

**Quality of Life Measurement of Patients
Receiving Treatments for Gallstone Disease:
Options, Issues and Results**

Johanna Cook

Senior Research Fellow

Jeff Richardson

Professor and Co-Director

September, 1993

ISSN 1038-9547

ISBN 1 875677 29 1

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 receives core funding from the National Health and Medical Research Council and 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.

ABSTRACT

The paper discusses the methods for determining health state values in the context of a study of treatment options for gallstone disease.* The study design is a cost utility analysis (CUA) and the technique used to measure quality of life had to meet the specific requirements of a CUA, viz that the units of measurement were appropriate for the calculation of quality adjusted life years. The paper examines the selection of a measurement technique and the related issue of how the measurement task was controlled to ensure the eliciting of valid responses. It also discusses the psychometric properties of the two measurement techniques used, namely the time-trade off technique (TTO) and the rating scale (RS). It supports the findings of other studies that the psychological quantity measured by the two scaling techniques is either quite different or measured in different units. Consequently, the magnitude of the measured change in QALYs would depend upon the choice of a scaling technique. It argues for the adoption of a *de facto* gold standard for measuring health state values for CUA via a generic multi attribute utility index.

* In two companion papers Street (1993) presents the methods and results from the analysis of treatment costs, and Cook and Richardson (1993) discuss the method adopted in the construction of health state scenarios used for the quantification of treatment outcomes. An overview of the study, its methods, results and the sensitivity of these to various assumed values is presented in Cook, Richardson and Street (1993).

TABLE OF CONTENTS

- 1 Introduction
- 2 Methodological Issues
 - 2.1 Selecting the Appropriate Measurement Technique
 - 2.1.1 Utility or Value
 - 2.1.2 Time Trade Off (TTO)
 - 2.1.3 Rating Scale (RS)
 - 2.2 Eliciting Valid Responses
 - 2.2.1 Survey Sample
 - 2.2.2 Controlling the Measurement Task
 - 2.2.3 Acceptability of TTO to Respondents
 - 2.3 Measuring Temporary Health States
- 3 Comparison of Measurement Techniques
 - 3.1 Comparison of Mean RS and TTO Values
 - 3.2 Comparison of Mean Values Between Health States
 - 3.3 Distribution of Health State Values
 - 3.4 Relationship Between the Measurement Techniques
 - 3.4.1 Ranking
 - 3.4.2 An Exchange Rate Between RS and TTO Values?
 - 3.4.3 Individual Values
 - 3.4.4 Means
- 4 Summary and Discussion

Tables

- 1 Population Sample by Age and Sex
- 2 Long Life with Sickness versus Shorter Life with Full Health: Acceptability of the Right to Choose
- 3 Differences in Mean Values for Chronic Health State Three (12 Months and 12 Years: TTO and RS Values)
- 4 Comparison of TTO and RS Values for Temporary Health States
- 5 t-Values of Differences in Mean Values of Temporary Health States
- 6 Utility Values from the Seven Health States: Means and Confidence Intervals
- 7 t-Values of Differences in Mean Health State Utilities
- 8 Distribution of RS and TTO Values: Pearson's Coefficient of Skew
- 9 Rank Order of Temporary Health States
- 10 Results of Linear and Power Function Regression of RS (Dependent) and TTO (Independent) Health State Values from Individuals
- 11 Results of Linear and Torrance Power Function Regression of RS (Dependent) and TTO (Independent) Mean Health State Values
- 12 Comparison of Mean Health State Values: TTO, RS and Predicted RS

Appendices

- 1A Quality of Life Measurement for Gallstone Disease Treatment Options Interview Schedule
- 1B Instructions to Interviewer
- 2 Health State Scenarios

Quality of Life Measurement of patients Receiving Treatments for Gallstone Disease: Options, Issues and Results

1 Introduction

This paper had its origin in a clinical trial of the three treatment options for gallstone disease, viz open cholecystectomy, laparoscopic cholecystectomy and extracorporeal shockwave lithotripsy (ESWL) which was conducted at St Vincent's Hospital, Melbourne during 1989 and 1990. Because of the importance of the quality of life (QoL) during post procedural recovery a cost utility analysis was selected as the appropriate framework for an economic evaluation of the three options. Cost utility analysis measures 'quality of life' in terms of 'utility', the strength of individual preferences for health states. Respondents are asked to trade-off quality of life (morbidity) and length of life (mortality), the unit of measurement being a quality adjusted life year (QALY) or a healthy year equivalent (HYE).¹ Consequently, it was necessary to select an instrument for measuring QoL that was appropriate for the calculation of healthy year equivalents (HYEs) or quality adjusted life years (QALYs). Despite the large number of CUA studies in the literature there is still no consensus about the appropriate gold standard for measuring quality adjusted life years. The choice of a measurement instrument is discussed in *section 2.1* below. The practical issues associated with eliciting valid responses are discussed in *section 2.2* and *section 2.3* examines the use of QALYs for measuring temporary health states. The data collected in the subsequent interview via the TTO and RS allows further exploration of a number of issues associated with the choice of measurement instrument and these are discussed in *section 3*.

¹ The term HYE has been by used by Mehrez and Gafni (1989) to distinguish the conventional QALY (which assumes that an individual's discounted health states over time may be combined additively) from a unit of outcome derived from the holistic evaluation of multi stage scenarios.

2 METHODOLOGICAL ISSUES

2.1 Selecting the Appropriate Measurement Technique

The techniques most commonly used to measure health state preferences for economic evaluation are the standard gamble (SG); the time trade off technique (TTO); and the rating scale (RS). The underlying assumption of these techniques is that they generate data with an interval property, ie that a given increment in the scale (for example, 0.2) has the same meaning anywhere along the scale (ie, a movement from 0.4 to 0.6 has the same meaning as a movement 0.8 to 1.0). This is a necessary property for the calculation of aggregate QALY gains if initial health states, and subsequent QoL benefits differ between individuals.

Preference for health states is not an observable phenomena and no measurement instrument can be tested for criterion validity. In the absence of a generally acceptable gold standard convergence between the values achieved by different measurement techniques is often used as one indicator of validity. Studies which have compared various approaches to measuring health have found, in general, that the standard gamble and the time-trade-off technique give comparable results but that the RS produces results which are significantly different from the former techniques (Torrance 1976; Richardson et al. 1989; Froberg & Kane 1989).

2.1.1 Utility or Value

Some economists have argued that the standard gamble should be regarded as the gold standard for measuring preferences as it was derived from the von Neumann and Morgenstern axiomatic approach to utility theory. According to its proponents it therefore takes account of an individual's attitude towards risk.

We started with the view that the much discussed difference between utility (measured under risk) and value (measured under certainty) is not helpful in the context of health state measurement (von Winterfeldt & Edwards 1986). The standard gamble has not been successful in measuring even one of the concepts of utility described in the literature (Richardson 1990). As a pragmatic measurement instrument the standard gamble has the disadvantage that it is not easily understood and thus accuracy of measurement may be seriously impaired. From this position the two major contenders for quality of life measurement are the TTO technique and the RS.

2.1.2 Time Trade Off (TTO)

The TTO is a certainty equivalent method derived from decision theory. Respondents are asked to trade off time in two health states. They are typically asked how many years of life in a defined ill health state they are willing to give up to be in full or normal health. The method has face validity as the functional relationship between time in the health state envisaged and the number of healthy year equivalents is measured directly as the scale itself is calibrated in terms of healthy year equivalents. In this respect the TTO is unique as other methods impose an intermediate scale (probability or RS) between the individual's preferences for healthy years and the measured health state. The units (expressed as weeks, months or years) have an unambiguous interval property - 6 weeks in *health state 1* is twice the *duration* of 3 weeks in *health state 2*. All else equal, the value of *health state 1* is twice the value of *health state 2*.

A disadvantage of the TTO is that it can only be administered in face to face interviews and it needs to be administered by appropriately trained interviewers. These requirements make the TTO relatively expensive to use. Like other methods available, it depends upon the respondents' capacity to understand the task they are asked to carry out.

2.1.3 Rating Scale (RS)

The rating scale (RS) was first used for psychometric measurement. The RS is often considered to be an efficient method for eliciting health state values because it can be used as a self-administered instrument (Brooks 1991). It is most often used in health status measurement as a line with clearly defined end points where 100 measures the best (or normal health) and 0 the worst health state (or death). Other health states are measured by placing a marker on the line relative to the best and worst health state. To achieve an interval scale respondents are instructed to place the health states in such a way on the line that the distance between the health states reflects how strongly they feel about each health state relative to the one above and below it. The limitation of rating scales are well covered in the literature. (See for example Gescheider 1988.) Chief among these is the contention that the resulting scale does not have an interval property. This is disputed by Anderson (1976) whose theory of functional measurement seeks to test and validate the interval property of the achieved scale. In addition, there is some evidence that the perceived ease of administration is more illusory than real. For the measurement of health states Torrance (1976) found the category RS to be substantially more difficult to administer than the TTO. Informal observations and feedback from our pilot study supported this finding. We found that respondents to the RS needed careful instruction to understand what was asked of them. The structure of our interview (discussed in *Section 2.2.2*) reflects the need to minimise these difficulties. Respondents had to be carefully guided through each of the steps of the measurement task. This evidence conflicts with the view that rating scales are less costly to administer.

2.2 Eliciting Valid Responses

Four issues are considered below that are relevant in this study to ensure that the values obtained from respondents represent the values of the target population that are reliable and valid. These are: (1) the appropriate population from which to obtain QoL values; (2) conveying the concept of the health state to be evaluated; (3) use of the measurement instrument and, finally; (4) the use of QALYs for temporary rather than permanent health state measurement.

2.2.1 Survey Sample

Potentially three groups could be sampled to obtain QoL values, namely: (1) patients, (2) the non-patient population with age, sex and SES matched with the general population and, finally, (3) the non-patient population with age, sex and SES characteristics matched against the patient population. The use of patients is supported by the argument that they have experienced the health state and, consequently, are uniquely capable of evaluating them. The counter argument is that, to a greater or lesser extent, patients may 'adapt' to the health state they have experienced thereby altering its numerical value relative to their previous perception of the health state value vis-a-vis other health states or full health (O'Connor 1993).

The case for using a general population sample is that as health care is financed largely by public money it should also be the general public that determines health state values. This argument was

not accepted here. While it is possible to accept a value system in which 's(he) who pays the piper calls the tune' it is not a necessary consequence of public funding. Payment is separable from detailed decision making and public hospital and medical services at the micro level in a public systems are generally guided by patient/doctor preferences and not by collective choice.

The implicit gold standard adopted in the present study was that, as with economic evaluation generally, the appropriate values to incorporate are those of the potential users. Consequently, the study adopted the third option of seeking a general population sample with similar characteristics to the patient population.

To simplify the sampling procedure SES profiles based on the indicators of socio-economic disadvantage (Ross et al. 1988) were drawn up from postcodes of the St Vincent's Hospital patients. The SES profiles were divided into an upper and lower group. The median SES value in both the upper and lower group was then used to identify the areas from which to recruit members of the general public. On the basis of the anticipated standard error (established from similar research reported in the literature), a target sample of 96 was sought to obtain health state values. Interviewing was carried out, by appointment, in the respondent's home. On the basis of quality control procedures discussed in 2.2.2 below three interviews were discarded. The results from the pilot survey were also discarded. The characteristics of the final sample of 93 are presented in *Table 1*.

— TABLE 1 —
Population Sample by Age and Sex

Sex	<i>Less than 40 Years %</i>	<i>41 - 55 Years %</i>	<i>56 Years and Over %</i>	<i>All Ages %</i>
Male	25.9	44.4	29.6	29.0
Female	37.9	47.0	15.2	71.0

2.2.2 Controlling the Measurement Task

The method used to measure health status needs to be as easy and comprehensible as possible and consistently applied. The validity and reliability of survey results depends, in large part, upon the way in which the instrument is administered. To increase the likelihood of a valid response visual aids were used to assist the implementation of the instrument. Three experienced interviewers were employed and trained to use the measurement techniques and visual aids. Interviewers were required to carry out a training interview with members of the research team. Subsequently, the first interview from each interviewer was observed by members of the research team to ensure quality control. If necessary, interviewers were given further training. A particular problem in the present context is the need to illicit 'true' preference judgements about life and death issues and about health states which could have been highly distressing to respondents. The literature and our own pilot research suggested a learning curve in the application of the chosen measurement instruments techniques. To minimise bias introduced in this way respondents were required to practise with both the RS and the TTO before being asked to evaluate the health states of interest.

Health scales are often calibrated with vaguely defined end points. For example, the best and worst health states in the EuroQol instrument are defined as 'best imaginable health state, worst imaginable health state'. The alternative procedure adopted here was to define the end points more precisely. The best health state was defined for respondents as 'You have no health problems. You are able to do everything that a person of your age can normally do.' The worst state for measuring temporary health states was chosen by the respondent from the health states presented to them. Health state scenarios are described in Appendix 2. The scale used for measuring the worst chronic health state was calibrated with death as the worst state (numerical value of zero) as described in Drummond, Stoddart and Torrance (1987).

The RS was presented to respondents as a 'feeling thermometer'. They were informed that the instrument measures how people feel about different health problems and that the numbers were used to indicate how strongly they might feel about being in a particular health state. That is, if they 'felt' that health was about $\frac{3}{4}$ of the way between the best and the worst health state they should mark the thermometer at 75. Alternatively, if it felt like being half way between the best and worst health state the mark should be placed at 50.

While administering the TTO we adopted the 'flip flop' technique developed by Torrance (1986). This alternates choices between the two extremes on the measurement scale and progressively narrows the difference between the choices, until the point is reached where the longer period in the ill-health state is considered equivalent to the reduced period of full health. This approach minimises the risk of a systematic bias from the manner of presentation of the options.

To ensure that respondent thought carefully and consistently about their preferences for each of the health states they were asked to imagine themselves in the health state described to them for a specified period of time. Specifically the interviewer stated: 'I want you to imagine that you are the person described on this card and that you would have these health problems for 12 weeks. After that you would return to normal health. Think about how you would feel about these health problems'. When respondents were asked to trade off life for full health they were asked to pause and think what it would mean to them if they were really to die, ie they were urged to regard the choices as real options they were facing. Health states were presented in random order and respondents were told not to worry about preferences they already had expressed but to view each evaluation task as separate from previous ones. Respondents were allowed to rate health states as equivalent if they could not choose between them. Interviews, on average, took between 30-45 minutes to complete. The interview schedule is reproduced in Appendix 1.

2.2.3 Acceptability of TTO to Respondents

A related issue is the extent to which the measurement task is agreed to by respondents. The validity of any measurement task depends upon respondents' acceptance of the legitimacy of the scaling technique. With the TTO respondents are asked to trade off time in the health states being compared. When the TTO is used to measure chronic health states the trade off is between time in the chronic state and time in full or normal health. A fundamental issue in the context of such life and death decisions is whether or not respondents agree with the basic premise that such a trade off is acceptable. If a large number of respondents, for whatever reason, are unwilling to trade-off life for quality of life then the TTO will be insensitive to changes in the QoL. Consequently, respondents were asked to indicate how they felt about trading off length of life for improvements in the quality of life. Specifically, they were asked if they were very sick and different treatments were available, should they have the right to choose between a long life with sickness and a shorter life with full health.

Table 2 demonstrates that the overwhelming majority of respondents (96 percent) agreed that they should have the right to make this choice. Only 2 percent strongly disagreed with this premise.

— TABLE 2 —
Long Life with Sickness versus Shorter Life with Full Health:
Acceptability of the Right to Choose

	Strongly Agree %	Agree %	Neutral %	Disagree %	Strongly Disagree %	Total %	n
Self							
Male	57.7	38.5	-	-	3.8	100	26
Female	51.5	40.9	6.1	-	1.5	100	66
Total	53.3	40.2	4.3	-	2.2	100	92

2.3 Measuring Temporary Health States

A requirement of cost utility analysis is that values for temporary health states are comparable with values for permanent or chronic health states. That is, temporary health states need to be measured on a scale where full or normal health is given a value of unity and death is given a value of zero. To measure the health state values for the temporary health states associated with gallstone disease we followed Torrance's (1986) two step approach. First, respondents were asked to rate each of the temporary health states relative to full health and the worst temporary health state as chosen by the respondent. The resulting values were then weighted by the value of the worst health state when measured as a chronic condition. In principle, this second task should have treated the worst health state as a chronic condition which lasted for the same time as the temporary health state, that is the health state is of the same duration but is now followed by death. In our case this would have required the chronic, morbid condition to be followed by death after only 12 weeks (the nominated period of temporary morbidity for the interview). This approach raises some interesting questions. Evidence from the literature on the effect of duration on health state values is contradictory. Kind and Rosser (1988) in their study of scale values for disability and distress found little difference in health state values when duration was changed from temporary to chronic. Sackett and Torrance (1978), on the other hand, demonstrated that the mean daily utility of a health state fell as the duration of time in the health state lengthened. Contradicting this, psychometric evidence indicates that adaptation is often associated with an increase in QoL values (O'Connor 1993).

In our case it is likely that the rating of the 12 week chronic health state could be distorted by the imminence of death to such an extent that answers would bear no relationship to the utility of the health state measured for a longer period of time. Because of the imminence of death the value of the health state could be biased upwards. Alternatively, measurement from the longer period might be contaminated by a perception that life in the worst health state would become unendurable - a saturation effect. This would bias the implied values downwards.

To test the effect of duration we compared 12 month values for the chronic health state three with values obtained for the same health state with an assumed life expectancy of twelve years. More than 82 percent of respondents nominated health state three (severe pain, diarrhoea and nausea) as the worst scenario. Table 3 shows the difference in mean TTO and RS values for the two periods as measured on a scale which uses normal health as the upper reference state and death as the lower reference state (values of 1 and 0 respectively). The results obtained from the two periods were not significantly different at the 5 percent level when measured with the TTO technique. The mean values (0.41 and 0.37 for month and year values respectively) would have been identical if a discount rate of 2 percent was assumed over the 12 year period. The difference between the mean RS values (0.38 and 0.31 for months and years values respectively) is almost twice as large in absolute terms and is significant at the 1 percent level. However, the results and,

particularly for the TTO, do not suggest such a large quantitative discrepancy that subsequent results would be overwhelmed by the use of the longer time period for estimating the true 12 week health state value. While encouraging, the results do not rule out the possibility of similar but offsetting errors in both observations.

— TABLE 3 —
Differences in Mean Values for Chronic Health State Three
(12 Months and 12 Years: TTO and RS Values)

Health State Duration	TTO		RS	
	Mean	Difference	Mean	Difference
12 months	.4126		.3799	
12 years	.3756	.037	.3122	.068†

† Significant at 1 percent level.

As the present data do not identify potential biases we selected the shorter, 12 month health state values in subsequent calculations. In the absence of a gold standard the validity of our approach can not be ascertained. However, alternative approaches to the conversion of temporary states probably have less validity. For example, the Quality of Well-Being Scale (QWB) developed by Kaplan et al. (1978) is based upon the RS evaluation of a single day's morbidity and the RS has death (a permanent state) as one of the anchor points. Nord et al. (1992) have argued that it is the logical inconsistency of moving from one morbidity day to death that may explain some of the perverse results from the QWB.

3 COMPARISON OF MEASUREMENT TECHNIQUES

In the absence of a gold standard for health status measurement, convergence between measurement techniques is often used to determine whether indices which purport to measure health status have criterion validity. Below we examine the relationship between the TTO and the RS using temporary health state values other than health state three. As noted, health state three was chosen by 82 percent of respondents as the worst health state or reference state. Temporary health states were therefore measured on a scale which used normal and health state three as the upper and lower reference states, these were assigned values of 1.0 and 0.0 respectively.

The RS relative utility values were read directly from a thermometer-type scale which showed a lower value of 0.0 corresponding with health state three and an upper value of 100 corresponding with normal health as defined earlier.

3.1 Comparison of Mean TTO and RS Values

Mean values for the temporary health states are given in Table 4. This demonstrates that the difference between the mean values obtained for each health state from the RS and the TTO is significant, a result which has also been found in other such comparisons. The TTO produced substantially and consistently higher values than the RS. The largest absolute difference of 0.32 was for health state four (severe pain) and the lowest absolute difference of 0.10 for health state two (laparoscopic cholecystectomy). The RS values are lower than the TTO values by an average of 35.5 percent with a range from 12 percent for health state two (laparoscopic cholecystectomy) to 56 percent for health state six (moderate pain and severe diarrhoea).

The 95 percent confidence intervals are relatively small but tend to increase with the lower values for the TTO indicating a greater degree of variability in responses to the more severe health states when using the TTO. Standard deviations for health states five, one and six range from 0.13 to 0.24 for the RS as compared with 0.28 to 0.32 for the TTO. Standard deviations for the higher rating health states two, seven and four are very similar for both methods (0.18 to 0.22 for the RS as compared with 0.18 to 0.23 for the TTO). The observed difference in mean TTO and RS health state values are all significant at the 1 percent level.

— TABLE 4 —
Comparison of TTO and RS Values for Temporary Health States

Health States*	Time Trade Off			Rating Scale			Difference in Means TTO - RS‡	Ratio RS/TTO
	Mean	SD	Confidence Interval†	Mean	SD	Confidence Interval†		
HS2 Laparoscopic cholecystectomy (normal)	0.82	.21	0.77 - 0.86	0.72	.19	0.68 - 0.75	.10	.88
HS7 Moderate pain	0.80	.18	0.77 - 0.84	0.54	.18	0.50 - 0.58	.26	.67
HS4 Severe pain (periodically)	0.74	.23	0.69 - 0.79	0.42	.22	0.37 - 0.46	.32	.58
HS5 Severe diarrhoea	0.64	.28	0.58 - 0.69	0.36	.15	0.33 - 0.39	.28	.56
HS1 Open cholecystectomy (normal)	0.62	.34	0.55 - 0.69	0.46	.24	0.41 - 0.51	.16	.74
HS6 Moderate pain/severe diarrhoea	0.43	.32	0.36 - 0.49	0.19	.13	0.16 - 0.22	.24	.44

† 95 percent confidence interval.

‡ All differences are significant at the 1 percent level.

* Health State Descriptions are presented in Appendix 2.

3.2 Comparison of Mean Values Between Health States

Temporary Health State Values

The differences between mean temporary health state values measured by the same technique was tested using a simple t-test. The null hypothesis is that the mean values for the different health states are equal. As shown by Table 5 the two techniques produce some notably different outcomes.

When measured with the TTO the difference of 0.02 between health state two (laparoscopic cholecystectomy) and health state seven (moderate pain) is not statistically significant. Using the RS the difference of 0.18 between these two health states is both statistically significant and quantitatively substantial in absolute terms. Similarly, the difference between health state five (severe diarrhoea) and health state one (open cholecystectomy) is only 0.02 and not statistically significant when measured with the TTO but the difference of 0.10 as measured by the RS is significant at the 1 percent level. By contrast there is no statistically significant difference between health state one (open cholecystectomy) and health state four (periodic severe pain) when measured with the RS (absolute difference 0.04) but the difference of 0.12 is significant at the 1 percent level when measured with the TTO.

Utility Values

For the calculation of quality adjusted life years the temporary health state values were converted into 'utility' values by recalibrating the temporary health state values on a 0-1 (death - normal health) scale. To do so the worst health state (usually health state three) was measured as a chronic health

state (ie for a period of 12 months followed by death) and other health state values were recalculated relative to this value.

The calculations were carried out as follows using the formula:

$$U_i = U_3 + V_i (1 - U_3)$$

where V_i is the value of the temporary health state S_i on the scale where $V_3 = 0$ (V_3 is the value of S_3 when it is the worst temporary health state).

U_3 = utility of S_3 measured as a chronic state on the scale where normal health and death are the upper and lower reference states.

For example, with $U_3 = 0.2$ and $V_i = 0.5$ then $U_i = 0.2 + 0.5 (0.8) = 0.6$.

Results are shown in Table 6.

— TABLE 5 —
t-Values of Differences in Mean Values of Temporary Health States

TTO Health State Values						
Health States	HS1	HS2	HS4	HS5	HS6	HS7
HS 1 Open cholecystectomy (normal)		-5.87**	-3.22**	-0.62	-4.39**	-5.74**
HS 2 Laparoscopic cholecystectomy (normal)			-2.67**	5.68**	-10.55**	0.66
HS 4 Severe pain (periodically)				2.82**	-8.29**	-2.35*
HS 5 Severe diarrhoea					-6.52**	5.55**
HS 6 Moderate pain/severe diarrhoea						-11.27**
HS 7 Moderate pain						
Rating Scale Health State Values						
Health States	HS1	HS2	HS4	HS5	HS6	HS7
HS 1 Open cholecystectomy (normal)		-11.87**	1.35	3.57**	-9.82**	-3.17**
HS 2 Laparoscopic cholecystectomy (normal)			-10.86**	16.27**	-24.78**	8.91**
HS 4 Severe pain (periodically)				2.19*	-9.69**	-5.79**
HS 5 Severe diarrhoea					-11.86*	8.87*
HS 6 Moderate pain/severe diarrhoea						-19.21*
HS 7 Moderate pain						

* Significant at the 5 percent level.

** Significant at the 1 percent level.

— TABLE 6 —
Utility Values from the Seven Health States:
Means and Confidence Intervals

Health States		TTO		RS	
		Mean	Confidence Interval†	Mean	Confidence Interval†
HS2	Laparoscopic cholecystectomy (normal)	0.90	0.87 - 0.93	0.81	0.77 - 0.84
HS7	Moderate pain (continuous)	0.90	0.87 - 0.92	0.70	0.66 - 0.74
HS4	Severe pain (periodically)	0.88	0.86 - 0.90	0.62	0.57 - 0.67
HS5	Severe diarrhoea	0.81	0.78 - 0.85	0.58	0.54 - 0.63
HS1	Open cholecystectomy (normal)	0.81	0.77 - 0.86	0.65	0.60 - 0.70
HS6	Moderate pain/severe diarrhoea	0.68	0.63 - 0.74	0.48	0.43 - 0.53
HS3	Severe pain/severe diarrhoea/nausea	0.47	0.40 - 0.54	0.39	0.33 - 0.45

† Confidence interval at 95 percent level.

Comparison of Temporary and Utility Health State Values

Although persons may perceive differences between the temporary health states when measured relative to other temporary health states these differences are necessarily reduced when the same health states are evaluated on a scale which includes the full range of health states, ie normal health to death. That is, a difference which is significant when viewed relative to other outcomes of gallstone disease may become insignificant when measured against the full spectrum of health related quality of life. For the purpose of resource allocation and the comparison of projects the latter perspective is the correct one.

To test the extent to which this was true for outcomes of treatments for gallstone disease, the significance of the differences between the recalibrated health states was calculated using the standard t-test. As shown in Table 7, when measured relative to full health and death the differences between health state four (severe periodic pain), health state two (laparoscopic cholecystectomy) and health state seven (continuous moderate pain), were no longer significant at the 5 percent level. The difference between RS values remained statistically significant with the exception of health state four (periodic severe pain) and health state one (open cholecystectomy) which ceased being statistically different at the 5 percent level.

— TABLE 7 —
t-Values of Differences in Mean Health State Utilities

TTO Health State Utilities		HS1	HS2	HS4	HS5	HS6	HS7
HS 1	Open cholecystectomy (normal)		-3.97**	-3.06**	-0.13	-4.48**	-4.26**
HS 2	Laparoscopic cholecystectomy (normal)			-1.49	3.94**	-7.50**	0.51
HS 4	Severe pain (periodically)				2.92**	-7.02**	-1.43
HS 5	Severe diarrhoea					-5.22**	4.11**
HS 6	Moderate pain/severe diarrhoea						-8.16**
HS 7	Moderate pain						
Rating Scale Health State Utilities		HS1	HS2	HS4	HS5	HS6	HS7
HS 1	Open cholecystectomy (normal)		-9.67**	1.21	3.43**	-8.24**	-3.10**
HS 2	Laparoscopic cholecystectomy (normal)			-8.70**	11.99**	-16.30**	7.36**
HS 4	Severe pain (periodically)				2.08*	-8.19**	-5.21**
HS 5	Severe diarrhoea					-9.93**	7.29**
HS 6	Moderate pain/severe diarrhoea						-13.08**
HS 7	Moderate pain						

* Significant at the 5 percent level.

** Significant at the 1 percent level.

3.3 Distribution of Health State Values

The distributions of the TTO value were distinctly non-normal. With the exception of health state six which had the lowest value, all TTO health states had negatively skewed distributions (Pearson's Coefficient of Skew > 0.50). Respondents were using the higher values relatively more often than the lower values. By contrast, health state values from the RS had a more normal distribution. With the exception of health state two and health state six (the highest and lowest rated health states respectively), RS distributions had a skewness index of less than + or - 0.50. That is, respondents tended to spread their answers more evenly across the scale (See Table 8).

— TABLE 8 —
Distribution of RS and TTO Values:
Pearson's Coefficient of Skew

Health States	TTO	RS
HS2 Laparoscopic cholecystectomy (normal)	-.216	-.95
HS7 Moderate pain	-1.64	-.42
HS4 Severe pain (periodically)	-1.42	.16
HS5 Severe diarrhoea	-.58	-.06
HS1 Open cholecystectomy (normal)	-.73	-.28
HS6 Moderate pain/severe diarrhoea	.19	.75

3.4 Relationship Between the Measurement Techniques

It is generally accepted in the psychometric literature that scale values produced by different methods may differ because the tasks asked from respondents may invoke different cognitive processes. It is argued, however, that the different methods should produce the same rankings (Froberg & Kane 1989). Below we examine the relationship between the health state values as measured by the RS and TTO.

3.4.1 Ranking

Table 9 shows that in this study both the TTO and RS method would have ranked health states similarly with the exception of health state one.

— TABLE 9 —
Rank Order of Temporary Health States

Health States	TTO Rank	RS Rank
HS2 Laparoscopic cholecystectomy (normal)	1	1
HS7 Moderate pain	2	2
HS4 Severe pain (periodically)	3	4
HS5 Severe diarrhoea	4	5
HS1 Open cholecystectomy (normal)	5	3
HS6 Moderate pain/severe diarrhoea	6	6

Health state one (open cholecystectomy) ranks below severe pain (health state four) when measured with the TTO and above this health state when measured with the RS. As individuals are expressing the same preferences with the two scales the discrepancy points to the possibility that at least one of the scales is less reliable in representing health state preferences accurately.

3.4.2 An Exchange Rate Between RS and TTO Values?

A number of studies have investigated the relationship between the RS and TTO values (Torrance 1976; Richardson et al. 1989). This is an important issue. If a stable functional relationship existed then we could predict the value of the health state on one measurement technique from knowledge of its value on another. It would offer some evidence that we may be measuring the same underlying construct and it would allow choice of the most efficient tool for measuring health state values.

To test whether individuals were expressing similar preferences with the two measurement techniques used here we investigated the association between the two measures using regression analysis. Prior work by Torrance (1976) found no systematic linear relationship between RS and TTO health state values at the level of the individual but identified a number of functional relationships between the mean TTO and mean rating values. He advocated the use of the power function of the form $RS = 1 - (1 - TTO)^\beta$. The function is consistent with the relationships found in psychometric research between subjective estimates of sensation magnitudes. In Torrance's research the coefficient β was found to have a value of 0.62 and the resulting function was recommended for general use.

We investigated the relationship between the two measurement methods using both mean and individual values for the health states, and using both a linear and a power function relationship between the TTO and RS values.

3.4.3 Individual Values

From Table 10 it can be seen that for individual values the only statistically significant linear relationships (at the 5 percent and 1 percent level respectively) are between the TTO and the RS for values of health state two (laparoscopic cholecystectomy) and health state one (open cholecystectomy). The highest correlation between the two measurements is achieved for health state one which has R^2 of 0.21739. However, even the significant results reported for health state one and health state two have β coefficients that differ by over 50 percent. This implies that an incremental change in the independent variable (TTO) does not correspond with the same incremental change in the dependent variable (RS) and that in general, a similar linear transformation does not exist between the two sets of scale values. That is, the psychological quantity measured by each scaling technique is either different or calibrated in different units.

The results obtained using a power function to transform data are also reported in Table 10. They further support the Torrance finding that no functional relationship exists between the two measures at the individual level. The R^2 is generally small and is significant at the 5 percent level only for health state one (open cholecystectomy).

— TABLE 10 —
Results of Linear and Power Function Regression of RS (Dependent)
and TTO (Independent) Health State Values from Individuals

Health States	Linear Function			Power Function†		
	R ²	β	SE	R ²	β	SE
HS2 Laparoscopic cholecystectomy (normal)	.05476 *	20.73	9.08	.02865	.11	.07
HS7 Moderate pain	-.00091	-3.05	10.71	.00003	.00	.04
HS4 Severe pain (periodically)	.00921	9.49	10.43	.00158	-.02	.05
HS5 Severe diarrhoea	.02211	8.85	6.2	.00002	.00	.03
HS1 Open cholecystectomy (normal)	.21739 **	33.85	6.81	.05545 *	.12	.06
HS6 Moderate pain/severe diarrhoea	.01838	5.59	4.31	.00063	.00	.02

† $RS = 1 - (1 - TTO)^\beta$
* Significant at the 5 percent level
** Significant at the 1 percent level

3.4.4 Means

A similar analysis of the linear and power functional relationship between the scaling techniques was carried out using the mean values obtained for each of the health states. Results are reported in Table 11. The results were significant at the 5 percent level.

— TABLE 11 —
Results of Linear and Power Function Regression of RS
(Dependent) and TTO (Independent) Mean Health State Values

Mean Values	R ²	B	S.E
Linear Function RS = $\alpha + B(TTO)$.79416 *	1.09	.28
Torrance Power Function RS = $1 - (1 - TTO)^{.62}$.80969 *	.68	.16

* Significant at 5 percent level.

The correlation achieved by using the power function where the exponent is 0.62. ($R^2 = 0.81$) was almost identical to that achieved with the linear regression results ($R^2 = 0.79$).

While these results indicate a satisfactory correlation between the two scaling techniques they tend to obscure the significant differences between the individual health state means predicted from these relationships. This is demonstrated in Table 12. The ratio of actual RS to TTO values ranges from a ratio of 0.88 for the highest value states to 0.44 for the lowest rating state, a difference of 44 percentage points (column 3). However, there is a similar divergence in mean values after transforming the RS values with the power function as used by Torrance. The relationship between the actual and predicted RS values range from a ratio of 0.90 for health state two to 1.53 for health state six (column 5). These findings further demonstrate that care must be taken in applying a general transformation to a range of health state means when the differences between them is not systematic. While the correlation co-efficients indicate a satisfactory relationship the table demonstrates that the actual values remain very dissimilar and that the transformation does not produce units of similar magnitude. This is analogous to the fact that a high correlation between variables does not indicate that the units in which the variables are measured are similar, a result noted above for the two cases where there was a significant linear relationship, viz health state one and health state two when measured at the level of the individual.

— TABLE 12 —
Comparison of Mean Health State Values:
TTO, RS and Predicted RS†

Health States		TTO (1)	RS (2)	RS/TT O (3)	^ RS (4)	^ RS/TTO (5)
HS2	Laparoscopic cholecystectomy (normal)	.82	.72	.88	.65	.90
HS7	Moderate pain	.80	.54	.67	.63	1.17
HS4	Severe pain (periodically)	.74	.42	.57	.57	1.36
HS5	Severe diarrhoea	.64	.36	.56	.47	.73
HS1	Open cholecystectomy (normal)	.62	.46	.74	.45	.98
HS6	Moderate pain/severe diarrhoea	.43	.19	.44	.29	1.53

^
† (RS) = 1 - (1 - TTO)^{.62}

4 SUMMARY AND DISCUSSION

The comparisons between the two measurement techniques used here highlight two important considerations in the choice of an appropriate measurement technique for QALY calculations.

The psychological quantity measured by each scaling technique is either quite different or is measured in different units. Importantly, for decision makers this means that changes in health states effected by a health program as measured by QALYs would depend upon the choice of the scaling technique.

Rating scales may not be an efficient method for eliciting health state values for resource allocation.

The controversy surrounding the two measurement techniques discussed above for the use of health status measurement in CUA is by no means resolved. Our study further confirms the ambiguous relationship between these scaling techniques and lends further support to the belief that they are measuring fundamentally different quantities. In the absence of a gold standard for such measurement and in the absence of a stable and systematic relationship between the scale measures, the choice between them must rest on other criteria which relate both to the objective of the analysis and the practical requirements of the task. The primary criteria should be the extent to which the two techniques had been 'validated' by a process of relevant hypothesis testing. While the psychometric properties of the RS have been extensively explored in other contexts, there has been little relevant testing in the context of resource allocation and some serious disconfirmation of the validity of RS based measurement instruments in this context (Nord et al. 1993). Secondly, all else being equal, preference might be given to the technique which imposed the simplest cognitive burden upon respondents. Rating scales are often considered to be more efficient in collecting population values because they are a simple and reliable instrument. Our study did not support this view. In the context of health status measurement which required respondents to trade-off length of life against quality of life, the RS was less well understood than the TTO. Additionally, the majority of respondents agreed that it was acceptable to trade life for improvement in the quality of life.

If the stated values produced by survey respondents were a correct reflection of the respondent's behaviour when faced with real decisions, choice between techniques might be based upon 'face validity', ie the extent to which the scales appear to measure what is required by the analysis. The objective of health status measurement in the context of CUA is to derive homogenous years of life as a unit of output. The outcome units produced by the measurement technique needs to have an interval property. It is our view that the TTO is best suited to achieve this precisely because it offers an objective standard for measuring duration. There is a clear unambiguous meaning attached to the unit of measurement (weeks, months or years) used in the TTO method. The meaning of the measurement unit achieved by the RS is less clear. The RS is valid only if two separate functional relationships are valid. First, respondents need to translate the value of the health states to corresponding values on the scale. And secondly, the units of the scale must be appropriate for trading off quantity and quality of life. Respondents are not provided with this information and there is no reason for believing that the scale values incorporate this property. In other words, the RS does not even achieve face validity.

It would be highly desirable for a standard measurement technique to be employed in CUA. It is a prerequisite to the comparison of different studies. We see the TTO as an important contender for this role. In a previous paper we have argued the case for the development of standardised health state description via a multi-attribute index descriptive system (Cook and Richardson 1993). Such an index needs to be scaled by a commonly accepted measurement technique. The resulting instrument appropriately validated on variety of populations would serve as the gold standard for allocative decision making for health programs affecting a wide range of health states.

REFERENCES

- Anderson, N.H. 1976, 'How functional measurement can yield validated interval scales of mental qualities', *Journal of Applied Psychology*, no. 61, pp. 677-692.
- Brooks, R.G. 1991, Health status and quality of life measurement: Issues and developments, The Swedish Institute for Health Evaluation, Lund.
- Cook, J. & Richardson, J. (1993), Quality of life during the treatment of gallstone disease: Issues in the development of health state scenarios, Working paper 32, NHMRC National Centre for Health Program Evaluation, Melbourne.
- Cook, J., Richardson, J. & Street, A. 1993, Cost utility analysis of treatments for gallstone disease: Interim results, Working paper 28, NHMRC National Centre for Health Program Evaluation, Melbourne.
- Drummond, M.F., Stoddart, G.L. & Torrance, G.W. 1987 *Methods for the Economic Evaluation of Health Care Programs*, Oxford Medical Publications, Oxford.
- Froberg, D.G. & Kane, R.L. 1989, Methodology for measuring health state preferences II: Scaling methods, *Journal of Clinical Epidemiology*, vol. 42, pp. 459-71.
- Gescheider, G.A. 1988, 'Psycho-physical scaling', *Annual Review of Psychology*, vol. 39, pp. 169-200.
- Kaplan, R.M., Bush, J.W. & Berry, C.C. 1976, 'Health status: Types of validity and the index of wellbeing', *Health Services Research*, vol. 11, no. 4, pp. 478-507.
- Kind, P. & Rosser, R. 1988, 'The quantification of health', *European Journal of Social Psychology*, vol. 18, pp. 63-77.
- Mehrez, A. & Gafni, A. 1989, Healthy year equivalents: How to measure them using the standard gamble approach, Working Paper 22, Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Ontario, Canada.
- Nord, E., Richardson, J. & Macarounas-Kirchmann, K. 1992, Social evaluation of health care versus personal evaluation of health states: Evidence on the validity of four health state scaling instruments using Norwegian and Australian survey data, Working paper 23, National Centre for Health Program Evaluation, Melbourne.

- O'Connor, R. 1993 Issues in the measurement of health related quality of life, Working paper 30, National Centre for Health Program Evaluation, Melbourne.
- Richardson, J. 1990 'What should we measure in health program evaluation', *Economics and Health 1990: Proceedings of the Twelfth Australian Conference of Health Economists*, ed C. Selby Smith, Australian Health Economics Society, Canberra, pp. 80-104.
- Richardson, J., Hall, J., & Salkeld, G. 1989, 'Cost utility analysis: The compatibility of measurement techniques and the measurement of utility through time', *Economics and Health 1989: Proceedings of the Eleventh Australian Conference of Health Economists*, ed C. Selby Smith, Australian Health Economics Society, Canberra, pp. 31-63.
- Ross, K.N., Farish, S. & Plunkett, M. 1988, Indicators of socio-economic disadvantage for Australian schools, Deakin Institute for Studies in Education, Geelong, Victoria.
- Sackett, D.L. & Torrance, G.W. 1978, 'The utility of different health states as perceived by the general public', *Journal of Chronic Diseases*, vol. 31, pp. 697-704.
- Street, A. 1993, Gallstone disease: The cost of treatment, Working paper 29, National Centre for Health Program Evaluation, Melbourne.
- Torrance G.W. 1976, 'Social preferences for health states: An empirical evaluation of three measurement techniques', *Socio-Economic Planning Science*, vol. 10, pp. 129-136.
- Torrance, G.W. 1986, 'Measurement of health state utilities for economic appraisal: A review', *Journal of Health Economics*, vol. 5, pp. 1-30.
- Von Winterfeldt, D. & Edwards, W. 1986, *Decision Analysis and Behavioural Research*, Cambridge University Press, Cambridge.

APPENDIX 1A

Quality of Life Measurement for Gallstone Disease Treatment Options Interview Schedule

Thank you for agreeing to take part in the interview.

As we explained to you earlier, we are trying to find out what you and other people in the community feel about different health problems. Your answers will help us to include people's feelings and preferences when we evaluate different medical treatments.

Do you have any questions?

We'll start with a bit of practice. Please feel free to ask any questions at any time.

I am going to give you some cards which describe how different patients feel. I am going to ask you to tell me which you think is better and which is worse. There are no right or wrong answers.

We will be using a rating scale.

SHOW RESPONDENT THE FEELING THERMOMETER FOR TEMPORARY HEALTH STATES

(HAVE SET OF CREAM-COLOURED PRACTICE CARDS READY TO SHOW RESPONDENT)

****This is called a rating scale. It is used to measure how strongly people feel about things. In our case we are using it to measure how you feel about different health problems. You see that it runs from 0 to 100 with 'Worst Health State' at the bottom and 'Best Health State' at the top. The higher up the scale - the better you feel about your health. The lower down the scale - the worse you feel.**

I am going to ask you how strongly you feel about the health problems of each of the patients described on these cards.

For example, if you feel that this person's health was about 3/4 of the way between the best and the worst you would mark it here (75) - we will use a pin. You see this is about 1/4 below the best health state. Here (50) you would feel it is half way between best and worst. And here (15) it is close to the worst.

The numbers are used to help you decide how far up or down the scale you feel it should be.

Do you understand these ideas?

We'll go ahead with our practice cards.

READ PRACTICE WORST HEALTH CARD (P.1) WITH RESPONDENT

Let us say this is the worst health state we are measuring. We would put the pin for this card at the bottom of the scale - at 0.

PUT MATCHING COLOURED PIN AT 0 AND PLACE CARD BESIDE IT

READ CARD FOR FULL HEALTH WITH RESPONDENT

This card describes full health. We will put the pin for this card at the top of the scale - at 100. This is the best health state.

PUT MATCHING COLOURED PIN AT 100 AND PLACE CARD BESIDE IT

READ THE PRACTICE INTERMEDIATE HEALTH STATE CARD (P.2)

How do you feel about this person's health problems? Where would you put the pin for this card on the scale between 'Worst Health' and 'Best Health'?

It is important that the distance between the pin and the 'Best' and the 'Worst Health' States shows how strongly you feel.

PROMPT FROM ** IF NECESSARY

Do you have any questions?

Now we'll go ahead with the task.

REPLACE PINS AT TOP OF FEELING THERMOMETER

PUT FEELING THERMOMETER IN FRONT OF RESPONDENT

SHOW RESPONDENT (WHITE SET OF) CARDS FOR HS.1 TO HS.9 AND FULL HEALTH

These are the cards that describe how real patients feel. I am going to ask you to rate each of these cards on this scale.

You will notice that the cards are colour coded and there is a coloured pin to match each card. We will take each card one at a time.

I want you to imagine that you are the person described on the card and that you would have these health problems for 12 weeks. After that you would return to full health. Think about how you feel about these health problems.

I am going to ask you first to decide which you feel is the worst. I'll read them through with you and then you can look at them again. Please take as long as you like. Remember, there are no right or wrong answers.

**READ CARDS ONE AT A TIME
GIVE EACH CARD TO RESPONDENT AFTER YOU'VE READ IT ALOUD WITH THEM
ALLOW RESPONDENT TIME TO THINK**

**PROMPTS: you can sort them in any way as we talk
- choose the worst by sorting them in pairs if it's helpful
- spread them out or rearrange them as you think etc....**

WAIT FOR RESPONDENT TO CHOOSE WORST HEALTH STATE

So this is the card you feel describes the worst health state? We will put the pin for this card at 0.

PUT MATCHING COLOURED PIN AT 0 AND PLACE CARD BESIDE IT

READ CARD FOR FULL HEALTH WITH RESPONDENT

This is the card that describes full health. We will put the pin for this card at 100, as before.

LEAVE CARDS IN PLACE AT TOP AND BOTTOM OF THE SCALE

PICK UP REMAINING CARDS (IF RESPONDENT HAS PUT THEM IN ORDER, RETAIN ORDER BUT MAKE NO COMMENT)

**GIVE RESPONDENT CARD YOU HAVE ON TOP
READ CARD AGAIN WITH RESPONDENT IF NECESSARY AT ANY TIME**

Now I am going to ask you to rate each of these cards on the scale between Best and Worst. Where would you put this card on the scale? Please take your time.

Remember the numbers are used to help you decide how far up or down the scale you feel it should be and it is important that the distance between the pin and the `Best' and the `Worst Health' States shows how strongly you feel.

PROMPT FROM ** IF NECESSARY

PLACE FIRST CARD (AND ANY SUBSEQUENT CARDS IF HELPFUL) NEXT TO NUMBER CHOSEN (SO THAT IT CAN BE REFERRED TO IF NEEDED)

GIVE REMAINING CARDS ONE AT A TIME AND ASK RESPONDENT TO PUT THE MATCHING COLOURED PIN ON THE SCALE (CHECKING THAT COLOURS MATCH)

ASK EACH TIME (IF NECESSARY):

Where would you put this card on the scale between `Best Health' and `Worst Health'?

You can give it the same or a different number and you can change them around if you want to.

**WHEN ALL CARDS HAVE BEEN RANKED AND RESPONDENT SEEMS SATISFIED
TAKE WORST HEALTH CARD FOR NEXT PART OF INTERVIEW**

**MAKE SURE ALL PINS ARE SECURE (FOR LATER RECORDING)
COLLECT FEELING THERMOMETER AND REMAINING CARDS**

**SHOW RESPONDENT THE FEELING THERMOMETER FOR
CHRONIC HEALTH STATES**

Now I would like to measure this card - which you feel has the worst health problems - on another scale. Please note that this scale has DEATH at the bottom and FULL HEALTH at the top.

POINT TO DEATH AND FULL HEALTH

Now I am going to ask you to imagine that you have only 12 years to live. I am not suggesting this is your real life expectancy but we need to know how you would feel about having these health problems for a longer time. So we are asking everyone to imagine they have only 12 years to live.

So, imagine you have only 12 years to live and you have the health problems of this patient. Your condition will remain unchanged for the rest of your life, ie 12 years.

How do you feel about having these health problems for 12 years? How does it compare to being in full health (which is our best health state) or dying immediately (being dead is now our worst health state)?

Do you understand what I'm asking?

Please place the red pin on the scale where you feel it should go between 'Full Health' and 'Death'. Please take your time.

RECORD NUMBER ON RESPONSE FORM

Now I need to measure a shorter time. I want you to imagine that you have only 12 months to live (and you have the health problems of this patient and that your condition will remain unchanged for the rest of your life, ie 12 months).

Please place the blue pin where you feel it should go between 'Full Health' and 'Death'. Take as long as you like. (You may give it the same number as the last one.)

RECORD NUMBER ON RESPONSE FORM

**COLLECT FEELING THERMOMETER AND PUT ASIDE
PUT WORST HEALTH CARD WITH TTO BOARD
SHUFFLE REMAINING WHITE CARDS SLIGHTLY
(SO THAT THEY WILL BE PRESENTED IN RANDOM ORDER)
HAVE PRACTICE CARDS READY**

Now we're going to begin the second part of the interview using some different ways of measuring how you feel.

**PREPARE TTO BOARD FOR TEMPORARY HEALTH STATES
(BLUE SLIDE IN TOP BAR, BLUE AND PINK SLIDE IN BOTTOM BAR)
(WEEKS)**

ATTACH PRACTICE CARD P.1 TO BOTTOM VELCRO STRIP - LIFE B

ATTACH PRACTICE CARD P.2 TO TOP VELCRO STRIP - LIFE A

NOW SHOW RESPONDENT THE TTO BOARD

These two bars represent the time you spend being either of these two patients. The numbers show how many weeks you spend with these health problems.

When the person has health problems (as described on the card), we show the time as blue. When the person returns to full health we show it as pink. If we want to show that the person dies, we show it as black.

Do you understand these ideas?

The upper bar represents LIFE A. This will always be 12 weeks followed by a return to full health.

The lower bar represents LIFE B. This patient too will return to full health at the end of 12 weeks but I will be changing the time spent with these health problems.

I am going to offer you a number of choices. I want you to imagine that the choices are real and that what you choose will actually happen to you each time.

I want you to think of each choice as being separate from all the others. It does not matter how you have chosen before or what choice you make afterwards. There are no right or wrong answers. We only want to know how you feel about each choice.

(Prompt: remember that the condition of each person stays the same until they return to full health.)

Please tell me if you would like to stop the interview at any time to read any of the descriptions again. There is no hurry.

Are you ready? We'll begin with a bit of practice.

We need to read the description of the person in LIFE A.

READ PRACTICE CARD (IN LIFE A) WITH RESPONDENT

Now we'll read the description of the person in LIFE B. Take your time.

READ PRACTICE CARD IN LIFE B WITH RESPONDENT

I am going to offer you a choice. If you think the choices are equal tell me and I will mark them both.

Please imagine yourself as the person and think carefully about how you would feel if you were in their place.

RUN THROUGH STEPS 1 TO 13 BELOW WITHOUT RECORDING RESPONSE

BEGIN TTO FOR TEMPORARY HEALTH STATES

We are now going to go through the same steps for each patient. Are you comfortable with what I'm asking you to do?

**REMOVE PRACTICE CARD FROM TOP STRIP
ATTACH FIRST WHITE (HS) CARD TO TOP VELCRO STRIP - LIFE A**

**REMOVE PRACTICE CARD FROM BOTTOM STRIP
ATTACH WORST HEALTH CARD TO BOTTOM VELCRO STRIP - LIFE B**

**FOR EACH NEW CARD READ THE DESCRIPTION OF (NEXT) PERSON IN LIFE A AND
PERSON IN LIFE B (IF NECESSARY)**

(PERSON IN LIFE B STAYS THE SAME FOR ALL TEMPORARY TTOS)

ALLOW RESPONDENT TIME TO THINK

RECORD LIFE A AS HS... ON RESPONSE FORM (FOR TTO TEMPORARY)

STEP 1

Imagine that you would spend 12 weeks with the health problems of each of these patients and that in both cases you would return to full health after 12 weeks.

Which life would you prefer?

IF RESPONDENT ANSWERS LIFE A MARK RESPONSE - OFFER STEP 2

IF RESPONDENT ANSWERS LIFE B PROMPT:

This means that you would prefer 12 weeks with the health problems of LIFE B to 12 weeks with the health problems of LIFE A. Is this correct?

IF RESPONDENT ANSWERS `YES':ASK `WHY' AND RECORD ANSWER VERBATIM

**DISCONTINUE TTO EXERCISE FOR THIS CARD AND GO TO THE NEXT CARD
IF RESPONDENT ANSWERS `NO' MARK RESPONSE (LIFE A) - OFFER STEP 2**

STEP 2

Now I am going to change the board to show that in Life B you would be in Full Health. Imagine that you would spend 12 weeks with the health problems of LIFE A after which you would return to full health or that you would spend the same 12 weeks in full health.

Which would you rather be? (I assume you would prefer full health?)

IF RESPONDENT CHOOSES FULL HEALTH MARK RESPONSE - OFFER STEP 3

IF RESPONDENT CHOOSES LIFE A PROMPT:

Do you mean you would prefer to live for 12 weeks with the health problems of LIFE A rather than have 12 weeks of full health?

IF `NO' MARK LIFE B AS PREFERRED - OFFER STEP 3

IF `YES' MARK LIFE A AS PREFERRED; ASK `WHY' AND RECORD RESPONSE VERBATIM - GO TO THE NEXT CARD

STEP 3

Now I have made the time you have to spend with the health problems of LIFE B eleven weeks after which you would return to full health.

Which life would you prefer now?

IF ANSWER IS LIFE B - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE A - MARK RESPONSE - OFFER RESPONDENT STEP 4

STEP 4

Now I have made LIFE B one week with these health problems after which you return to full health.

Which life would you prefer now?

IF ANSWER IS LIFE A - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE B - MARK RESPONSE - OFFER RESPONDENT STEP 5

STEP 5

Now you would have ten weeks with the health problems of LIFE B after which you would return to full health.

Which would you prefer?

IF ANSWER IS LIFE B - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE A - MARK RESPONSE - OFFER RESPONDENT STEP 6

STEP 6

Now I have made the time you have to spend with health problems in LIFE B two weeks after which you return to full health.

Which would you prefer now?

IF ANSWER IS LIFE A - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE B - MARK RESPONSE - OFFER RESPONDENT STEP 7

STEP 7

Now I have made the time you have to spend with health problems in LIFE B nine weeks after which you would return to full health.

Which would you prefer now?

IF ANSWER IS LIFE B - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE A - MARK RESPONSE - OFFER RESPONDENT STEP 8

STEP 8

Now I have made the time with health problems in LIFE B three weeks after which you return to full health.

Which would you prefer now?

IF ANSWER IS LIFE A - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE B - MARK RESPONSE - OFFER RESPONDENT STEP 9

STEP 9

Now I have made the time with health problems in LIFE B eight weeks after which you would return to full health.

Which would you prefer?

IF ANSWER IS LIFE B - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE A - MARK RESPONSE - OFFER RESPONDENT STEP 10

STEP 10

Now I have made the time with health problems in LIFE B four weeks after which you return to full health.

Which would you prefer now?

IF ANSWER IS LIFE A - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE B - MARK RESPONSE - OFFER RESPONDENT STEP 11

STEP 11

Now I have made the time with health problems in LIFE B seven weeks after which you would return to full health.

Which would you prefer now?

IF ANSWER IS LIFE B - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE A - MARK RESPONSE - OFFER RESPONDENT STEP 12

STEP 12

Now I have made the time with health problems in LIFE B five weeks after which you return to full health.

Which would you prefer now?

IF ANSWER IS LIFE A - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE B - MARK RESPONSE - OFFER RESPONDENT STEP 13

STEP 13

Now I have made the time with health problems in LIFE B six weeks after which you return to full health.

Which would you prefer now?

IF ANSWER IS LIFE B - MARK RESPONSE - AND GO TO NEXT CARD

IF ANSWER IS LIFE A - MARK RESPONSE - AND GO TO NEXT CARD

***When All Patients Have Been Measured as Temporary Health States, Begin
TTO for Chronic Health States***

Instructions to Interviewer

WHAT TO DO IF RESPONDENT ASKED FOR INFORMATION ON THE TREATMENT OPTIONS

Ask respondent to wait until end of interview when you will explain the different treatment options.

WHAT TO DO AT END OF INTERVIEW

1. Explain treatment options if asked.

Stress you are a lay person and any further information must be sought from doctor.

Open cholecystectomy:

Surgical removal of the gallbladder under a general anaesthetic.

Laparoscopic cholecystectomy:

Surgical removal of the gallbladder under a general anaesthetic through a small cut around the navel.

Gallstone lithotripsy:

With lithotripsy gallstones are broken up into tiny fragments with shockwaves from a machine outside the body. The gallbladder is cleared by giving patients tablets that will help to dissolve the tiny bits of gallstones and they pass out of the body naturally.

2. Stress that times in after treatment health states are artificial.
They have been made 12 months or 12 years so that we can measure preferences for the health states in a consistent way.

Again any further information about the duration of actual health states will need to come from their doctor.

WHAT TO DO IF WORST HEALTH STATE CHANGES

TTO

When administering the TTO method you will have asked the respondent to name the worst possible health state which then becomes the reference state against which all other health states are to be evaluated. Because it is the worst health state you would expect that the respondent always indicates a preference for less time spent in the worst health state than time in the health state being evaluated.

It is possible, however, that during the course of the interview the respondent will perceive another health state as worse than the reference state. If this is the case you will need to reverse the order of the health states. (That is B now becomes A, and, A becomes B) and continue with the TTO as before.

MAKE SURE THAT YOU MAKE A CORRESPONDING CHANGE TO THE HEALTH STATES RECORDED AS A AND B ON THE TTO MARKING FORM.

WHAT TO DO IF RESPONDENT WANTS TO GIVE A TTO VALUE FOR B WITHOUT GOING THROUGH ALL STEPS

Tell respondent that it is important for our methodology that we ask questions in a particular order. Suggest you will shorten procedure. Start with the TTO question which precedes the given value by 2.

eg. If respondent says 'I think I would prefer 8 months in health state B to 12 months in health state A' (Step 9) - Start with Trade-Off question - 9 weeks in Life B (Step 7).

Continue until you reach an instruction and stop.

If respondent insists that he/she wants to give you a direct answer, record answer on response form and make note to this effect.

WHAT TO DO IF RESPONDENTS WANTS CLARIFICATION ABOUT MEANING OF WORDS USED TO DESCRIBE HEALTH STATE SCENARIOS.

Tell respondent that the description of the health states are based on what patients have told us about how they felt after treatment. We want respondents to put themselves in the patient place and imagine how they would feel if they used these same words to describe their own health state.

DURING INTERVIEW

Encourage the respondent at all times to take their time to think about the health states before responding. You can adlib on how to do this, eg. suggest that you re-read the health state to them if this appears appropriate.

When they are looking at the card initially encourage 'hands on' use of cards, ie. if they wish to sort or order the cards encourage this.

DEATH

On the question about death, ask them to think carefully for a minute about the possibility of dying tomorrow and what this would imply if this was a serious option.

RATING SCALE

If patients are very unsure of the rating scale *adlib* with an example, eg. 'If I was bedridden I might place the pin here (Number 25), ie. at a low number because it is closer to the worst health state and I feel that it is at least three quarters as bad as the worst health state' etc etc as needed.

APPENDIX 2

Health State Scenarios

Health State	Category	Scenario
1	Open Cholecystectomy (normal)	You have had a successful operation. You get tired very easily and you don't sleep very well at night. The wound gives you a continuous dull sort of pain. You find that you cannot carry out most of your normal activities.
2	Laparoscopic Cholecystectomy (normal)	You have had a successful operation. You feel a little tired. It is uncomfortable for you to move. You cannot exercise or lift heavy things. You find that you can do most of your normal activities.
3	Severe pain Severe diarrhoea Nausea	You are having specialist medical treatment. Your treatment gives you 2 or 3 attacks of continuous agonising pain in your chest and back. The pain can last from half an hour to 4 hours. You can do nothing to relieve the pain. When the pain goes you can return to your normal activities. The treatment also gives you uncontrollable diarrhoea 2 or 3 times a week. You need to be near a toilet most of the time. About once a week the diarrhoea is very painful. About once a week you feel a bit nauseous for a few hours.
4	Severe pain (periodically)	You are having specialist medical treatment. Your treatment gives you 2 or 3 attacks of continuous agonising pain in your chest and back. The pain can last from half an hour to 4 hours. You can do nothing to relieve the pain. When the pain goes you can return to your normal activities.
5	Severe diarrhoea	You are having specialist medical treatment. You have uncontrollable diarrhoea 2 or 3 times a week. You need to be near a toilet most of the time. About once a week the diarrhoea is very painful.
6	Moderate pain Severe diarrhoea	You are having specialist medical treatment. The treatment gives you an uncomfortable heavy feeling in your stomach most of the time. About once a month you also have a cramping in your chest and back. You have uncontrollable diarrhoea 2 or 3 times a week. You need to be near a toilet most of the time. About once a week the diarrhoea is very painful.
7	Moderate pain	You are having specialist medical treatment. The treatment gives you an uncomfortable heavy feeling in your stomach most of the time. About once a month you also have a cramping in your chest and back.