

# **The Measurement and Valuation of Quality of Life in Economic Evaluation**

## **An Introduction and Overview of Issues and Options**

**Professor Jeff Richardson**

Director, Health Economics Unit,  
Centre for Health Program Evaluation

**Professor Jan Abel Olsen**

Visiting Professor, University of Tromsø, Norway

**Dr. Graeme Hawthorne**

Deputy Director, Program Evaluation Unit,  
Centre for Health Program Evaluation

**Mr Duncan Mortimer**

Research Fellow, Health Economics Unit,  
Centre for Health Program Evaluation

**Mr Richard Smith**

Senior Lecturer, Health Economics Unit,  
Centre for Health Program Evaluation

September, 1998

ISSN 1325 0663

ISBN 1 876662 09 3

## CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

## PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator  
Centre for Health Program Evaluation  
PO Box 477  
West Heidelberg Vic 3081, Australia  
**Telephone** + 61 3 9496 4433/4434      **Facsimile** + 61 3 9496 4424  
**E-mail** CHPE@BusEco.monash.edu.au

## ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

## AUTHORS' ACKNOWLEDGMENTS

This paper reports findings from a literature review commissioned by the Commonwealth of Australia Department of Health and Aged Care to promote informed discussion of the measurement and valuation of utility-based quality of life. The review falls within the consultative process of revising the November 1995 edition of *The Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee: including major submissions involving economic analyses*. The views expressed in this paper are not necessarily those of the Department, the Pharmaceutical Benefits Advisory Committee or its sub-committees, and do not necessarily reflect the final outcome of related aspects of the next edition of the PBAC guidelines.

## ABSTRACT

This Working Paper is intended to provide a relatively quick overview of the issues associated with the measurement of Health Related Quality of Life (HRQoL) for economic evaluation. By contrast with other such overviews it emphasises a number of the more contentious issues where economists have not reached agreement. These include the broad measurement strategy, technical issues associated with measurement, the role of ethical values and psychometrics. On many of these issues we state our consensus view which, in some cases, differs from economic orthodoxy. These views are justified in a critical review of the literature which is reproduced as a companion paper.

The paper includes a glossary of terms and examples of the “multi-attribute utility” (MAU) instruments which are used in one of the approaches to the measurement of HRQoL.

# TABLE OF CONTENTS

<b>1</b>	<b>Background</b>	<b>1</b>
<b>2</b>	<b>Overview of the Present Role of Utility Measurement</b>	<b>2</b>
2.1	Measuring QoL and Economics	2
2.2	Evaluation, Psychometrics and Decision Theory	4
<b>3</b>	<b>Options and Dimensions of Choice</b>	<b>6</b>
3.1	Scaling Techniques	7
3.2	Other Considerations	8
3.3	QALYs, DALYs and HYEs	8
<b>4</b>	<b>Issues and their Significance</b>	<b>10</b>
4.1	The Place of QALYs in the Cost Benefit Framework	10
4.2	Health State Descriptions	11
4.3	Preference Weights	12
4.4	Extending QALYs: The Ethical Critique	13
<b>5</b>	<b>Conclusion and Determinations</b>	<b>15</b>
<b>Appendix 1:</b>	<b>Glossary of Terms</b>	<b>16</b>
<b>Appendix 2:</b>	<b>MAU Instruments</b>	<b>23</b>
A2.1	AQoL	24
A2.2	EQ-5D	27
A2.3	Summary HUI II and III	29

# The Measurement and Valuation of Quality of Life in Economic Evaluation

## An Introduction and Overview of Issues and Options

### 1 Background

The Australian Government requires that drugs being considered for inclusion in the Pharmaceutical Benefits Scheme (PBS) should undergo an economic evaluation. To facilitate this process, the Commonwealth Department of Health and Aged Care has issued guidelines for economic evaluation, and these are currently under review. As part of the review process, the Health Economics Unit was commissioned to outline and evaluate the issues associated with the measurement of Health Related Quality of Life (HRQoL) for the purposes of economic evaluation and to make recommendations with respect to the methods and instruments which are most appropriate for inclusion in the revised guidelines. More specifically, the review and recommendations were to consider the methods for quantifying the *strength of preference* – utility – for different health states and the methods which can, therefore, be used to measure Quality Adjusted Life Years (QALYs) which are the basis of Cost Utility Analysis (CUA). The present Working Paper is an adaptation of the introductory sections of this review. Other parts are reproduced in companion Working Papers.

In detailing guidelines/recommendations, our approach was to critically review the relevant literature and to synthesise the chief issues and options to select guidelines that, in the authors' judgement, are best supported by this literature. The overriding criteria for the recommendation of guidelines was that they will maximise the likelihood of producing utility scores such that cost/QALY ratios derived from them will, as far as possible, result in a correct ranking of the importance of the inclusion of drugs on the PBS. This, in turn, implies that utility scores derived from the recommended procedures provide the best possible index of the desirability of the health states before, during and after the use of the relevant drug. The criteria for 'best' in this context is that the utility scores possess the following properties:

- a valid and reliable measure of the strength of preference for a health state
- sensitivity to changes in the health state
- the existence of an interval property
- ability to verify and reproduce results and minimise possible gaming by interested parties
- comparability between scores derived in different studies.

With these properties or criteria in mind, this paper provides a brief overview of the economic evaluation of health programs and the role of utility measurement. It specifies the broad options for the measurement of utility; summarises many of the major issues and provides the authors' conclusions with respect to these issues. These, in turn, were the outcome of the authors' review and assessment of the literature. This is reproduced in two companion papers. The final recommendations reached are reproduced in a third paper.

---

## 2 Overview of the Present Role of Utility Measurement

This section attempts to place utility measurement in a broad perspective by summarising:

- the present role of utility measurement in economic evaluation
- the progenitors of present measurement techniques

The reason for such an overview is that a full appreciation of a number of the current issues in utility theory stems from different expectations of the role of utility measurement and the fact that both the expectations and the role of utility measurement are evolving.

### 2.1 Measuring QoL and Economics

Economic evaluation consists of a comparison of the costs and benefits of an enterprise, where these correspond with the inputs and consequences of the enterprise respectively. Conceptually, costs (or opportunity costs) refer to the value of the opportunities that have been lost by using resources for this enterprise rather than for an alternative enterprise. Generally, costs can be measured in dollars. This is because the value of resources used can normally be measured by their market price<sup>1</sup>. It is important to note that the measurement of utility and QoL are of relevance on the benefits side of the cost benefit equation. They should not be confused with dollar costs.

By contrast with costs, the (positive or negative) benefits or consequences of an enterprise are often difficult to monetise as, for example, in the case of environmental degradation or the loss of heritage value where there is no market price that captures societal benefits. As a consequence, as in the case of the early evaluation of health programs, benefits are often non-monetised and expressed as 'intangibles'. The different forms of economic evaluation of health programs have arisen as a direct consequence of the way in which these 'intangibles' have been treated. QoL, narrowly defined, is only one of the intangibles associated with health interventions. The other factors of social concern, as discussed in this paper and its companion papers are not quantified by the present techniques of CUA and, consequently, the outcomes of a narrowly defined CUA should not be regarded as the only relevant benefits of an economic evaluation.

Importantly, the measurement of benefits in economic evaluation involves two steps that are conceptually distinct and normally distinct in practice. The first is the measurement of the consequences of a health-related intervention as measured in natural units such as additional years of life, change in blood pressure, etc. Second, there is the determination of the *value* of these changes. Economics is concerned with the second of these steps and it is the role of epidemiologists or clinical researchers to determine outcome in natural units. This implies that economic evaluation does not compete with or intrude upon clinical or epidemiological research. Rather the two forms of evaluation are complementary.

---

<sup>1</sup> In theory the market price in a competitive market represents the opportunity cost as prices in such markets are equal or close to the marginal cost and, in such a market this will be equal to the marginal benefit. Where markets are not competitive, adjustments should be made to the observed price.

---

Conventionally, three types of economic evaluation of health programs have been distinguished and, broadly, these define the different stages of the unfinished evolution of economic evaluation. First, Cost Benefit Analysis (CBA) is defined by the monetisation of benefits, including the benefits of life and the quality of life. Various techniques have been used to monetise the value of life, the most common being the human capital and the willingness to pay approaches. It has commonly been asserted that the latter—the willingness to pay technique—represents the gold standard in measurement of health benefits as elsewhere in economics, utility is measured by this willingness to pay. Dissatisfaction with these techniques led to the development of the second type of evaluation; viz, Cost Effectiveness Analysis (CEA). In CEA, benefits are expressed in natural units of outcome and, commonly, by lives saved or the number of life years obtained. As these natural units ignore the quality of life, CEA evolved into Cost Utility Analysis (CUA) which was intellectually marketed by asserting that it weighted life years by an index of the quality of life to produce 'Quality Adjusted Life Years' (QALYs). In reality, the weights employed in CUA are, conceptually, an index of the strength of preference for a health state. As 'utility' is defined in economics as the strength of preferences CUA therefore produces 'utility adjusted life years'.

As CUA, in turn, only incorporates quality (utility), economic evaluation has recently entered a fourth stage that might be described as 'cost welfare analysis'. In this, the concept of utility is being modified to include other potentially relevant consequences. These include the distribution of QALYs per sé and the distribution between different social groups (and particularly the aged or disadvantaged), the initial severity of the health state and the effect of such 'process disutilities' as anxiety, hope, etc. Suggestions have already been made for the weighting of QALYs. This further stage is given relatively little attention in the current report as there is not yet a consensus on these issues.

As described in Richardson (1998), these forms of economic evaluation do not simply represent an increased sophistication in the application of economic theory. The unit of outcome in each case is different and there are ethical arguments for and against the selection of different units. In particular, when CBA attempts to use the willingness to pay as the basis of outcome it is, implicitly, accepting the orthodox 'welfarist' view that the appropriate unit of outcome is 'utility' and that this should be measured, as elsewhere, by the willingness to pay. By contrast, cost effectiveness analysis (CEA) represents a shift from the welfarist tradition to the 'material welfare' tradition of Marshall, Pigou and mainstream economics before the 1930's. This tradition has re-emerged as 'extra welfarism'. In this, the welfarist objective may be supplemented or replaced by some non-utility objective such as life per sé. As judged by stated policy objectives, there is little doubt that extra *welfarist values* are commonly held. Cost utility analysis (CUA) is a hybrid between these two traditions as its unit of outcome seeks to weight life years (extra welfarism) by an index of utility (welfarism). What we have called 'cost welfare analysis' is the exploration of an enlarged set of outcomes that may (or may not) more accurately represent community values.

This situation is somewhat confusing and common terminology is particularly misleading. 'Cost benefit analysis' would be more appropriately labeled 'cost utility analysis' as it seeks to measure utility (albeit, as revealed by the willingness to pay). 'Cost effectiveness' would be better described as cost benefit analysis as the unit of outcome is a non-utility benefit. CUA is, strictly, a sub-set of CEA where the outcome is (utility) weighted life years.

---

Setting aside fine academic issues of definition the important message here is that there are alternative outcomes and it is an ethical issue as to which of these should be adopted in the health sector. In our view, economists who argue that particular forms of analysis are 'theoretically correct' and that other forms of analysis are simple approximations of the 'theoretically correct' outcome as implied by 'economic theory' misunderstand the role of social value judgements and the potential importance of extra welfarism.

We do not dwell on this theoretical issue in the report. It is included to highlight the fact that it is not for the present authors to decide social values. Our recommendations are based upon what appear to be universally held values, viz the importance of patient preferences, life years and the quality of life. How the latter should be measured is not, in our view, an issue that can be determined by value free, positive economic theory. Recommendations therefore allow some flexibility with respect to non-core outcome measurement.

## **2.2 Evaluation, Psychometrics and Decision Theory**

Current evaluation techniques are not the result of the steady evolution of a single, coherent research program. Rather, they would be better described as an ad hoc amalgam of inputs from economics, psychometrics, sociology and decision analysis. Possibly due to the tradition in economics of assuming full information and 'rationality', economists have usually assumed that the literal meaning of scenarios will be fully understood. Moreover, it is usually assumed that the meaning of the scenarios will be fully absorbed into the attitudes and beliefs of respondents and that answers to scaling questions will be an accurate response to the precise question that has been asked. For this reason, a significant part of CUA has ignored the need for the validation of scenarios and utility scores. Rather, validity has been assumed or, as in the case of Gafni and Birch (1995), redefined to mean 'consistency of measurement procedures with a set of axioms' where consistency is ensured, not by validation studies, but by the use of the standard gamble. Increasingly, however, it has been recognised that the validation techniques of psychometrics and the qualitative techniques of sociology have an important role in CUA.

Quite separate from the evolution of economic evaluation, health outcomes have been measured by clinicians using one of the very large number of disease specific instruments. These have generally consisted of a set of ad hoc scales or sets of questions that require either a yes/no answer or the use of simple category rating, (eg five options from best to worst). These scales have become more sophisticated through the use of the correct psychometric techniques for instrument construction. Through time, a smaller number of generic instruments have also been created, the more recent of which have also been developed using correct psychometric techniques for instrument construction (the NHP, SIP, MOS, SF 36, SF 12, WHOQoL).

Multi attribute utility techniques evolved from this latter intellectual tradition. The first of these instruments (the QWB and Rosser Kind Index) were developed by psychologists not economists and, probably for this reason, these instruments were scaled using the two most common psychometric techniques, viz the rating scale and magnitude estimation. Also reflecting their intellectual origins, these techniques took little account of the instrument properties needed for

---

valid utility measurement as laid out by decision analytic theory, viz preference and structural independence between items or dimensions and the use of an appropriate model for combining dimensions. These requirements were first taken into account in the development of the HUI (version 1) although, as with subsequent versions of the HUI, the descriptive systems were not developed using correct theory or instrument construction. Consequently, none of the MAU instruments in common use have adopted fully satisfactory construction methods. Descriptive systems have been ad hoc and/or scaling and modeling techniques for instrument construction have been inappropriate.

In sum, our present methods have evolved from different traditions that have incorporated different objectives, different and untested assumptions and different units of outcome. Despite this, there is common agreement that the strength of preference (utility) is an important component of outcome and there has been some agreement about the techniques that should be used for its measurement. Both holistic and MAU approaches have evolved through the ad hoc adoption of methods from other disciplines.

The significance of these historical facts is that they help to explain the differences in the perspectives of many researchers. The difference is, perhaps, most starkly illustrated by the psychometricians' continuing emphasis upon empirical validation and the emphasis by some economists upon consistency with the axioms of orthodox theory and definition of objectives in terms of welfarism. The position of the present authors is that there is consensus over the importance of some but not all objectives, and that objectives are a matter for social decision making and not the product of 'economic theory'. Moreover, validation in the psychometric sense is central to good measurement and instrument construction should be based upon the correct application of psychometrics and decision analytic theory. Consistent with the consensus view, descriptive systems should be scaled using a technique that measures preferences in an appropriate way.

---

### 3 Options and Dimensions of Choice

There are three broad measurement strategies for the quantification of the utility value of HRQoL.

- (i) **The composite (holistic) approach.** With this approach, a scenario or vignette of the health state is constructed usually on the basis of information directly collected from patients in a focus group or with the use of other qualitative techniques. The vignette is then directly rated with the use of one of the techniques discussed below (the standard gamble, time trade-off, etc).
- (ii) **The Multi Attribute Utility (MAU) approach.** With this approach, an MAU ‘instrument’ is constructed. This has two parts. First, the ‘descriptive system’ consists of a set of items or questions about different dimensions of health and a variable number of response categories for each item. Second, there is a scoring system which adds utility weights to the responses and then combines these individual scores into an overall utility score on a scale on which zero and unity represent death and normal (good or best) health respectively. The entire instrument is constructed in the initial research phase. It is subsequently applied by asking patients to fill in the descriptive system. Their utility score is then derived from the pre-determined scoring system. Two of the descriptive systems (the EuroQoL, the Assessment of Quality of Life (AQoL) are reproduced in the Appendix and also a description of the Health Utilities Index (HUI) (Mark II)).
- (iii) **Direct utility elicitation** is the practice of asking patients to directly apply a utility score to their current health state using one of the scaling instruments. Despite its a priori appeal very few studies have employed this technique and, in the absence of satisfactory evidence or argument we are unable to consider this option as a contender for standard use.

Despite the prima facie difference in these three options, each consists of the same two basic components. First, a health state must be described. In the three options this is done with a vignette, with the M-A descriptive system and by informing the patient that the health state under consideration is their own. Second, a utility score is attached to the health state. With the three options this corresponds with the scaling of the scenario, the application of the pre-determined scoring system and self-rating respectively.

The choice between these options should be determined by (i) the validity of the descriptive system; (ii) the sensitivity of the descriptive system; (iii) the cognitive burden upon subjects when they assign utility scores; and (iv) the ease of instrument application. These criteria do not result in a general preference for one option over the others. MAU instruments are comparatively easy and cheap to apply. Consequently they may be used repeatedly in longitudinal studies. Their descriptive systems can, in principle, be independently validated as can, in principle, their utility scoring system. In contrast, with the holistic approach the validity of the vignette has seldom, if ever, been tested and the approach is significantly more costly to apply. However the vignette is capable of providing an extremely detailed and disease specific description which may, in

---

principle, incorporate statements of risk, process and prognosis. Our general conclusion, which mirrors current practice, is that the different methods are appropriately used in different contexts.

### 3.1 Scaling Techniques

Five major techniques have been used for 'scaling' health state descriptions, ie for the determination of an importance weight. Perhaps the largest issue in cost utility analysis has been which, if any, of these techniques produce importance weights that correspond with the economist's concept of utility. The five techniques are:

- (i) **The Rating Scale (RS).** Respondents are asked to indicate the strength of their preference for a health state on a straight line calibrated between zero (death) and unity (full health). The distance above and below the selected value represents the strength of preference relative to full health and death.
- (ii) **Magnitude Estimation (ME).** Respondents are asked to state how many times worse that one health state is than a standard scenario that takes a score of 100. A series of health states, including death, are compared all using the question 'How many times worse than the previous health state is this one?'. Subsequent rescaling of scores is claimed to result in a ratio scale.
- (iii) **The Time Trade-Off (TTO).** Respondents are asked to indicate the proportion of their remaining life (or the life expectancy nominated by the interviewer) they would be prepared to give up to be returned from a life time (or the nominated time) in the health state under review to full health. The proportion then indicates the disutility of the health state. For example, if a person was willing to give up 20 percent of their remaining life to avoid a health state, then the health state disutility would equal 0.2 and the utility would equal 0.8.
- (iv) **The Standard Gamble (SG).** Respondents are asked to select between a life time (or a stated number of years) in the health state under review or a gamble in which there is a probability,  $p$ , that they would live in full health and a probability  $(1-p)$  that they would almost immediately die. The probability,  $p$ , is varied until the individual can no longer make up his or her mind. At this point the probability  $p$  represents the importance weight or index of utility. This result occurs when the utility value of full health and death are set equal to unity and zero respectively. From this it follows that  $U_s = p.U(1) + (1 - p) U_0 = p.1 + (1-p).0 = p$ .
- (v) **Person Trade-Off (PTO).** Once known as the equivalence technique, this method asks respondents to consider two ways of spending a fixed budget. The first will save the life of  $x$  people and return them to full health. The second will save the life of a stated number (eg 100) people but leave them in the health state to be assessed.  $x$  is varied until the two programs are considered to be equally valuable. The utility of the health state is then inferred from the fact that  $x$  people receiving full health (utility gain of 1.0) is equal to 100 people receiving a utility gain of  $U_{hs}$ , where this represents the utility of the health state.

---

These scaling techniques do not produce the same results and it necessarily follows that they do not all produce true utility scores. As discussed in the next section some have argued that the standard gamble represents a gold standard as a result of the set of definitions. While we reject this approach, this leaves no clear criterion for the measurement of utility. However, amongst economists, there is now near consensus that one of the final three techniques should be employed as, unlike the first two, each of these involves explicit choice and a choice that requires a consideration of the value of life per sé and the quality of life. As the QALY represents an exchange rate between quantity and quality of life (lower QoL scores have an identical effect upon QALY scores as a lower length of life) this is considered to be an essential property of a utility scaling instrument.

### 3.2 Other Considerations

The combination of one of the techniques for *describing* health states and one of the *scaling* techniques for *quantifying preferences* will produce a weighted life year that may claim to be a QALY. These are not the only outcome from a health program, and neither are they the only consideration when comparing the 'benefits of health programs'. These other issues are discussed more fully later in the paper.

### 3.3 QALYs, DALYs and HYE

The World Bank and World Health Organisation (WHO) have supported the development of a particular QALY-like unit of outcome which has been entitled the 'Disability Adjusted Life Year' (DALY). In its more sophisticated 1996 version, DALYs are derived, as with QALYs, by the multiplication of life years by an index of utility. The index is derived from a rating scale on which the utility scores of 22 health states (diseases) are shown. These utilities were derived using the person trade-off technique. Other health states (diseases) were then scaled by asking a group of experts to locate the health state on the rating scale in relation to the 22 marker states. In principle, this procedure is simply a variant of the usual QALY procedure. The unique features of the DALY are that (i) it is the only instrument to date to employ the PTO scaling technique; (ii) the PTO values were derived after debate and discussion (by experts) of the quality of life in each of the disease states; (iii) rating was carried out by medical experts and not by the public; (iv) the results have been used to derive the 'burden of disease', ie the DALYs lost because of different diseases. As a result, statistics detailing burden of disease are now available for many diseases and for the majority of countries.

A further QALY-like unit of outcome has been entitled the Healthy Year Equivalent (HYE). With this approach to measurement the health scenario is used which both describes the health state and provides information on its duration. The entire scenario is then converted into "healthy year equivalents" in a two stage procedure. In the first, the standard gamble is used to convert the multi-year scenario into a single index of utility. In the second, the standard gamble is again used but this time to convert the index of utility (which is numerically identical to the probability value obtained in the first stage) into a number of "healthy year equivalents" in normal health. The economists who first proposed this approach (Gafni and Mehrez) argued that the HYE was

---

“theoretically superior” firstly, because it allowed for a change in the strength of a person’s preferences through time and secondly, because it employed the standard gamble as the scaling technique with measurement being undertaken “under risk”. Following a protracted debate in the literature, it is now generally believed that there is little theoretical difference between the HYE and the QALY when the latter is obtained using the time trade-off (TTO) technique over the same period of time.

The existence of the DALY and the HYE indicate a more general conclusion. This is that there are, potentially, a large number of “QALY-like” units of outcome which are associated with different scaling techniques; time period of the assessment; inclusion or exclusion of prognosis; inclusion or exclusion of multiple health states and the inclusion or exclusion of pre-outcome emotions such as anxiety, risk avoidance and anticipated regret. Attention has focused upon the DALY and the HYE for a variety of reasons which do not include a demonstration, in terms of evidence, that they are psychometrically superior instruments or a more appropriate reflection of social values than other QALY-like measures.

---

## 4 Issues and their Significance

In this section we summarise the major issues that are discussed in greater detail in the companion reviews of the QALY and MAU literatures. Additionally, we note our own view and the extent to which the issues are of significance for our recommendations.

### 4.1 The Place of QALYs in the Cost Benefit Framework

There is currently some debate about whether or not QALYs are the appropriate unit of output. Some economists argue that orthodox, welfarist theory implies the superiority of the “willingness to pay” (WTP) index of utility. Extra welfarists maintain that this unduly favours the wealthy. In our view this is an ethical issue that cannot be resolved by positive economic theory. The balance of evidence suggests that the population rejects income as the basis for access to health care and considers that health per sé, not utility, is the appropriate object of measurement. The issue is not central to a review that focuses upon methods for QALY measurement.

**(i) What is the relationship between psychometric and economic measures of utility?**

Both purport to measure utility but only the economist’s approach purports to measure this utility directly. It does so by giving significance to the importance weights derived from the scaling techniques, the use of which are the defining difference between an economic and a psychometric instrument.

**(ii) Do QALYs take into account the possibility that different people may obtain different levels of utility from a year of life and, particularly, when they have greater income to spend?**

No. QALYs treat the capacity of each individual to enjoy ‘utility’ as being the same. It is our view that the decision whether or not to treat individuals as having equal utility is an ethical or social/political judgement which cannot be answered by economic theory. The evidence available strongly suggests that, in the context of a national health scheme, there is strong support for the QALY assumption and that the very reason why national schemes have been introduced is to prevent differential treatment on the basis of income and most other characteristics.

**(iii) Can the utility of a multi year health state scenario and/or a scenario with changing health states be calculated by the addition of the utilities experienced in each year and measured without reference to the utility of other years?**

There is universal agreement that this is unlikely and that additive separability is assumed to simplify the calculation of total utility scores.

---

**(iv) Can MAU instruments that measure only context free, timeless utility scores measure the utility of a multi period scenario?**

This is the same issue as the previous one. There is universal agreement that the MAU will only approximate the result of theoretically ideal multi period measurement. Unfortunately little effort has been made to quantify the magnitude of the error introduced by the use of an MAU instrument.

**(v) What is the criterion for valid measurement?**

The first view is that 'validity' means consistency with orthodox economic theory. The second view is that 'valid' measurement occurs when we have correctly measured the concept we wished to measure. Whether or not this concept is appropriate is an ethical issue, ie it is a value judgement which economists cannot make on the basis of value free positive analysis. The authors accept the second of these interpretations.

## **4.2 Health State Descriptions**

**(i) What are the respective advantages of the holistic and MAU approaches?**

The MAU reduces the cognitive burden on respondents/patients when they complete the instrument. It may therefore be used in longitudinal studies with comparative ease. The properties of the instrument are known and replication of results is, in principle, possible. However the cognitive burden on respondents during the construction of the instrument may lead to bias; the validity of some instruments is suspect and instruments have variable sensitivity to different health states. The holistic approach, while more expensive, may include more realistic health state scenarios including factors such as anxiety, risk and changing health status.

**(ii) Can direct utility elicitation be used?**

Yes, but insufficient experience with this technique has been compiled to allow its comparison with holistic or MAU measurement.

**(iii) Do the holistic, MAU and direct elicitation methods produce the same results?**

The overall correspondence is poor in studies conducted to date. However, these studies are limited and used the early generation of MAU instruments.

**(iv) Should utility measures incorporate *ex ante* factors, viz anxiety, anticipation, dread and dislike of risk associated with an intervention?**

This is a social judgement that must be made and cannot be answered by technical economic analysis. Most studies to date do not include these factors.

- 
- (v) Does the addition of a patient's utility experienced each year give the same overall result as the simultaneous evaluation of a multi year scenario?**

The limited evidence indicated that the two procedures would give different answers.

- (vi) Can generic utility instruments be substituted for condition specific outcome measures (CSOM's)?**

Generally the CSOM will be more sensitive and the generic instrument may be unable to detect clinically relevant differences.

- (vii) Do framing and labeling effects have a quantitatively significant impact upon utility scores?**

Yes. Care should be exercised to avoid these.

- (viii) Does the labeling of endpoints on a utility scale affect utility measurement?**

Yes. For comparison, standard endpoints must be employed.

- (ix) How should states worse than death be treated?**

There is no gold standard. Altering the lower level of the possible negative utility scores may significantly effect results. Sensitivity analysis is therefore necessary.

- (x) Patients adapt to health states. Should pre or post adaptation utilities be measured?**

There is some difference of opinion on this issue. Most argue that actual utility should be measured and adaptation should lead to the measurement of higher utility scores.

- (xi) What are the advantages and disadvantages of the holistic (composite) measurement of health states as compared with the MAU (decomposed) approach that underlies an MAU descriptive system?**

These issues were discussed in the previous section. There is no consensus on this question. It is generally recognised that while the holistic approach is capable of producing a more context and/or disease specific description it may result in cognitive overload and corresponding invalidity and unreliability. While this problem is overcome by the decomposed approach, there has been little discussion and even less empirical enquiry about the accuracy of the modeling devices needed to combine the separate elements of the decomposed approach into a single score.

---

## 4.3 Preference Weights

### (i) Whose values should be incorporated in utility scores?

There is no consensus on this issue. Once again, this is an ethical issue for social resolution. There is compelling reason for using patient values because it is patients who experience the health states. For comparability between holistic and MAU instruments (which have used the values of the population) it is necessary for holistic measurement to use the values of the general public.

### (ii) Is the standard gamble the gold standard for eliciting utilities?

Most practising health economists no longer accept the special status of the standard gamble. It is one of three scaling techniques that reflect choice between the quality and quantity of life, viz the Standard Gamble (SG), the Time Trade-Off (TTO) and the Person Trade-Off (PTO).

### (iii) Is the Disability Adjusted Life Year (DALY) an acceptable measure of the quality and quantity of life?

The DALY is one of many possible QALY like instruments and in some contexts (especially the quantification of the burden of disease) it appears to have a comparative advantage over other QALY like measures. However it has not been subject to comparison and validation against alternative measures. Its use of age weights and the PTO are likely to ensure results that are not comparable with other QALYs.

### (iv) Is the Health Year Equivalent (HYE) superior to the QALY?

The HYE is another QALY like measure defined by its simultaneous measurement of a multi year scenario and the use of a two-stage standard gamble. Most theorists consider the two-stage standard gamble to be identical to the time trade-off. In practise the scores produced by the HYE are likely to differ because they have been generated by two sets of decisions both confounded by risk that is unrelated to the health state.

### (v) Which of the three scaling instruments (SG, TTO, PTO) has the best psychometric properties?

There is little evidence on the properties of the PTO although some preliminary studies suggest significant unreliability. SG and TTO appear to have better properties; the SG having somewhat greater test re-test correlation and the TTO scoring better on psychometric tests of validity (correlation with other measures; consistency with patient ranking of health states).

---

**(vi) Can the results from the application of different scaling techniques be reconciled using a functional relationship?**

A small number of such studies have been undertaken but with variable results. There is only one example where a transformation in one study proved satisfactory in a second independent study. Consequently, there are no acceptable universal transformations.

**(vii) Should the application of scaling techniques use visual aids?**

General experience suggests that these are useful.

**(viii) Do purely hypothetical stated preference techniques give a valid representation of revealed preferences?**

Comparisons of stated and revealed preferences for health outcomes are very difficult. Studies reported in the transportation and environmental literature clearly suggest that well conducted stated preference studies predict revealed preferences.

#### **4.4 Extending QALYs: The Ethical Critique**

**(i) How widespread is the acceptance of QALYs as the basis of economic evaluation?**

Some commentators totally reject the ethical basis of QALYs, viz that life years and the preference for different health states should be 'traded' off against each other. Generally commentators accept that the quality and quantity of life are both important in economic evaluations but either argue that the technical methods for eliciting QALYs should be improved or that additional factors should be taken into account in economic evaluations.

**(ii) Which other issues have been nominated for inclusion in economic evaluations?**

There is prima facie evidence of:

- the independent importance of the severity of the initial health state;
- for the need to disregard strict economic criteria in an emergency;
- for the equitable distribution of health benefits;
- for non discrimination on the basis of long term disability; and
- for the different treatment of life years for people of different ages.

---

## 5 Conclusion and Determinations

This paper has provided a brief overview of the economic evaluation of health programs and the role of utility measurement. It outlines the broad options for the measurement of utility and it summarised many of the major issues. Critical reviews of the quality adjusted life year (QALY) and multi-attribute utility (MAU) literatures, as well as recommendations/guidelines arising from the review are given in three companion CHPE working papers.

Several issues arising from the discussion in this paper had a significant influence on the final recommendations/guidelines. With respect to the present role of, and options for, utility measurement these were as follows:

- (i) Despite its a priori appeal very few studies have employed this technique and, in the absence of satisfactory evidence or argument we are unable to consider direct utility elicitation this option as a contender for standard use.
- (ii) The MAU and holistic approaches each have their particular advantages. Our general conclusion, which mirrors current practice, is that these methods are each appropriate when used in suitable contexts.
- (iii) Each of the TTO, SG and PTO involves explicit choice and a choice that requires a consideration of the value of life per sé and the quality of life. Consequently, these three techniques are deemed appropriate for 'scaling' health state descriptions.
- (iv) Economists who argue that a particular form of analysis (CUA, CBA, or CEA) is 'theoretically correct' and that other forms of analysis are simple approximations of the 'theoretically correct' outcome as implied by 'economic theory' misunderstand the role of social value judgements and the potential importance of extra welfarism.
- (v) There is consensus over the importance of some but not all objectives reflected in outcomes for economic evaluation, and that objectives are a matter for social decision making and not the product of 'economic theory'.
- (vi) Validation in the psychometric sense is central to good measurement and instrument construction should be based upon the correct application of psychometrics and decision analytic theory.

An important conclusion, which is evident from the scope and significance of the issues discussed in this paper, is that Cost Utility Analysis remains in its developmental stage. There is no consensus over a number of important issues of theory and practice. There is even debate about the use of Utility Adjusted Life Years as distinct from the willingness to pay for health programmes. While Cost Utility Analysis is now a fact of life in the economic evaluation of health programmes this final and most important conclusion indicates the need for sensitivity analysis in the conduct of CUA and caution in the interpretation of the results.

# APPENDIX 1

## Glossary of Terms

---

## A1.1 Glossary of Terms

### 15D

A 15 dimensional instrument designed and developed by Harry Sintonen in Finland. This pioneering instrument, developed in the early 1980s remains one of the most sensitive of the instruments currently available. It is scaled using a rating scale and dimensions are combined with an additive model.

### Composite (holistic) utility measurement

This term describes one of the three ways in which utility may be measured. With this approach a complete health state (consisting of multiple dimensions) is described in a vignette or brief narrative describing the health state. This is then used in a survey in which respondents are asked to evaluate and score the health state using one of the utility scaling techniques (TTO, SG, PTO). Utility may also be measured using a multi-attribute utility instrument such as the AQoL or it may be measured directly by asking patients to rate their own health state.

### DALY (Disability adjusted life year)

This is a concept introduced recently by the World Bank/WHO study team lead by Chris Murray and Alan Lopez. The DALY is a QALY like measure in which life years are adjusted by a utility index derived from the PTO. More specifically, PTO values were obtained for 22 health states during a series of focus groups carried out with health experts during which the health states were discussed. The 22 states were located on a rating scale and all other major diseases were then located on the rating scale in relationship to the 22 marker states, again by health experts. Results were used to estimate the burden of disease for all major disease groups in all countries.

### Decision analysis

This is a branch of theory that assists with decision making. In essence, complex decisions are broken down into their constituent parts (a complex health state is broken down into its multiple dimensions). The value or utility of each aspect of the decision (item or dimension of health) is independently estimated (using one of the numerous scaling techniques evolved by decision theorists). The decomposed values or utility scores are recombined using one of several possible models (additive or multiplicative being those that have been adopted by MAU instrument makers).

---

<b>Descriptive system</b>	This refers to the questionnaire-like set of statements and response categories that define the health domain of a QoL instrument
<b>Dimension</b>	A sub-set of items which define a coherent sub-set of the quality of life. For example, physical health, mental health, social health.
<b>Disease specific instruments</b>	Instruments structured to emphasise dimensions of particular relevance in assessing outcomes for a specific condition. Disease specific instruments are more likely to be sensitive to changes in health status for their target patient group than are generic instruments such as the EQ-5D or HUI.
<b>Economic costs</b>	In principle, economic (or opportunity) costs represent the benefits foregone as a result of any economic activity. The economic cost of reading this glossary is the benefit that may have been obtained from reading something else or undertaking another activity. In practice, economic costs are generally measured by the market price of a commodity where this is available or by an estimated market price equivalent when it is not. Economic costs are compared with dollar benefits in cost benefit analysis and a ranking of the ratio of cost to benefits is obtained in cost effectiveness analysis and cost utility analysis.
<b>EuroQoL or EQ-5D</b>	This MAU instrument consists of six items that define 243 response categories (Dolan, 1997). The instrument was designed for cross European comparisons of the quality of life. It was subsequently the basis for one of the largest scaling exercises to date, using the time trade-off technique. The research, carried out by Alan Williams and his colleagues at York, resulted in several sets of 'utility tariffs' each applying to a different age, sex and SES group.
<b>Factor analysis</b>	Factor analysis collapses many items or variables into a few underlying dimensions or constructs. That is, several items assessing hearing, speech, and vision might load highly on a factor that captures the underlying HRQoL dimension of 'Physical Senses'. The association between items and factors determines the semantic content of dimensions. Factor analysis allows the structure of an instrument to be verified or empirically derived.
<b>Health Utility Index (HUI)</b>	The HUI Mark I, II and III are three MAU instruments developed by George Torrance and his colleagues at McMaster University. Along with the AQoL, the HUI instruments are the only ones to employ a multiplicative model for combining utilities.

---

---

<b>HRQoL (Health related quality of life)</b>	This is a relatively loose concept with refers to the sub-set of quality of life which may be affected by a health program. As health programs may, potentially, affect all aspects of life, HRQoL could, arguably, be considered identical with the concept of quality of life.
<b>Healthy year equivalent (HYE)</b>	This is one particular form of QALY which is derived by using the standard gamble to evaluate, not one but all of the years of life that will be experienced in a given health state. The concept has been vigorously promoted by Gafni at McMaster University. There has been a major debate to determine whether or not the HYE is conceptually different from measurement over multiple years using the TTO.
<b>Instrument</b>	This term has its normal meaning <i>vis</i> a device for achieving a particular objective. In the present context the device is the descriptive system all questionnaire – type set of items with which a respondent can indicate – describe – their health state. The instrument may or may not have a corresponding set of utility weights. If it does not, it is commonly described as a 'psychometric instrument' (ie where unweighted response scores can be summed). Where multiple dimensions are described and utility weights or scores are included, it becomes 'a multi-attribute utility' instrument.
<b>Item</b>	A single statement dealing with a single element or concept. Each item typically has multiple response categories (eg: not at all, a little, a lot, a great deal).
<b>Latent variable</b>	An unobserved variable which is postulated as a construct and is defined by a sub-set of items. Each health dimension represents a latent variable as does the global concept of quality of life.
<b>Multi-attribute Utility (MAU) Instrument</b>	This consists of a set of questions and a corresponding set of scores that can be combined into a single index number. The 'instrument' is the questionnaire that asks people to indicate, for each item or health related statement in the questionnaire which response most closely corresponds with their own health. The instrument has multiple attributes if it is a generic instrument that can describe many dimensions of health. It is also a utility instrument if the importance weights attached to each response have been derived using a technique for utility measurement (standard gamble, time trade-off, person trade-off, rating scale, magnitude estimation).

---

---

<b>Orthogonality (Independence)</b>	Decision analytic theory, upon which the modelling (combination) of utility scores is based requires that the different dimensions of health should be orthogonal (uncorrelated) with each other. This independence is also known as structural independence. (See structural independence).
<b>Person Trade-off (PTO)</b>	This scaling device was once known as the equivalence technique but has been (more descriptively) redefined and popularised by Erik Nord. The technique asks respondents to consider two programs. The first returns a defined number (eg 100) patients from imminent death to the health state being evaluated). The second program returns a variable number, N, patients from imminent death to full health. N is varied until the two programs appear equally attractive. It is then possible to infer that the utility of the health state is equal to $n/100$ .
<b>Preference independence</b>	This refers to the requirement of multi-attribute utility theory that the utility value of an item should not depend upon a person's position on an unrelated item or combination of items. Preference independence would be violated if the dis-utility of anxiety were greater when anxiety occurred in combination with pain. There are three forms of preference independence (see Keeney and Raiffa, 1976; Winterfeldt and Edwards, 1993; or Feeney et al 1995).
<b>Psychometrics</b>	This is the branch of psychology that deals with the quantification of psychological quantities such as intelligence. It is derived from psychophysics that deals with the quantification of the strength of perception of physical stimuli such as heat, light and sound. The psychometric entities dealt with by psychometrics are considered to be a psychological 'construct' and, as such, are analogous to the construct we call 'quality of life'. As these constructs cannot be directly observed, various techniques have evolved for their measurement and validation. All of the original MAU instruments employed psychometric techniques.
<b>Quality adjusted life year (QALY)</b>	This is almost a self-explanatory term. QALYs are a metric obtained by multiplying the number of calendar years of life by an index number that reflects the utility of strength of preference for the health state of the person involved. Strictly speaking, the metric should be described as a 'utility adjusted life year' or as 'healthy year equivalents'. The term may either be used narrowly to refer to life years multiplied by an index number which refers to average utility or the term may be used more broadly to refer

---

to the generic set of metrics which involve the adjustment of life years. The broader concept subsumes: 'healthy years equivalents', DALYs, years adjusted by any of the utility measuring devices, utility measured before or after the health state has been experienced, utility of a defined number of years measured as a single outcome, or the utility of a health state that varies through time.

**Quality of Well Being (QWB instrument)**

Along with the Rosser-Kind index this was the earliest of the MAU instruments and was based upon an ad hoc survey of quality of life. It is based upon a rating scale and an additive model. The QWB was the basis for the utility scores used in the State of Oregon to rank all of the procedures available for the poor through the State's Medicaid program.

**Rating scale**

This is the scaling technique that has been most commonly used by psychometricians. Typically it presents the respondents with a vertical or horizontal line which is calibrated between zero and 1 or zero and 100 where the end points are clearly and unambiguously defined (usually as normal health and death). Respondents are asked to indicate how far between the end points the health state being evaluated should be. Respondents are typically urged to consider the distance on the rating scale to represent the strength of their feeling. Rating scales are seldom used by economists because of the ambiguity in the interpretation of the results.

**Standard gamble (SG)**

This is a scaling technique that is considered by some to be the gold standard because it employs the axioms of von Neuman and Morgenstern. The standard gamble consists of a choice between (typically, a life time) in the health state of interest and a gamble between life and death in which the probability of life is varied until the gamble is equally attractive as the certainty of life in the inferior health state. At this point of equivalence, the probability of the favourable outcome is taken as an index of the strength of preference (or utility).

**Scaling**

The process whereby a set of numbers representing relative importance is attached to a descriptive system. If an appropriate scaling technique is adopted (time trade-off, standard gamble) then the scale values may claim to represent utility weights.

**Structural independence**

This is a requirement of decision theory that is the basis for combining or modelling the utility scores of different health dimensions. Structural independence – the lack of correlation between dimensions – is required so that a particular aspect of health will not be, in effect, double counted. For example, if

---

strong pain was always associated with anxiety (these items were correlated in a patient survey) then the importance of anxiety could be measured both directly and indirectly through the disutility assigned to pain.

### **Time Trade-off (TTO)**

A scaling technique to derive utility weights. In the typical time trade-off, the respondent is offered a choice between the health state of interest for a defined number of years (typically the respondent's life expectancy) and a lesser period of full or normal health. The second option is varied until the respondent indicates that the two options are of equal value. At this point, the utility of the health state may be defined as the ratio of years of full health divided by the larger number of years in the health state.

### **Utility**

An index of the strength of a person's preference. Initially, and as envisaged by Jeremy Bentham, John Stuart Mill and the Victorian Utilitarians, utility referred to the psychological dimension of pleasure/pain. From the 1930s onwards this concept was largely replaced in economics by the concept of preference utility. In philosophy, the original concept similarly competes with preference based utility, although this concept is somewhat different from the economist's concept.

### **WHO QoL**

A large generic instrument designed by a cross-national team coordinated by the WHO. The full WHOQoL is much broader than a conventional MAU instrument. The WHOQoL (brief) is available for use as a psychometric (unweighted) instrument. Utility weights have not yet been derived.

---

## **APPENDIX 2**

### **MAU Instruments**

---

## A.1 Assessment of Quality of Life (AQoL) Instrument

For the questions in this section, please TICK  ONE BOX for each question to show which alternative best describes you during the last week.

- 1 Concerning my use of prescribed medicines:
  - I do not or rarely use any medicines at all
  - I use one or two medicinal drugs regularly
  - I need to use three or four medicinal drugs regularly
  - I use five or more medicinal drugs regularly.
  
- 2 To what extent do I rely on medicines or a medical aid? (NOT glasses or a hearing aid.) (*For example: walking frame, wheelchair, prosthesis etc.*)
  - I do not use any medicines and/or medical aids
  - I occasionally use medicines and/or medical aids
  - I regularly use medicines and/or medical aids
  - I have to constantly take medicines or use a medical aid.
  
- 3 Do I need regular medical treatment from a doctor or other health professional?
  - I do not need regular medical treatment
  - although I have some regular medical treatment, I am not dependent on this
  - I am dependent on having regular medical treatment
  - My life is dependent upon regular medical treatment.
  
- 4 Do I need any help looking after myself?
  - I need no help at all
  - Occasionally I need some help with personal care tasks
  - I need help with the more difficult personal care tasks
  - I need daily help with most or all personal care tasks.
  
- 5 When doing household tasks: (*For example: preparing food, gardening, using the video recorder, radio, telephone or washing the car.*)
  - I need no help at all
  - occasionally I need some help with household tasks
  - I need help with the more difficult household tasks
  - I need daily help with most or all household tasks.

---

6 Thinking about how easily I can get around my home and community:

I get around my home and community by myself without any difficulty

I find it difficult to get around my home and community by myself

I cannot get around the community by myself, but I can get around my home with some difficulty

I cannot get around either the community or my home by myself.

7 Because of my health, my relationships (*for example: with my friends, partner or parents*) generally:

Are very close and warm

Are sometimes close and warm

Are seldom close and warm

I have no close and warm relationships.

8 Thinking about my relationship with other people:

I have plenty of friends, and am never lonely

Although I have friends, I am occasionally lonely

I have some friends, but am often lonely for company

I am socially isolated and feel lonely.

9 Thinking about my health and my relationship with my family:

My role in the family is unaffected by my health

There are some parts of my family role I cannot carry out

There are many parts of my family role I cannot carry out

I cannot carry out any part of my family role.

---

10 Thinking about my vision, including when using my glasses or contact lenses if needed:

I see normally

I have some difficulty focusing on things, or I do not see them sharply. *For example: small print, a newspaper or seeing objects in the distance.*

I have a lot of difficulty seeing things. My vision is blurred. *For example: I can see just enough to get by with.*

I only see general shapes, or am blind. *For example: I need a guide to move around.*

11 Thinking about my hearing, including using my hearing aid if needed:

I hear normally

I have some difficulty hearing or I do not hear clearly. *For example: I ask people to speak up, or turn up the TV or radio volume.*

I have difficulty hearing things clearly. *For example: Often I do not understand what is said. I usually do not take part in conversations because I cannot hear what is said.*

I hear very little indeed. *For example: I cannot fully understand loud voices speaking directly to me.*

12 When I communicate with others: (*For example: by talking, listening writing or signing.*)

I have no trouble speaking to them or understanding what they are saying

I have some difficulty being understood by people who do not know me. I have no trouble understanding what others are saying to me.

I am understood only by people who know me well. I have great trouble understanding what others are saying to me.

I cannot adequately communicate with others.

13 Thinking about how I sleep:

I am able to sleep without difficulty most of the time

My sleep is interrupted some of the time, but I am usually able to go back to sleep without difficulty

My sleep is interrupted most nights, but I am usually able to go back to sleep without difficulty

I sleep in short bursts only. I am awake most of the night.

14 Thinking about how I generally feel:

I do not feel anxious, worried or depressed

I am slightly anxious, worried or depressed

I feel moderately anxious, worried or depressed

I am extremely anxious, worried or depressed.

---

15 How much pain or discomfort do I experience:

- None at all
- I have moderate pain
- I suffer from severe pain
- I suffer unbearable pain.

## A.2 EQ - 5D

*By placing a tick (thus: ) in one box in each group below, please indicate which statements best describe your own health state today.*

### 1. Mobility

- I have no problems in walking around
- I have some problems in walking around
- I am confined to bed

### 2. Personal Care

- I have no problems with personal care
- I have some problems washing or dressing myself
- I am unable to wash or dress myself

### 3. Usual Activities (e.g. work, study, housework, family or leisure activities)

- I have no problems with performing my usual activities
- I have some problems with performing my usual activities
- I am unable to perform my usual activities

### 4. Pain/Discomfort

- I have no pain or discomfort
- I have moderate pain or discomfort
- I have extreme pain or discomfort

### 5. Anxiety/Depression

- I am not anxious or depressed
- I am moderately anxious or depressed
- I am extremely anxious or depressed

---

### A.3 Summary HUI Mark II & III

The HUI II and III are under copyright. The descriptive system and scoring formula may be obtained from the instrument developers: D. Feeny, W. Furlong and G. Torrance<sup>2</sup>. The descriptive systems are also reproduced in Spilker (1996)<sup>3</sup>.

The items included in the HUI II cover the following dimensions:

**Sensation:** Vision (2 items), Hearing (2 items), Speech (2 items)

**Emotion:** Happiness - irritability, anxiety, depression (2 items)

**Pain:** Pain/discomfort - impact on function, Pain/discomfort - drug dependence (2 items)

**Mobility:** Ambulation, dexterity (2 items)

**Cognition:** Memory, problem solving (2 items)

**Activities of daily living:** Self-care (1 item)

---

<sup>2</sup> Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Bio-statistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5.

<sup>3</sup> Spilker, B., Ed. (1996). Quality of Life and Pharmacoeconomics in Clinical Trials, Lippencott Rover.