \times 8 = 960$) and this presents special problems in scaling such a large volume of information. Torgeson (1958) has put forward several strategies for resolving such difficulties, but in this instance Torrance adopted a procedure based on MAUT (Multi-attribute Utility Theory) (Keeny and Raiffa, 1976). Individual subjects rated levels within each dimension using category scaling methods. Time trade-off techniques were then used to establish the relationship between dimensions. These data were then combined according to MAUT rules, and a function was derived which enabled

Figure 6 : The Nottingham Health Profile

<u>Dimensions</u>	<u>No. of statements</u>
Physical Mobility	8
Pain	8
Sleep	5
Energy	3
Social Isolation	5
Emotional Reactions	9

Example statements

I find it hard to stand for long (Physical Mobility)

I'm in pain when I walk (Pain)

I'm waking up in the early hours of the morning (Sleep)

I soon run out of energy (Energy)

I feel lonely (Social Isolation)

I lose my temper easily these days (Emotional Reactions)

Figure 7 : Extract from Torrance's Health State Classification
(lower and upper categories for each dimension)

Physical function

- Level 1 Being able to get around the house without help from another person; having no limitation in physical ability to lift, walk, run, jump or bend.
- Level 6 Needing help from another person in order to get around the house; not being able to use or control arms and legs.

Role function

- Level 1 Being able to eat, dress, bathe and go to the toilet without help; having no limitations when playing, going to school, working or in other activities.
- Level 5 Needing help to eat, dress, bathe or go to the toilet; not being able to play, attend school or work.

Socio-emotional function

- Level 1 Being happy and relaxed most or all of the time and having an average number of friends and contacts with others.
- Level 4 Being anxious or depressed some or a good bit of the time, and having very few friends and little contact with others.

Health problems

- Level 1 Having no health problems.
- Level 6 Being blind or deaf or not able to speak.

values for all states to be calculated. The classification and its associated valuations formed the basis of an evaluation of neonatal intensive care (Boyle, 1983).

The Quality of Well-Being Scale (Patrick, Bush and Chen, 1973) is based upon one of the earliest and most influential contributions to health status indicators research (Fanshel and Bush, 1970). It consists of 3 ordinal scales on dimensions of daily activity - mobility, physical activity and social activity (Figure 8). Combinations of each scale define discrete function levels which are in turn linked with a classification of symptoms and problems. A typical function level consists of:

Did not drive or had help to use bus (mobility)

Walked with physical limitations (physical activity)

Performed self-care activities but not work, school or housework (social activity)

Pain in chest, stomach, side, back or hips (symptom/problem)

Weights for each function level and symptom/problem complex were computed from categorical scaling data produced by a set of predominantly healthy raters.

The QWB is a fairly lengthy, observer-completed instrument that requires specially trained interviewers, is typically obtained by taking the average score over a 4 day period. It has survived the past 13 years with some minor modifications and has recently been used in a study of chronic obstructive pulmonary disease (Kaplan, et al., 1984). The study was designed to test the effects of improving adherence to exercise programs. Patients were examined prior to being randomly assigned to behaviour modification or control groups. The QWB was assessed at this

Figure 8 : Function Levels in the Quality of Well-Being Scale

<u>Mobility</u>	<u>Physical Activity</u>	<u>Social Activity</u>
Drove car and used bus or train without help	Walked without physical problems	Did work, school or housework and other activities
Did not drive or had help to use bus or train	Walked with physical limitations	Did work, school or housework but other activities limited
In house	Moved own wheelchair without help	Limited in amount or kind of work, school or housework
In hospital	In bed or chair	Performed self-care but not work, school or housework
In special care unit		Had help with self-care

stage together with various measures of lung function and exercise tolerance. These measurements were repeated at regular periods throughout the 18 months of the study. Changes in exercise tolerance were significantly correlated with change in QWB score after 3 months. Lung function changes were also highly correlated with QWB scores across the entire study period ($r = 0.63$ $p < 0.001$). Epstein et al. (1981) examined the cost-effectiveness of biochemical testing in the detection and diagnosis of thyroid disorders, using the QWB classification to describe patients' health status. The study was based upon a review of case notes of patients with abnormal serum levels and a control group, selected at random. The "health output" of the screening test was computed on the basis of estimated likely scenarios for patients presenting with thyroid complaints.

The measures described so far have, with exception of the NHP, originated with North America research groups. The problems of translating America instruments into a UK setting have obstructed the import of most of the products of these groups, and the emergence of a British generic measure is therefore doubly significant. Rosser initially developed a set of descriptions of states of illness for use in measuring hospital output (Rosser and Watts, 1972). Two dimensions - disability and distress - were identified in discussions with doctors who had been asked to identify the salient aspects of a patient's conditions which influenced judgements about the severity of their illness. Combinations of 8 levels of disability and 4 levels of distress were used to describe 29 states of illness (Figure 9). It was thought that an unconscious patient (disability level VIII) would experience no distress, hence combinations with other distress levels (B-D) were excluded. Subsequent psychometric scaling, using magnitude estimation methods, produced a set of valuations for each of these states (Rosser and Kind, 1978). Although the sample size was small, the judges in the scaling

Figure 9 : Rosser's Descriptions of Illness States

Disability

- I No disability
- II Slight social disability
- III Severe social disability and/or slight impairment of performance at work. Able to do all housework except very heavy tasks
- IV Choice of work or performance at work very severely limited. Housewives and old people able to do light housework only but able to go out shopping
- V Unable to undertake any paid employment. Unable to continue any education. Old people confined to home except for escorted outings and short walks and unable to do shopping. Housewives able only to perform a few simple tasks
- VI Confined to chair or to wheel chair or able to move around in the home only with support from an assistant
- VII Confined to bed
- VIII Unconscious.

Distress

- A No distress
- B Mild
- C Moderate
- D Severe

exercise were drawn from groups with different current experiences of illness and included medical and psychiatric patients, as well as doctors and nurses. This work identified a number of states regarded as worse than death (Kind and Rosser, 1979). Results of using the Rosser disability/distress scale have not been widely published although it has been incorporated in a number of patient studies in psychotherapy, chronic obstructive airways disease and end-stage renal failure. A study in 2 Teaching Hospitals examined the reliability of nurses' judgements in using the classification (Benson, 1978). The scale has been used alongside an established measure of neurological state in patients with traumatic head injury, and as a comparative instrument with the Nottingham Health Profile in a study of patients with intracranial disease (Kind, 1986). It is currently being used in a prospective study of patients with posterior fossa lesions. Perhaps the major role for this measure has been in the computation of quality-adjusted life years (QALYs), initially described in an evaluation of coronary artery by-pass surgery (Williams, 1985) and recently reviewed in a CHE Discussion Paper (Gudex and Kind, 1988).

Summary

Measures of health-related quality of life are capable of describing and valuing health outcomes, either in the context of specific disease settings or more generally. Both forms of measurement require a means of distinguishing different levels or health states and this may be achieved by defining such states directly, or by combining elements from 2 or more attributes/dimensions. The descriptive content of a QoL measure may originate with a number of different reference groups, which do not always include patients. The process of setting up the descriptions is not value-free.

The measurement potential of this descriptive system can be enhanced by weighting states, so that the magnitude of any changes, as well as the direction, can be expressed. The choice of method used to establish the weights can play a significant part in determining the range and dispersion of scores. Different scaling methods give rise to scales with different properties. Whilst the descriptive systems in individual QoL measures show some degree of similarity, there is a corresponding lack of consensus regarding the valuation of dimensions/items. Since no absolute set of values exists as a comparative standard it seems probable that the question of scaling procedure will remain unresolved unless further basic research is carried out.

To the uninitiated, the selection and application of a QoL measure can appear to entail much uncertainty. The state of the art is admittedly crude, but there is accumulating evidence of congruent validity with other measures. The use of a single QoL measure in isolation should therefore be avoided. If a given disease or population group is being investigated then it would be logical to select at least one measure which was selectively targetted and to use this in tandem with one or other of the generic measures that can be reduced to a single index.

For those who are tempted to view the methodological problems inherent in designing new measures as insurmountable, or who are overfaced by the sheer range of options in selecting an appropriate QoL measure, then the QoL research group at the Centre for Health Economics is always ready to offer advice.

The development and use of QoL measures is central to a proper evaluation of the impact of health care. It can generate data capable of

informing decision-makers at all levels. Collaborative studies by clinicians and health economists will add to an emerging consensus regarding its usefulness and general applicability. QoL measurement is here to stay!

Measuring Quality of Life - a Bibliography

General

- Apgar, V. (1953) A proposal for a new method of evaluation of the newborn infant. **Current Researches in Anaesthesia and Analgesia**, July-August, 260-267.
- Bates, T.D. (1985) A Clinician's View on Palliative Care, Terminal Care and Quality of Life. **Effective Health Care**, 2(5), 211-215.
- Bergner, M. (1985) Measurement of Health Status. **Medical Care**, 23(5), 696-704.
- Boyle, M.H. et al (1983) Economic evaluation of neonatal intensive care of very-low-birth-weight infants, **New England Journal of Medicine**, 308, 1330-1337.
- Campbell, A., Converse, P.E. and Rodgers, W.L. (1976) **Quality of American Life: Perceptions, Evaluations and Satisfaction**. Russell Sage Foundation, New York.
- Cohen, C. (1982) On the Quality of Life: Some philosophical reflections. **Circulation** 65 (supplement 3), 29-33.
- Chen, M.M., Bush, J.W. and Patrick, D.L. (1975) Social Indicators for Health Planning and Policy Analysis. **Policy Sciences**, 6, 71.
- van Dam, F.S. et al (1981) Quality of Life: Some theoretical issues. **Journal of Clinical Pharmacology**, 21:suppl 8-9, 1662-168s.
- Diamond, R. (1985) Drugs and the Quality of Life: The patient's point of view. **Journal of Clinical Psychiatry**, 46:5(2), 29-35.
- Dubos, R. (1976) The State of Health and Quality of Life. **Western Journal of Medicine**, 125, 8.
- Flanagan, J.C. (1982) Measurement of Quality of Life: Current state of the art. **Archives of Physical Medical Rehabilitation**, 63(2), 56-59.
- Fries, J.F. (1983), Towards an understanding of patient outcome measurements. **Arthritis and Rheumatism**, 26, 697-704.
- Gill, W.M. (1984) Measuring a Patient's Quality of Life. **New Zealand Medical Journal**, 97, 329-330.
- Goldsmith, S.B. (1972) The Status of Health Status Indicators. **Health Services Research** 87, 212-220.
- Grogono, A.W. and Woodgate, D.J. (1971) Index for Measuring Health. **Lancet**, 1024-1026.
- Hornquist, J.O. (1982) The Concept of Quality of Life. **Scandinavian Journal of Social Medicine**, 10, 57-61.
- Hurst, J.W. (1984) Measuring the Benefits and Costs of Medical Care: The contribution of health status measurement. **Health Trends**, 16, 16-19.

- Jones, M.B. (1977) Health Status Indexes: The trade-off between quantity and quality of life. **Socio-Economic Planning Sciences**, 11, 301-305.
- Kagan, A.R. and Kagan, J.D. (1983) The Quality of Which Life? **American Journal of Clinical Oncology**, 6(1), 117-118.
- Klonoff, P.S., Snow, W.G. and Costa, L.D. (1986) Quality of Life in Patients 2 to 4 Years After Closed Head Injury. **Neurosurgery**, 19(5), 735-743.
- Krauth, J. (1982) Objective Measurement of Quality of Life. **Experientia**, 41: suppl, 402-451.
- Krupinski, J. (1980) Health and the Quality of Life. **Social Science and Medicine**, 14a, 203-211.
- Liu, B.C. (1985) Quality of Life: Concept, measure and results. **American Journal of Economics and Sociology**, 34, 1-13.
- Malm, U., May, P. and Dencker, S.V. (1981) Evaluation of the Quality of Life of the Schizophrenic Outpatient: A checklist. **Schizophrenic Bulletin**, 7, 477-485.
- McNeil, B.J., Weichselbam, R. and Pauker, S.G. (1981) Speech and Survival: Trade-offs between quality and quantity of life in laryngeal cancer. **New England Journal of Medicine**, 305(17), 982-987.
- Martini, C.J. and McDowell, I.W. (1976) Health Status: Patients and physician judgements. **Health Services Research**, 11, 508-515.
- Moriyama, I.M. (1968) Problems in the Measurement of Health Status, in **Indicators of Social Changes: Concepts and Measurements**, Sheldon, E.B. and Moore, W.E. (eds.), Russell Sage Foundation, New York.
- Nelson, E., Conger, B. et al (1983) Functional Health Status Levels of Primary Care Patients. **Journal of the American Medical Association**, 249-3331-3338.
- Najman, J.M. and Levine, S. (1981) Evaluating the Impact of Medical Care and Technologies on the Quality of Life. **Social Science and Medicine**, 15F, 107-115.
- Nikias, M.K., Sollecito, W.A. and Fink, R. (1979) An Oral Health Index Based on Ranking of Oral Status Profiles by Panels of Dental Professionals. **Journal of Public Health Dentistry**, 39(1), 16-26.
- Parkerson, G.R., Gehlbach, S.Y. et al (1981) The Duke-UNC Health Profile: an adult health status instrument for primary care. **Medical Care**, 19, 806.
- Sackett, D.L., Chambers, L.W., MacPherson, A.S. et al (1977) The Development and Application of Indexes of Health: I. General methods and a summary of results. **American Journal of Public Health**, 67, 423-428.
- Schipper, H. (1983) Why Measure Quality of Life? **Canadian Medical Association Journal**, 128, 1367-1370.
- Sellin, T. and Wolfgang, M.E. (1964) **The Measurement of Delinquency**. Wiley, New York.

Sengupta, R.P. et al (1975) Quality of Survival Following Direct Surgery for Anterior Communicating Artery Aneurysms. **Journal of Neurosurgery**, 43, 58-64.

Starzl, T.E., Koep, L.J. et al (1979) The Quality of Life After Liver Transplantation. **Transplant Proceedings**, 11, 252-256.

Starfield, B. (1974) Measurement of Outcome. **Millbank Memorial Fund Quarterly**, Winter, 39-50.

Sullivan, D.f. (1966) Conceptual Problems in Developing an Index of Health. **National Centre for Health Statistics**, Series 2, 17.

Ware, J.E., Brook, R.H. et al (1981) Choosing Measures of Health Status for Individuals in General Populations. **American Journal of Public Health** 71, 620-625.

Weinstein, M.C. (1980) Estrogen Use in Post-Menopausal Women - Costs, Risks and Benefits. **New England Journal of Medicine**, 303, 308.

Weinstein, M.C. and Stason, W.B. (1977) Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. **New England Medical Journal**, 96, 716.

Weissert, W.G., Wan, T.T. et al (1980) Cost-Effectiveness of Homemaker Services for the Chronically Ill. **Inquiry**, 17, 230-243.

Wilcock, G.K. (1978) Benefits of Total Hip Replacement to Older Patients and the Community. **British Medical Journal**, July 1, 37.

Willems, J.S., Sanders, C.R. et al (1980) Cost-Effectiveness of Vaccination against Pneumococcal Pneumonia. **New England Journal of Medicine**, 303, 553.

Measurement/Scaling

Dawes, R.M. (1972), **Fundamentals of Attitude Measurement**. Wiley, New York.

Edwards, A.L. (1957). **Techniques of attitude scale construction**. **Appleton Century Crofts**.

Kaplan, R.M., Bush, J.W. and Berry, C.C. (1979) Health Status Index: Category ratings versus magnitude estimation for measuring levels of well-being. **Medical Care**, 18, 501.

Keeney, R.L. and Raiffa, N. (1976). **Decisions with multiple objectives : preferences and value trade-offs**. Wiley, New York.

Kind, P. (1982) A Comparison of Two Models for Scaling Health Indicators. **International Journal of Epidemiology**, 3, 271-275.

Llwyn-Thomas, H., Sutherland, H.J. et al (1982) The Measurement of Patients' Values in Medicine. **Medical Decision Making**, 2, 449-462.

Llwyn-Thomas, H., Sutherland, H.J. et al (1984) Describing Health States. **Medical Care**, 22, 543.

von Neumann, J. and Morgenstern, O. (1953) **Theory of Games and Economic Behaviour**, Wiley, New York.

Sintonen, H. (1981) An Approach to Measuring and Valuing Health States. **Social Science in Medicine**, 15c, 55-65.

Stevens, S.S. (1946). On the theory of scales and measurement. **Science**, 103, 667-680.

Stevens, S.S. (1966). A metric for the social consensus. **Science**, February 4th, 151, 530-541.

Torgerson, W.S. (1958) **Theory and Methods of Scaling**. John Wiley, New York.

Torrance, G.W. (1976) Social Preferences for Health States: An empirical evaluation of three measurement techniques. **Socio-Economic Planning Sciences**, 10, 129.

Veit, C.T., Rose, B.J. and Ware, J.E. (1982) Effects of Physical and Mental Health on Health-State Preferences. **Medical Care**, 20, 386.

Quality of Well-being

Blischke, W.R., Bush, J.W. and Kaplan, R.M. (1975) A Successive Intervals Analysis of Social Preference Measures for a Health Status Index **Health Services Research**, 10, 181.

Bush, J.W., Fanshel, S. and Chen, M.M. (1972) Analysis of a Tuberculin Testing Program using a Health Status Index. **Journal of Socio-Economic Planning Science**, 6, 49.

Bush, J.W., Chen, M.M. and Patrick, D.L. (1973) Cost-Effectiveness Using a Health Status Index: Analysis of the New York State PKU screening program using a health status index, in Berg, R. (ed.) **Health Status Indexes**, 172-208, Hospital Research and Educational Trust.

Bush, J.W., Anderson, J.P., Kaplan, R.M. et al (1982) "Counterintuitive" Preferences in Health-related Quality of life. **Medical Care**, 20, 516-525.

Epstein, K.A. et al (1981). The abnormal screening serum thyroxine (T4) : Analysis of physician response, outcome, cost and health effectiveness. **Journal of Chronic Diseases**, 34, 175-190.

Fanshel, S. and Bush, J.W. (1970) A Health Status Index and its Application to Health Services Outcomes. **Operations Research**, 18, 1021-1066.

Kaplan, R.M., Bush, J.W. and Berry, C.C. (1976) Health Status: Types of validity for an index of well-being. **Health Services Research**, 11, 478-507.

Kaplan, R.M., Bush, J.W. and Berry, C.C. (1978) The Reliability, Stability and Generalizability of a Health Status Index. American Statistical Association, Proceedings of the Social Statistics Section, 704-709.

Kaplan, R.M., Bush, J.W. (1982) Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis. **Health Psychology**, 1, 61.

Kaplan, R.M. , Atkins, C.J. and Timms, R. (1984). Validity of a quality of well-being scale as an outcome measure in COPD. **Journal of Chronic Disease**, 37(2), 85-95.

Patrick, D.L., Bush, J.W. and Chen, M.M. (1973) Methods for Measuring Levels of Well-Being for a Health Status Index. **Health Services Research**, 8, 228.

Patrick, D.L., Bush, J.W. and Chen, M.M. (1973) Towards an Operational Definition of Health. **Journal of Health and Social Behaviour**, 14, 6-23.

Patrick, D.L., Bush, J.W. and Chen, M.M. (1973) Methods of Measuring Levels of Well-Being for a Health Status Index. **Health Services Research**, 8, 228.

Sickness Impact Profile

Bergner, M., Bobbitt, R.A., Pollard, W.E. et al (1976) The Sickness Impact Profile: validation of a health status measure. **Medical Care**, 14, 157.

Bergner, M., Bobbitt, R.A. et al (1976) The Sickness Impact Profile: Conceptual formulation and methodology for the development of a health status measure. **International Journal of Health Services**, 6(2), 393-415.

Bergner, M., Bobbitt, R.A. et al (1981) The Sickness Impact Profile: Development and final revision of a health status measure. **Medical Care**, 19(8), 787-805.

Bergner, L., Bergner, M. et al (1984) Health Status of Survivors of Out-of-Hospital Cardiac Arrest Six Months Later. **American Journal of Public Health**, 74, 508.

Carter, W.B., Bobbitt, R.A., Bergner, M. et al (1976) Issues and Methods in Validation of an Interval Scaling: the Sickness Impact Profile. **Health Services Research**, 11(4), 516-528.

Follick, M.J., Smith, T.W. and Ahern, D.K. (1985) The Sickness Impact Profile: A global measure of disability in chronic low back pain. **Pain**, 21, 67-76.

Gilson, B.S., Gilson, J.S., Bergne, M. et al (1975) The Sickness Impact Profile: Development of an outcome measure of health care. **American Journal of Public Health** , 65, 1304-1310.

Greenwald, H.P. (1987). The specificity of quality of life measures among the seriously ill. **Medical Care**, 25(7), 642-651.

Johnson, J.E., King, K.B. and Murray, R.A. (1983) Measuring the Impact of Sickness on Usual Functions of Radiation Therapy Patients. **Oncology Nursing Forum**, 10, 36-39.

McCusker, J. and Stoddard, A. (1984) Use of a Surrogate for the Sickness Impact Profile. **Medical Care** 22(9), 789-795.

Ott, C.R., Sivarjan, E.S. et al (1983) A Controlled Randomised Study of Early Cardiac Rehabilitation: The Sickness Impact Profile as an assessment tool. **Heart Lung**, 12, 162-170.

Patrick, D.L. et al (1985). A cross-cultural comparison of health status values. **American Journal of Public Health**, 75(12) 1402-1407.

Pollard, W.E., Bobbitt, R.A. Bergner, M. et al (1976) The Sickness Impact Profile: Reliability of a health status measure. **Medical Care**, 14, 146.

Sugarbaker, P.H., Barofsky, I. et al (1982) Quality of Life Assessment of Patients in Extremity Sarcoma Clinical Trials. **Surgery**, 91, 17.

Nottingham Health Profile

Hunt, S.M. (1981) The Development of Quality of Life Profiles: The Nottingham Health Profile in Berfenstam, R. and Jonsson, E. (eds.). **Measurement of the Quality of Life**, SPRI, Uppsala, Sweden.

Hunt, S.M. and McEwen, J. (1980) The Development of a Subjective Health Indicator. **Sociology of Health and Illness**, 2, 231.

Hunt, S.M., McKenna, S.P. et al (1980) A Questionnaire Approach to Perceived Health Status: A validation study. **Journal of Epidemiology and Community Health**, 34, 281-286.

Hunt, S.M., McEwen, J. and McKenna, S.P. (1986). **Measuring Health Status**. Croom Helm, London.

McDowell, I.W., Martini, C.J.M. and Waugh, W. (1978) A Method for Self-Assessment of Disability Before and After Hip Replacement Operations. **British Medical Journal**, September 23, 2, 857-859.

Rosser Index

Benson, T.J.R. (1978) Classification of Disability and Distress by Ward Nurses: a reliability study. **International Journal of Epidemiology**, 7, 359-361.

Kind, P., Rosser, R.M. and Williams, A. (1982) Valuation of Quality of Life: Some psychometric evidence, in **The Value of Life and Safety**, Jones-Lee, M.W. (ed.), North-Holland, Geneva.

Kind, P. and Rosser, R.N. (1979). Death and dying: scaling of death for healthy status indices. In Barber, B. et al (eds.), **Lecture Notes on Medical Informatics**, Berlin, Springer-Verlag 5, 28-36.

Kind, P. (1986). Measuring health status in the York CT Scanner Project, Centre for Health Economics, (Mimeo).

Kind, P. and Rosser, R.M. (1988). The quantification of health. **European Journal of Social Psychology**, 18, 63-77.

Rosser, R.M. and Watts, V. (1972) The Measurement of Hospital Output. **International Journal of Epidemiology**, 1, 361-368.

Rosser, R.M. and Kind, P. (1978) A Scale of Valuations of States of Illness: Is there a social consensus? **International Journal of Epidemiology**, 7, 347.

Torrance

Chambers, L.W. et al (1982) The McMaster Health Index Questionnaire as a Measure of Quality of Life for Patients with Rheumatoid Disease. **Journal of Rheumatology**, 9(5), 780-784.

Sackett, D.L. and Torrance, G.W. (1978) The Utility of Different Health States as Perceived by the General Public. **Journal of Chronic Disease**, 31(11), 697-704.

Torrance, G.W. (1976) Towards a Utility Theory Foundation for Health Status Index Models. **Health Services Research**, 11(4), 349-369.

Torrance, G.W., Thomas, W.H. and Sackett, D.L. (1972) A Utility Maximisation Model for Evaluation of Health Care Programs. **Health Services Research**, 7(2), 118-133.

Karnofsky

Evans, C. and McCarthy, M. (1985) Prognostic Uncertainty in Terminal Care: Can the Karnofsky Index help? **Lancet**, May 25th, 1204-1206.

Hutchinson, T.A., Boyd, N.F. and Feinstein, A.R. (1979) Scientific Problems in Clinical Scales as Demonstrated in the Karnofsky Index of Performance Status. **Journal of Chronic Diseases**, 32, 661-666.

Karnofsky, D.A., Abelmann, W.H. et al (1948) The Use of Nitrogen Mustards in the Palliative Treatment of Carcinoma. **Cancer**, November, 634-656.

Milstein, J.M., Cohen, M.E. and Sinks, L.F. (1985) The Influence and Reliability of Neurologic Assessment and Karnofsky Performance Score on Prognosis. **Cancer**, 56, 1834-1836.

Yates, J.W. et al (1980) Evaluation of Patients with Advanced Cancer Using the Karnofsky Performance Status. **Cancer**, 45, 2220-2224.

Arthritis

Balaban, D.J., Sagi, P.C. et al. (1986). Weights for scoring the quality of well-being (QWB) instrument among rheumatoid arthritics. **Medical Care**.

Bombardier, C., Ware, J. et al (1986). Auranofin therapy and quality of life:- patients with rheumatoid arthritis. **American Journal of Medicine**, 81, 565-578.

Deyo, R.A. et al (1982). Physical and psychosocial function in reumatoid arthritis. **Archives of Internal Medicine** 142, 879-882

Deyo, R.A. and Inui, T.S. (1984) Towards Clinical Applications of Health Status Measure: Sensitivity of scales to clinically important changes. **Health Services Research**. 198

Deyo, R.A., Inui, T.S. et al (1983) Measuring Functional Outcomes in Chronic Disease: A comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. **Medical Care**, 21(2), 180-192.

Fillenbaum, G.G. and Smyer, M.A. (1981) The Development, Validity and Reliability of the OARS Multidimensional Functional Assessment Questionnaire. *Journal of Gerontology*, 36, 428-434.

Fries, J.F., Spitz, P., Kraines, G. and Holman, H.R. (1980), Measuring 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(5), 789-793.

Helewa, A., Goldsmith, C.H. and Smythe, H.A. (1982) Independent Measurement of Functional Capacity in Rheumatoid Arthritis. *Journal of Rheumatology*, 9(5), 794-797.

Jette, A.M. (1980) Functional Status Index: Reliability of a chronic disease evaluation instrument. *Archives of Physical and Medical Rehabilitation*, 61, 395-401.

Liang, M.H. and Jette, A.M. (1981). Measuring functional ability in chronic arthritis. *Arthritic and Rhetumatism* 24, 80-86.

Liang, M.H. et al (1982) In Search of a More Perfect Mousetrap. *Journal of Rheumatology*, 9(5), 775-779.

Meenan, R.F. (1982) The AIMS Approach to Health Status Measurement: Conceptual background and measurement properties. *Journal of Rheumatology*, 9(5), 785-788.

Meenan, R.F., Gertman, P.M. and Mason, J.H. (1982) The Arthritis Impact Measurement Scales: Further investigation of a health status measure. *Arthritis and Rheumatology*, 25, 1048-1053.

Cancer

Bell, D.R., Tannock, I.F. and Boyd, N.F. (1985) Quality of Life Measurement in Breast Cancer Patients. *British Journal of Cancer*, 51(4), 577-580.

Brinkley, D. (1985) Quality of Life in Cancer Trials. *British Medical Journal*, September 14th, 291:6497, 685-686.

Burge, P.S. et al (1975) Quality and Quantity of Survival in Acute Myeloid Leukaemia. *Lancet*, ii, 612-614.

Carlens, E., Dahlstrom, G. and Nou, E. (1970) Comparative Measurements of Quality of Survival of Lung Cancer Patients After Diagnosis. *Scandanavian Journal of Respiratory Disease*, 51, 268-275.

Danoff, B., Kramer, S. et al (1983) Assessment of the Quality of Life in Long-Term Survivors After Definitive Radiotherapy. *American Journal of Oncology*, 6, 339-345.

de Haes, J.C. and von Knippenberg, F.C. (1982) The Quality of Life of Cancer Patients: A review of the literature. *Social Science and Medicine*, 20(8), 809-817.

de Haes, J.C. and Welvaart, K. (1985) Quality of Life after Breast Cancer Surgery. *Journal of Surgical Oncology*, 29(2), 123-125.

Edelstyn, G.A., MacRae, K.D. and MacDonald, F.M. (1979) Improvements of Life Quality in Cancer Patients Undergoing Chemotherapy. **Clinical Oncology**, 5, 43-49.

Fayres, P.M. and Jones, S.d. (1983) Measuring and Analysing Quality of Life in Cancer Clinical Trials: a review. **Statistics in Medicine**, Vol. 2, 429-446, Bristol, John Wiley and Sons.

Furberg, C.D., Mattson, M.E. (1984) Assessment of Quality of Life in Clinical Trials in International Symposium on Gastric Cancer: **Methodology of clinical studies and therapeutic approaches**, pp. 17-23.

Geddes, D.M. (1985) Quality of Life. **Clinics in Oncology**, 4(1), 161-176.

Gough, I.R., Furnival, C.M. et al (1983) Assessment of the Quality of Life of Patients with Advanced Cancer. **European Journal of Clinical Oncology**, 19, 1161-1165.

Irwin, P.H., Gottlieb, A. et al (1982) Quality of Life After Radiation therapy: A study of 309 cancer survivors. **Social Indicators Research**, 10, 187-210.

Linssen, A.C. et al (1982) The Development of a Well-Being Questionnaire at the Netherlands Cancer Institute. **Proceedings of the 3rd EORTC Workshop on Quality of Life**, Paris.

McArdle, C.S., Calman, K.C. et al (1981) The Social, Emotional and Financial Implications of Adjuvant Chemotherapy in Breast Cancer. **British Journal of Surgery**, 68, 261-264.

Meyerowitz, B.E. et al (1983) Quality of Life for Breast Cancer Patients Receiving Adjuvant Chemotherapy. **American Journal of Nursing**, 2, 232-235.

Nou, E. and Aberg, T. (1980) Quality of Survival in Patients with Surgically Treated Bronchial Carcinoma. **Thorax**, 35, 255-263.

Padilla, G., Presant, C.A. et al (1981) Assessment of Quality of Life in Cancer Patients. **Proceedings of the American Society of Clinical Oncology**, 22, 297.

Presant, C.A. (1984) Quality of Life in Cancer Patients: Who measures what? **American Journal of Clinical Oncology**, 7, 571-573.

Priestman, T.J. (1984) Quality of Life After Cytotoxic Chemotherapy: Discussion paper. **Journal of the Royal Society of Medicine**, 77(6), 492-495.

Priestman, T.J. and Baum, M. (1976) Evaluation of Quality of Life in Patients Receiving Treatment for Advanced Breast Cancer. **Lancet**, April 24th, 899-901.

Schipper, H. et al (1984) Measuring the Quality of Life of Cancer Patients: The Functional Living Index. **Journal of Clinical Oncology**, 2(5), 472-483.

Selby, P.J., Chapman, J. et al (1984) The Development of a Method of Assessing the Quality of Life for Cancer Patients. **British Journal of Cancer**, 50(1), 13-22.

Spitzer, W.O., Dobson, A.J. et al (1981) Measuring the Quality of Life of Cancer Patients. **Journal of Chronic Diseases**, 34, 585-597.

Spunberg, J.J., Chang, C.H. et al (1981) Quality of Long-Term Survival Following Irradiation for Intracranial Tumours in Children under the Age of Two. **International Journal of Radiation Oncology and Biological Physics**, 7, 727-736.

Williams, N.S. and Johnston, D. (1983) The Quality of Life after Rectal Excision for Low Rectal Cancer. **British Journal of Surgery**, 70(8), 460-462.

Young, K.J. and Longman, A.J. (1983) Quality of Life and Persons with Melanoma: A pilot study. **Cancer Nursing** 6(2), 219-225.

Cardiovascular

Bulpitt, C.J. (1982) Quality of Life in Hypertensive Patients, in Amery, A. (ed), **Hypertensive Cardiovascular Disease: Pathophysiology and Treatment**, pp. 931-948, Martinus Nijhoff Publishers, The Hague.

Croog, S.W., Levine, S. et al (1986) The effects of antihypertensive therapy on the quality of life. **New England Journal of Medicine**, 314(26) 157-1664.

Gensini, G.G. (1983) A More Meaningful Scoring System for Determining the Severity of Coronary Heart Disease. **American Journal of Cardiology**, 51, 606.

Harris, R. (1970) Assessing the Cardiac Patient for Rehabilitation. **New York State Journal of Medicine**, 70, 511-515.

Jackhuck, S.J. et al (1982) The Effects of Hypertensive Drugs on Quality of Life. **Journal of the Royal College of General Practitioners**, 32, 103-105.

Jenkins, C.D., Stanton, B.A. et al (1983) Coronary Artery Bypass Surgery. Physical, Psychological, Social and Economic Outcomes. **Journal of the American Medical Association**, 250, 782-788.

La Mendola, W.F. and Pellegrini, R.V. (1979) Quality of Life and Coronary Artery Bypass Surgery. **Social Science and Medicine**, 13A, 457-461.

Ross, J.K. et al (1978) Wessex Cardiac surgery Follow-up Survey: The quality of life after operation. **Thorax**, 33, 3.

Smith, H.C., Frye, R.L. and Piehler, J.M. (1983) Does Coronary Bypass Surgery have a Favourable Influence on Quality of Life? Controversies in Coronary Artery Disease. **Cardiovascular Clinic**, 13, 253-264.

Weinstein, M.C., Pliskin, J.S. and Stason, W.b. (1977) Coronary Artery Bypass Surgery: Decisions and policy analysis, in Bunker, J.P. et al (eds.) **Cost, Risk and Benefits of Surgery**, 342-371, Oxford Press, New York.

Wenger, N.K., Mattson, M.E. et al (1984) Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies. **American Journal of Cardiology**, 54, 908-913.

Williams, A.H. (1985) Economics of Coronary Artery Bypass Grafts, **British Medical Journal**, August 3, 291, 6491, 326-329.

Renal Disease

Evans, R.W., Manninen, D.L. et al (1985) The Quality of Life of Patients with End-Stage Renal Disease. **The New England Journal of Medicine**, 312, 553-559.

Gutman, R.A., Stead, W.W. and Robinson, R.R. (1981) Physical Activity and Employment Status of Patients on Maintenance Dialysis. **New England Journal of Medicine**, 304, 309-313.

Johnson, J.P., McCauley, C.R. and Copley, J.B. (1982) The Quality of Life of Haemodialysis and Transplant Patients. **Kidney International**, 22, 286-291.

Rozenbaum, E.A., Pliskin, J.S. et al (1985) Comparative Study of Costs and Quality of Life of Chronic Ambulatory Peritoneal Dialysis and Haemodialysis Patients in Israel. **Israel Journal of Medical Sciences**, 21(4), 335-339.

Rehabilitation/elderly

Crew, N.M. (1980) Quality of Life: The ultimate goal in rehabilitation. **Minnesota Medicine**, 63, 586-589.

Goldberg, R.T., Bernard, M. and Granger, C.V. (1980) Vocational Status: Prediction by the Barthel Index and PULSES Profile. **Archives of Physical and Medical Rehabilitation**, 61, 580-583.

Gresham, G.E., Phillips, T.F. and Labi, M.L. (1980) ADL Status in Stroke: Relative merits of three standard indexes. **Archives of Physical and Medical Rehabilitation**, 61, 355-358.

Katz, S., Ford, A.B. et al (1963) Studies of Illness in the Aged: The index of ADL: a standardised measure of biological and psychological function. **Journal of the American Medical Association**, 185, 914.

Kerner, J.F. and Alexander, J. (1980) Activities of Daily Living: Reliability and validation of gross vs specific ratings. **Archives of Physical and Medical Rehabilitation**, 61, 161-166.

Kottke, F.J. (1982) Philosophic Considerations of Quality of Life for the Disabled. **Archives of Physical and Medical Rehabilitation**, 63, 60-62.

Lawrence, L. and Christie, D. (1979) Quality of Life After Stroke: A three-year follow-up. **Age and Ageing**, 8, 167-172.

Mahoney, F.I. and Barthel, D.W. (1965). Functional evaluation: The Barthel Index. **Rehabilitation**, 22/23, 61-65.

Wade, D.T., Skillbeck, C.E. and Hewer, R.L. (1983) Predicting Barthel ADL Score at 6 Months After an Acute Stroke. **Archive of Physical and Medical Rehabilitation**, 64, 24-28.

Respiratory Disease

Daughton, D.M., Fix, J. et al (1982) Maximum Oxygen Consumption and the ADAPT Quality of Life Scale. **Archives of Physical and Medical Rehabilitation**, 63(12), 620-622.

Heaton, R.K., Grant, I.G. et al (1983) Psychologic Effects of Continuous and Nocturnal Oxygen Therapy in Hypozemic Chronic Obstructive Pulmonary Disease. **Archives of Internal Medicine**, 143, 1941-1947.

McSweeney, A.J., Grant, I. et al (1982) Life Quality of Patients with Chronic Obstructive Pulmonary Disease. **Archives of Internal Medicine**, 142, 473-478.

Prigatano, G.P., Wright, E.C. and Levin, D. (1984) Quality of Life and its Predictors in Patients with Mild Hypoxemia and Chronic Obstructive Pulmonary Disease. **Archives of Internal Medicine**, 144, 1613-1619.

Toevs, C.D., Kaplan, R.M. and Atkins, C.J. (1984) The Costs and Effects of Behavioral Programs in Chronic Obstructive Pulmonary Disease. **Medical Care**, 22(12), 1088-1100.