CENTRE FOR HEALTH PROGRAM EVALUATION

WORKING PAPER 88

Exploring the Relationship Between Time Trade-off and Willingness-to-Pay: An Empirical Investigation

Mr Richard Smith
Senior Lecturer, Health Economics Unit,
Centre for Health Program Evaluation,
Monash University

May, 1999
ISSN 1325 0663
ISBN 1 876662 00 X

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.

AUTHOR(S) ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor Jeff Richardson for providing comments on the ideas contained in this research and earlier drafts of the paper, to Dr Graeme Hawthorne for graciously allowing the use of the AQoL data and instrument, to Helen McNeil for conducting the interviews and managing the data, to Neil Day for statistical advice, and finally, of course, to all those who completed the interviews. Needless to say, the usual caveat applies.

ABSTRACT

This paper explores the relative *sensitivity* of the Time Trade Off (TTO) and Willingness-To-Pay (WTP) valuation techniques to changes in health status.

This was achieved by administering a WTP survey to a population of 50 subjects who had previously completed a TTO. These two sets of data were then analysed for relative sensitivity in valuing changes in health state, and also for any trend within the data which might illustrate a level of comparability between TTO and WTP values.

The main findings were that: (i) TTO appeared to perform well in differentiating significantly between different *levels* of health within each dimension, but not so well in differentiating between different *dimensions* of health at the same nominal 'level' of health status; (ii) WTP, compared with TTO, appeared to perform far *less* well in differentiating significantly between different *dimensions* of health at the same nominal 'level' of health status, but performs slightly better at differentiating between different *levels* of health within each dimension; and (iii) there was no significant correlation found between the two values.

Subject to two important caveats (that these results can only be considered indicative, rather than definitive, as they are based on a small sample, and that the WTP values reflect only a limited range of benefit - that of health status), this study indicates that, with the possible of exception of very poor health states, these two techniques should not be considered comparable, although, overall, WTP seems to be more sensitive to changes in *levels* of health status.

However, the directions in which the data point are interesting, novel and worthy of continued research. These are: (i) that WTP values show greater dispersion around the mean/median as states become 'worse', indicating perhaps the increasing impact of the budget constraint; (ii) that neither TTO nor WTP differentiates very well between broad dimensions of health (likely due to either the dimensions used, or that it is the level of health status which is important, and the dimension within which that level occurs is irrelevant); (iii) that WTP better differentiates between levels of health status within a dimension than TTO, implying that it is a more sensitive measure in the valuation of small changes in health status; and (iv) any comparability between techniques would have to be at the levels of health which are commonly viewed as being the worst, where the WTP in time or money is substantial to avoid them.

TABLE OF CONTENTS

Abstract **i**

1 Introduction **1**

2 QALYs and WTP as Measures of Benefit **3**

3 Methods **6**
- 3.1 Health States Values 6
- 3.2 Sampling and Survey Design 9
- 3.3 Interview Procedure 9
- 3.4 Sample Size 10
- 3.5 WTP Estimation 10
- 3.6 TTO Estimation 12
- 3.7 Demographic Influence 12
- 3.8 Statistical Analyses 12

4 Results **13**
- 4.1 Sample Size, Response Rate and Characteristics 13
- 4.2 WTP in Dollar Values 14
- 4.3 WTP Expressed as Proportion of Income 16
- 4.4 TTO Values 17
- 4.5 Sensitivity of TTO and WTP to Changes in Health State 19
- 4.6 Correlation Between TTO and WTP 24

5 Discussion **31**

References **33**

Appendix 1 WTP Studies in Health Care: 1985-1998 **36**

Appendix 2 WTP Questionnaire **41**

List of Tables

Table 1:	AQoL Dimensions and Items	7
Table 2:	Breakdown of Scenarios Used	8
Table 3:	Demographic Details	13
Table 4:	WTP Dollar Values	14
Table 5:	WTP Expressed as a Proportion of Income	16
Table 6:	TTO Values	18
Table 7:	Sensitivity of TTO Values Across Items	20
Table 8:	Sensitivity of TTO Values within Items	21
Table 9:	Sensitivity of WTP Values Across Items	23
Table 10:	Ranking of Questions by TTO, WTP $ and WTP %	25
Table 11:	Rankings Presented as Questions Corresponding to Rank Number	26
Table 12:	Pearson Correlation's Between TTO Versus WTP $ and WTP %	29

List of Figures

Figure 1:	Boxplot of WTP $ Values	15
Figure 2:	Boxplot of % WTP	17
Figure 3:	Boxplot TTO Values	19
Figure 4:	Scatterplot for TTO vs WTP $	28
Figure 5:	Scatterplot TTO Versus WTP %	28

Exploring the Relationship Between Time Trade-off and Willingness-to-Pay

An Empirical Investigation

1 Introduction

Historically, cost-utility analysis (CUA), predominantly expressing benefit in terms of Quality-Adjusted Life Years (QALYs), has been the mainstay of incorporating the valuation of health outcomes in economic evaluation (Drummond et al, 1997). However, valuation in monetary terms, using willingness-to-pay (WTP), and thus the use of cost-benefit analysis (CBA), has increased exponentially in popularity in health economics over the last decade (see appendix 1 for a list of studies in health care published from 1985 to 1998). Despite both forms of evaluation seeking to express individual valuation of benefits, and the classification of both as forms of economic evaluation, there are theoretical differences. It is argued, for example, that cost-benefit analysis is the only form of evaluation based on welfare economic theory (Pauly, 1995), whilst CUA is based on the decision-maker school (Sugden & Williams, 1978), that only CBA will allow judgements of allocative efficiency to be determined, and that CBA allows a wider scope of benefits to be incorporated than simply 'health' (Olsen & Smith, 1999). However, there have been few studies which have attempted an empirical comparison of the two techniques.

To date only three studies (O'Brien & Viramontes, 1994; Thompson, 1986; Bala et al, 1998) have compared the assessment of utility weights and WTP values. The study by O'Brien & Viramontes (1994) concluded that "there is some evidence of convergent validity for WTP with preferences measured by SG" (p289). The most recent paper, by Bala et al (1998) was particularly concerned with the variation in WTP and Standard Gamble (SG) techniques according to the incorporation of time: QALYs combining time with utility weights as a separate stage in construction, whereas WTP assessing a specific health-time profile. This paper aimed to assess another factor in the debate concerning the relationship between WTP and QALYs: that of the relative *sensitivity* of TTO and WTP techniques to small and large changes in health status. The analysis presented also differs from those previously in that: (i) time has been

‘standardised’ by assessing the TTO and WTP values for a health status change lasting exactly one year; (ii) it is concerned with comparing WTP with TTO, rather than SG; and (iii) in assessing sensitivity this study is looking for *discriminant* validity, rather than simply *convergent* validity of previous studies.

Specific hypotheses tested are that: (i) there is little correlation between WTP and TTO; (ii) TTO is less sensitive than WTP to small changes in health status; and (iii) TTO and WTP are equally sensitive for large changes in health status.

The paper is split into 5 sections. Following this introduction, section 2 provides a brief background summary of the use of WTP and QALYs as measures of health benefits. Section 3 then covers the methods used in this study, section 4 outlines the results and section 5 concludes.

2 QALYs and WTP as Measures of Benefit

Traditionally health outcome measurement has relied upon changes in morbidity and mortality, but increasingly there is recognition of the wider outcomes of health programs, including changes in physical, emotional and social well-being (Drummond et al, 1989). This argument recognises that health outcome assessment should reflect the multiplicity of health outputs which are **valued** by patients.

Cost-utility analysis (CUA) and cost-benefit analysis (CBA) are related techniques of economic evaluation for estimating the **value** of such multiple outcomes (Drummond et al, 1997; Johannesson, 1996). They both try to value heterogeneous multi-dimensional outcomes to form a single unit (and as such are therefore distinct from cost-effectiveness analysis (CEA) which uses only unidimensional outcome measures).

CUA expresses benefits most commonly in terms of 'Quality Adjusted Life Years' (QALYs) (although other measures may be used, such as the 'Healthy Year Equivalent' (HYE), these are more infrequent) (Mehrez & Gafni, 1992). QALYs attempt to measure benefit by combining duration and quality of life in a single estimate. The procedure typically adopted involves assignment of a utility, or preference, weight between 0 and 1[1] to a health 'state', where 1 corresponds to 'good' or 'normal' health, and 0 to death. These health states (incorporating various aspects of quality of life, such as physical, emotional and social well-being) are described in a series of well-defined 'scenarios' (Smith et al, 1993). Often, but not exclusively, patients with experience of these health states are used in the construction of such scenarios (Smith & Dobson, 1993). The weight derived for each scenario is then multiplied by the duration of that state, and might be multiplied by the probability of experiencing that state. Thus,

$$Q = (HS_1 - HS_0) * T_D * p$$

where Q is the outcome. However, the only ***utility*** based parameters are those relating to the health states; HS_1 and HS_0. Preferences for the duration of the health improvement, T_D , may be accounted for if a preference based discount factor is used. Preferences related to the probability of successful outcome, p , could be accounted for if standard gamble were used as the elicitation-technique. A compound utility measure for the three terms (health states, time and probability) is offered by the HYE (Mehrez & Gafni, 1989).

Several techniques are available to estimate utility weights corresponding to H_0 and H_1, such as standard gamble, rating scale and person trade-off (Torrance, 1986). However, Time Trade-Off (TTO) has arguably been the most widely used, and has assumed the role of a *de facto* 'gold standard' (Drummond et al, 1997). The TTO technique offers respondents a choice between a fixed length of time in a 'poor health' state versus varying, shorter, times in 'good health'. Respondents then trade off the amount of time they would be prepared to give up in the poor

[1] Note that although there are theoretically states which receive a weight less than 0 (and, in theory, above 1), in practice this is extremely rare - most utility estimation exercises are formulated to restrict weights to being in the 0 to 1 range (Smith & Dobson, 1993).

health state to be in the good health state. That is, they are offered a choice, for example, of 10 years in poor health verses 9, 8, 7 etc years in good health. At the point at which they become indifferent to the choice, the ratio of the time in the poor health state to the number of years forgone provides the utility value of the poor health state. For instance, if the respondent would be willing to forgo 5 years in good health to avoid 10 years in poor health, then the utility weight would be 0.5 (5/10). This weight is then multiplied by the number of years in that state to obtain the number of QALYs. In essence these preferences therefore act as an 'exchange rate' between quality and length of life (Richardson, 1994).

In contrast, CBA expresses the value of the benefit resulting from a program in monetary terms. There are several means by which this monetary value may be obtained, but consensus opinion is that 'willingness-to-pay' (WTP) is the preferred method, having a strong basis in the theory of a constrained utility maximisation (Johansson, 1995), and being founded on principles of modern welfare economics (Boadway, 1974; Boadway & Bruce, 1984). This theory states that the decision concerning whether to implement a program or not should be made with reference to the preferences of those likely to be affected by the intervention, and that these values are best summarised in terms of the amount each individual would be willing to pay for the change brought about by the intervention (Drummond et al, 1997). This is an approximation of what would occur under a competitive market, and therefore leads to the optimal provision of the public good. This is because, as long as the gainers can *potentially* compensate the losers, net societal benefit will be greater than net social cost, and thus society overall experiences an increase in welfare. WTP is typically assessed by means of direct survey of such values using a hypothetical market construct, termed 'contingent valuation'.

Contingent valuation involves asking individuals directly in a survey the maximum amount they are prepared to pay to have the commodity in question, or the minimum amount they would accept in compensation to be deprived of it. This technique has been subject to much development in economics over the last 30 years, and is a standard means to estimate benefit in the areas of transport and environmental economics (Olsen et al, 1999).

Although used less frequently in valuing health care programs, recent literature on economic evaluation of health care has shown increasing interest in the use of willingness-to-pay (WTP) as a measure of health benefits (see Appendix 1). It is argued that WTP is 'superior' to QALYs in assessing beenefits for three primary reasons. First, that WTP is the 'theoretically correct' approach, because of its foundation in welfare economics. Second, that WTP imposes no restrictions as to which attributes of a programme people are allowed to value, as opposed to QALYs where only *health states* are being valued. Third, that benefits are valued in the same unit as costs which is required for advising decision makers on improvements in *allocative* efficiency. The author has considered these arguments in detail elsewhere (Olsen & Smith, 1999).

However, although both forms of evaluation have much in common (as they are both individual preference weighted approaches to the valuation of health outcomes), there has been no clear theoretical, or empirical, conclusion concerning their relationship (Bala et al, 1998). Although some authors consider the two techniques, in theory, to have their foundations within welfare

economics, and thus share the same conceptual underpinnings, others view them as deriving from, or responding to, fundamentally different constructs (Pauly, 1995; Sugden & Williams, 1978). However, it is not the purpose of this paper to comment upon these conceptual issues, but *rather to consider the relationship which may exist on empirically. In particular, to assess the relative sensitivity* of each measure to changes in health status, and to assess the degree of *compatibility* of the two techniques in providing a relative valuation over different health states.

3 Methods

3.1 Health States Valued

The health states valued in this exercise were those valued for the construction of the Assessment of Quality of Life (AQoL) instrument, developed by the Health Economics Unit of Monash University (Hawthorne & Richardson, 1995; Hawthorne et al, 1996, 1997a, 1997b). This is a multi-attribute utility (MAU) instrument, which comprises 5 dimensions of 5 items each (ie 15 items), each with 4 levels (from A to D, corresponding to best through worst health), as illustrated in Table 1. Health states are derived from combinations of these dimensions/items.

The AQoL has a unique nested structure in which each of the five structurally *independent* health dimensions are composed of three *interdependent* items. This structure increases sensitivity while preserving dimension independence. The items within each dimension form a single factor as determined by factor analysis, and each of the five dimensions is orthogonal to the other four. Given that all items load on the principle component analysis factor a single quality of life score can be determined. The reliability of the instrument (assessed through internal consistency) is good (Cronbach's standardised α = 0.8). Confirmatory factor analysis indicates that the hypothesised structure explains virtually all of the interrelationship between variables. The descriptive system of the AQoL is therefore unsurpassed in the construction of MAU instruments. The utility weights attached to the possible health states derived from the instrument were achieved using the TTO technique, and recent work has shown the reliability and validity of these utility weights (Hawthorne et al, 1997b). Further information concerning the development of these items, and the TTO weighting technique, can be found in Hawthorne et al (1997b).

Table 1: AQoL Dimensions and Items

Dimension	Item
Illness	Use of prescribed medicines Reliance on medical aids Receiving regular medical treatment
Independent living	Household tasks Self-care Mobility
Social relationships	Relationships with others social isolation Family role
Physical senses	Seeing Hearing Communication
Psychological well-being	Sleep Anxiety and depression Pain

These health states were chosen for two reasons. First, it provided a convenient sample of individuals who had previous experience of valuing these states, and who had provided TTO values for them. This enabled the WTP survey to be both more extensive than if it had also included the TTO valuation exercise, as well as remaining fairly short and simple to understand. Second, using this sample, who had previously given TTO values, meant that there was no concern over the ordering of questions; that the TTO exercise may bias the subsequent WTP values or vice versa. The WTP interviews were conducted at approximately 2 months subsequent to the TTO interviews, and it is reasonable to assume that individual circumstances would not have changed significantly during that period. Thus, there is little concern for potential bias introduced by having the two sets of valuations distanced by time.

From the 1.1 billion possible scenarios (combinations of items) available from the AQoL, 18 were chosen for this exercise. These were chosen to enable assessment at four levels: (i) small changes in quality of life along specific items within a dimension; (ii) large changes in quality of life along specific items with a dimension; (iii) changes in quality of life across dimensions; and (iv) an extreme change in quality of life from the worst state possible within AQoL to 'good health'[2]. These four areas, and the scenarios which relate to them, are outlined in table 2.

[2] The "Good Health" classification refers to either the best possible health attainable within that item (questions 1 to 6), or a composite description of Good Health (remaining questions) used when more than one item/dimension is being valued.

Table 2: Breakdown of Scenarios Used

Assessment	Question	Dimension	Item(s)	Severity[1]
(i)	1	Illness	2	C
(i)	2	Independent Living	5	C
(i)	3	Social Relationships	7	C
(i)	4	Physical Wellbeing	12	B
(i)	5	Psych. Wellbeing	14	B
(i)	6	Pain	15	B
(ii)	7	Illness	2	D
(ii)	8	Independent Living	5	D
(ii)	9	Social Relationships	7	D
(ii)	10	Physical senses	12	D
(ii)	11	Psych. Wellbeing	14	D
(ii)	12	Pain	15	D
(iii)	13	Illness	1-3	D
(iii)	14	Independent Living	4-6	D
(iii)	15	Social Relationships	7-9	D
(iii)	16	Physical senses	10-12	D
(iii)	17	Psych. Wellbeing	13-15	D
(iv)	22	ALL	ALL	D

[1] Severity is classified according to a scale of A to D, where A is the least severe, and D the most severe.

As one can see from Table 2, questions 1 to 6 are all valuations along one specific item within a dimension, split such that there are 3 each of small change from either 'B' (which may be considered a 'mild' poor health state along that dimension) or 'C' (a 'moderate' poor health state) to 'good health'. Questions 7 to 12 are within these same items, but are assessing the value of the worst state within that item (a level of 'D'). These questions allow assessment of the relative value of movements between magnitudes of 'poor health' within a specific dimension, as well as assessment of the relative value between different dimensions of quality of life. Questions 13 to 17 are composite valuations of all items within a dimension, and these have been set to the worst state. Again, this allows comparison between dimensions. The final question, 22, assesses the extent of the value of the largest possible change in health status from the 'all worst' to 'good health'.

This combination of assessments, from small to large changes, and within and across dimensions, enables assessment of the sensitivity of the two measures, TTO and WTP, to small and large changes in health status.

3.2 Sampling and Survey Design

Respondents were drawn from a convenience sample, comprising those who had previously completed the AQoL survey. This was for two reasons. First, as such respondents had completed the AQoL survey shortly before, there existed good demographic and background data which did not have to be reassessed. This meant that the questionnaire could concentrate on the valuation questions and avoid being overly burdensome on respondents. Second, as these subjects had completed the same exercise valuing the **same** health states in terms of TTO there already existed a set of TTO values for these health states, thus providing opportunity to conduct a comprehensive, yet manageable WTP interview.

Another consideration was that using a sample drawn from the initial AQoL survey made it more likely that respondents would be a representative cross-section of the Australian population. The original AQoL sample was recruited through random selection from Victorian electoral rolls within census collector divisions, stratifying for SES (Hawthorne et al, 1996, 1997a). Although the sample used in this study is therefore one of 'convenience', it is expected to be more representative than a traditional convenience sample as subjects were drawn from this prior randomised stratified sample.

Prior to administration of the survey, the instrument was tested on a random sample of 10 subjects taken from the AQoL sample (who were not interviewed during the main survey), who were subjected to additional questioning in a 'debriefing' session after the interview. This was undertaken to ensure that respondents understood the questions as posed, they understood the implications of their WTP values as given, and to take account and address any ambiguities or problems which were felt to exist in the survey instrument. From these session it was clear that respondents had little difficulty in understanding the questions, although there were some minor ambiguities in the questionnaire which were addressed.

3.3 Interview Procedure

All interviews were conducted by trained interviewers (who had also administered the AQoL survey) at times and places suitable for subjects. Subjects were informed of the nature of the survey, both at initial contact via telephone to establish their willingness to participate, and also at the time of interview. They were then asked to complete a consent form prior to the interview.

Although interviewers were trained, and a 'pilot' study had been conducted, it was possible that the WTP questions would not be readily understood by respondents. Because of this, audio tapes of interviews were collected and monitored to ensure that no undue 'prompting' by the interviewer was occurring (ie interviewer bias), that respondents appeared to understand the questions, and also to help in assessment of the feasibility of such an exercise.

3.4 Sample Size

As this study was required to sample those who had previously valued the **same** health states, the sample size was dictated by the number who had previously done so. Thus the initial sample size was the total number of respondents available who met this criteria, which was 50.

3.5 WTP Estimation

Respondent WTP values were derived using an 'open-ended' CV survey, involving asking respondents directly the maximum amount they would be willing-to-pay to move from one health state to another.

The form of the questionnaire was to present the respondent with two health states, and ask them to express their maximum willingness to pay to ensure they would be in 'good health' state rather than the 'poor health' state over a period of one year. ie a WTP to avoid being in the poor health state. Respondents were told that each health state (good or poor health) was to last for a period of 1 year, after which time they would revert to 'good health'. Thus, their WTP was expressed as a yearly figure. This period was used to enable direct comparison with utility weights, which also use a single year as the unit of time involved.

The questionnaire comprised four sections. First respondents were reminded of the previous AQoL survey they had completed and how they were asked to choose between different health states and express the time they would be prepared to give up to avoid a health state. They were then told that this exercise would be exactly the same except that rather than giving up time to avoid a health state they would be required to give up personal income (ie money). The questions were deliberately kept as close as possible to the AQoL survey to yield comparative data.

Second, in order to try to remove any possible 'contamination' of the health state valuation according to the means by which that improvement in health state would be brought about (ie mechanism by which they could avoid a health state) they were asked not to consider that. If this proved impossible, the interviewer was authorised to explain that they may assume it would be a pharmaceutical treatment which would be quick, painless and involve no side-effects. Responds were finally reminded that there would be no right or wrong answers, only their opinion.

Third, prior to the valuation questions all health states were shuffled such that the order with which respondents valued them was randomised. This was to be used to assess any degree of 'question order' effect. Valuation questions all followed the same format: the respondent was presented with a pair of health states and asked which they preferred (those preferring the poor health state would be considered 'irrational', although in the event all chose 'good health' as the preferred state). In each case they were asked to assume they were in health state A (the poor health state), and asked their **maximum** willingness-to-pay to secure a change from this state A to the other state, state B (good health), presented. The WTP figure was to be thought of as being paid out-of-pocket (OOP), and expressed as either fortnightly (the traditional pay period for

most people in Australia), monthly or annually as respondents felt comfortable. These values where then calculated as annual figures and given back to respondents for verification. Respondents were reminded to consider how much they could afford to pay when giving their response. As shown in table 2, respondents were asked 18 questions concerning their WTP for such trade-offs.

In terms of payment 'vehicle' (ie use of OOP), it was felt that alternatives, such as insurance based payment was not appropriate for this comparison. Although WTP can be assessed in either a 'point-of-use' or 'insurance' context, TTO weights are only able to be assessed in a point-of-use' context. That is, what they are willing to pay (in terms of forgone years of life) for a change in health status ***assuming they have the disease which will place them in the poor health state.*** Thus, whilst it may be desirable in a societal decision making context to assess WTP in an insurance based context (O'Brien & Gafni, 1996), given the objective of this paper in comparing the two preference-based measurement techniques, it was necessary to express both in a 'point-of-use' context.

Questions were posed in an open-ended format[3], utilising a prompt if necessary. The prompt is a version of one used by Jones-Lee (1989; Jones-Lee et al, 1985) and the author (Smith et al, 1993) previously. The specific WTP question and prompt is provided in Appendix 2. The prompt amount was varied to enable testing for 'starting point' bias, and was only used when the respondent fails to provide any response without it for a period of approximately 30 seconds. Respondents who provided a valuation of zero were prompted to clarify whether they simply were not willing to pay anything, that they did not actually value the change in health state (ie are indifferent between states A and B), or there is another reason - such as income constraint, ethical objections to such valuation, feel that this payment may become policy, etc. Respondents also ranked health states to enable a consistency check to be applied to the valuations.

Fourth, once the WTP questions relating to the change in health state were complete, the respondent completed several questions concerning the value of one year of life in 'good health'. This was to provide a comparison with the WTP values provided to avoid the health states mentioned. Respondents also at this stage provided a ranking of all the 'poor health' states, again to be used as a consistency check.

Finally, of course, the respondents were assured anonymity, and the purpose of the research outlined, to ensure full cooperation from the respondent.

[3] There is currently a debate concerning the relative merits of different questionnaire formats, particularly the used of closed ended (discrete, referendum, dichotomous or binary choice) questionnaires and variations in open-ended formats, such as bidding, payment cards and use of prompts (Smith, 1997). The author does not wish to become embroiled in this debate at this juncture, suffice to raise the issue as one for consideration.

3.6 TTO Estimation

The utility weights were derived using the TTO method. A full account of this process is reported in Hawthorne et al (1997b).

3.7 Demographic Influence

Estimation of the impact of respondents' demographic characteristics upon the TTO/ and WTP values was assessed. However, this demographic information was not collected in this survey, as respondents, as mentioned, were those who had already previously completed the AQoL survey. This data was therefore already available from that survey. Variables of interest in this analysis were: age/life expectancy, income, gender, current health status, marital status, and the importance of 'faith'.

3.8 Statistical Analyses

As data was heavily skewed, for both WTP and TTO values, non-parametric statistics were used. A visual representation of the spread, skewness and outlying data points is also presented as a boxplot (Hoaglin et al, 1983). This is a graphical means of presentation which is based on the quartile range of data, and is therefore less influenced by the distribution than, for example, the variance, which is a more traditional means of assessment (Hoaglin et al, 1983). For statistical estimation the Wilcoxon paired comparison test was used, which assesses differences between two related sample's. Associations between WTP and TTO values were analysed using the Pearson product-moment correlation.

All analyses were conducted using SPSS for Windows, version 6.1.

4 Results

4.1 Sample Size, Response Rate and Characteristics

Between 5/8/96 and 2/6/97 47 interviews were completed of the 50 individuals approached, a recruitment response rate of 94%. This was undoubtedly high due to the nature of subjects having completed the previous TTO interview. All 47 provided useable responses. However, 9 of those undertaking the TTO exercise had provided zero responses (ie no trade) for ethical or protest reasons (ie they would not engage in the TTO exercise, even though they clearly preferred 'good health' to 'poor health'). These were excluded from the estimation of mean TTO scores as reported. In contrast only 3 refused to participate in the WTP survey, and again these are excluded from the mean WTP value estimation (these three also refused to undertake the TTO survey). This issue is returned to in the discussion.

Demographic and background details are reported in table 3.

Table 3: Demographic Details

Variable	Value
Age	43 (±1.9)
Life expectancy	37 (±2)
Income	$46,170 (±$4,013)
Gender	47% male
Current Health Status	89% ≥ 'average'
Marital status	83% married/defacto
Time to complete interview	65 minutes (±2.8)
Importance of 'faith'	49% ≥ important

4.2 WTP in Dollar Values

Table 4 below provides a comprehensive breakdown of WTP values provided to each question posed.

Table 4: WTP Dollar Values

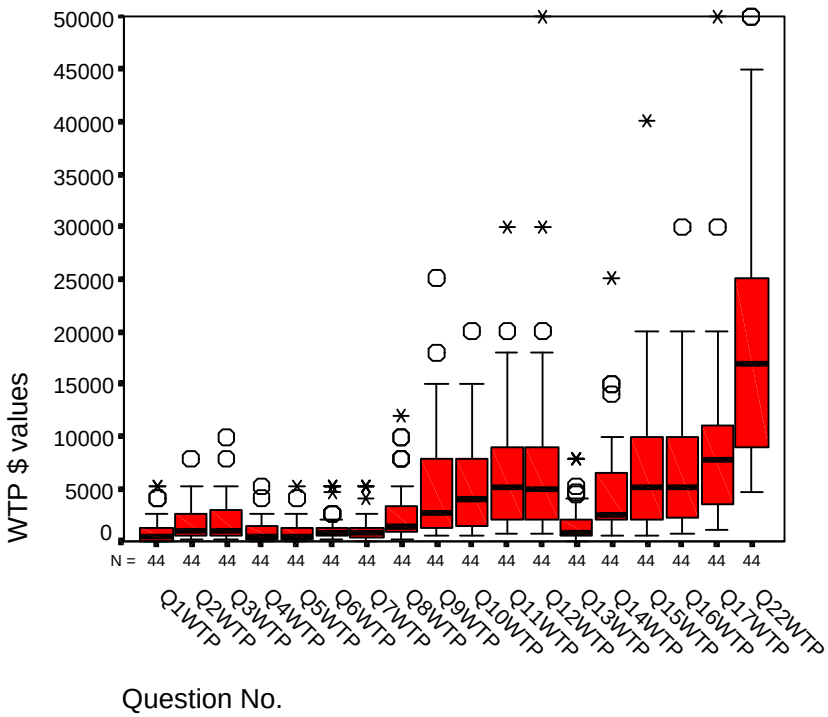
Qst	Mean (S.E)		Median	Mode	Std dev	Min	Max
Q1	1,020	(216)	500	0	1,435	0	5,200
Q2	1,498	(231)	1,040	520	1,531	130	7,800
Q3	2,164	(330)	1,040	520	2,191	130	10,000
Q4	1,089	(179)	520	0	1,189	0	5,200
Q5	1,007	(175)	520	520	1,158	0	5,200
Q6	1,359	(229)	780	520	1,517	130	5,200
Q7	1,303	(228)	780	520	1,511	0	5,200
Q8	2,831	(447)	1,430	1,040	2,966	260	12,000
Q9	4,845	(770)	2,800	7,800	5,106	520	25,000
Q10	4,903	(664)	3,940	5,200	4,406	520	20,000
Q11	6,358	(897)	5,200	2,080	5,951	780	30,000
Q12	7,601	(1,355)	4,940	2,080	8,989	780	50,000
Q13	1,878	(343)	910	520[a]	2,280	0	7,800
Q14	4,789	(747)	2,490	2,080	4,954	520	25,000
Q15	7,607	(1,122)	5,200	2,080[a]	7,443	520	40,000
Q16	7,150	(932)	5,200	5,200	6,182	780	30,000
Q17	9,417	(1,360)	7,800	7,800	9,023	1,200	50,000
Q22	20,140	(2,270)	17,000	20,000	15,060	4,600	80,000

N=44, as 3 refused to make any payment for ethical/protest reasons.
[a]Multiple modes exist. The smallest value is shown.

The large differences shown between mean and median values indicates a skewed distribution, whilst large standard errors and deviations a wide spread of responses. As with most WTP surveys, a few high values are generating higher mean values overall for each question. However, there appears to be a considerable degree of consistency across median and mean values, and minimum and maximum values across question 'groups'.

A graphical representation of the levels of precision and dispersion can also be gained from the boxplot diagram presented in figure 1.

Figure 1: **Boxplot of WTP $ Values**



From figure 1 we can see that there is a far lower level of dispersion for questions 1 through 7, and it is only where the states become far worse that such large dispersions begin to be found. One possible explanation for this is that the respondents were only willing to pay small amounts for changes of a minor nature, and at this level was an amount that all could afford. However, many may wish to pay far more for greater improvements, ie express greater WTP for the worse health states, but are prevented from doing so. That is, it might be that as the health state worsens the income constraint becomes more deterministic of the valuation provided.

However notwithstanding this, both mean and median values show similar trends.

First, values do not vary greatly for questions 1 through 6, which are valuations of single items at levels of B (4, 5 and 6) or C (1, 2 and 3). Interestingly it is the dimension of social relationship which receives the highest mean WTP to avoid, although both this and independent living demonstrate equivalent median WTP.

Second, and as, *a priori*, expected the values for questions 7 through 12, which are the same items as questions 1 through 6 but at a worse health state level of D, are all significantly higher, although this is more pronounced for questions 10, 11 and 12, than for questions 7, 8 and 9, which again is to be expected as the differences in 'quality of life' here are greater (D versus B, as opposed to D versus B).

Third, and perhaps most interestingly, adding more items to dimensions, as is the case with questions 13 through 17, does not add substantially to the WTP values given. That is, there appears to be a greater increase in WTP value from moving from a higher level of quality of life to a lower one (eg B to D) within an item than there does from addition of more items to a dimension.

Fourth, over higher levels of quality of life, in questions 1 to 6, there does not appear to be a great distinction between dimensions of quality of life, with values for questions 1, 2, 4, 5 and 6 virtually identical in mean and median terms. There appears to be little sensitivity at these levels.

4.3 WTP Expressed as a Proportion Of Income

In order to 'adjust' for differences in income, the WTP values were expressed as a proportion of each individuals annual income. Results for this are presented in table 5 below.

Table 5: WTP Expressed as a Proportion of Income

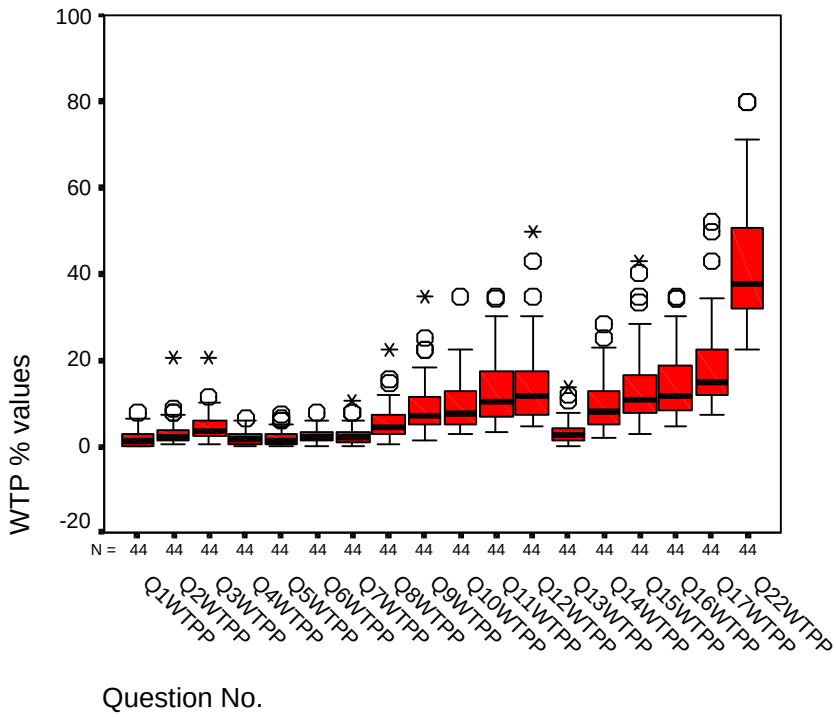
Qst	Mean (S.E)		Median	Mode	Std dev	Min	Max
Q1	1.83	0.31	1.23	0.00	2.05	0.00	8.00
Q2	3.39	0.52	2.31	0.87	3.45	0.37	21.00
Q3	4.43	0.56	3.47	3.47	3.68	0.74	21.00
Q4	1.95	0.26	1.60	0.00	1.73	0.00	6.40
Q5	1.93	0.28	1.40	0.00	1.88	0.00	7.40
Q6	2.52	0.25	2.08	0.87	1.63	0.24	8.00
Q7	2.62	0.34	2.08	0.00	2.27	0.00	10.40
Q8	5.78	0.67	4.62	3.47	4.41	0.74	22.30
Q9	9.43	1.03	7.18	3.47	6.86	1.42	34.70
Q10	9.91	0.97	7.62	5.20	6.42	2.84	34.70
Q11	12.81	1.19	10.20	6.90	7.90	3.47	34.70
Q12	14.51	1.50	11.60	8.32	10.03	4.62	50.00
Q13	3.65	0.48	2.50	3.47	3.19	0.00	13.87
Q14	9.60	0.97	7.90	5.20	6.42	1.90	28.60
Q15	14.10	1.46	10.90	6.90	9.70	2.97	42.86
Q16	14.30	1.21	11.56	6.90	8.03	4.62	34.67
Q17	18.81	1.60	15.12	17.33	10.05	7.56	52.00
Q22	42.71	2.16	37.86	34.67	14.31	22.22	80.00

N=44, as 3 refused to make any payment for ethical/protest reasons.

These data reflect a similar pattern to the dollar amounts presented above. However, it is of interest to note that the worse the quality of life becomes the higher the WTP as a proportion of income becomes up to a very high level of over 40% to avoid the worst state (question 22).

Again, a visual representation of dispersion and precision is provided by a boxplot in figure 2.

Figure 2: Boxplot for % WTP



4.4 TTO Values

Table 6 provides a similar breakdown of TTO values by question, both in terms of mean and median values, with associated measures of precision. As indicated above, the population sample for these questions was 38, with 9 refusing to make any trade-off at all for ethical/protest reasons.

Table 6: TTO Values

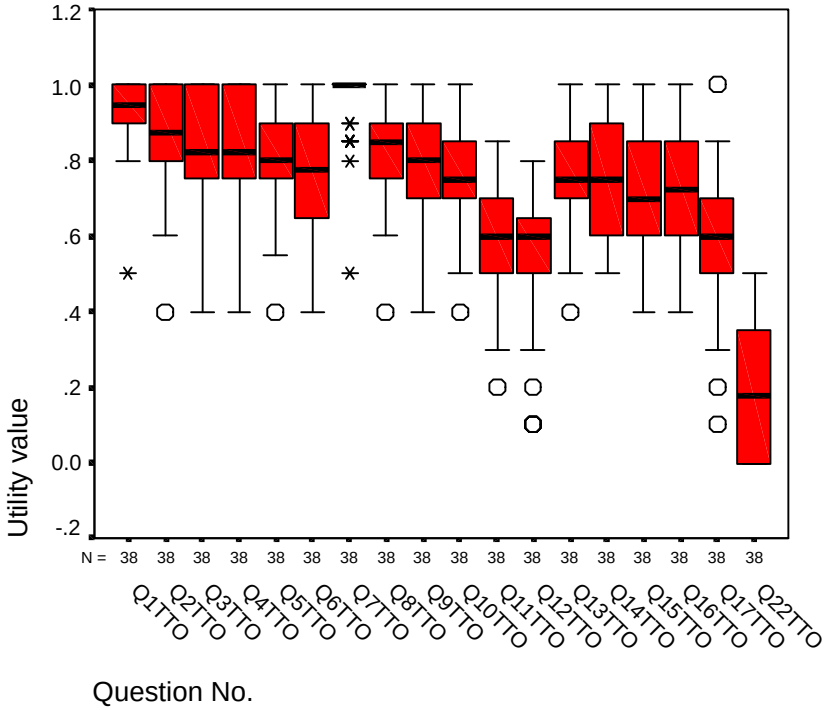
Qst	Mean	Median	Mode	Std dev	Std error	Min	Max
Q1	0.94	0.95	1.00	9.437E-02	1.531E-02	0.50	1.00
Q2	0.87	0.88	1.00	0.14	2.208E-02	0.40	1.00
Q3	0.83	0.83	1.00	0.14	2.312E-02	0.40	1.00
Q4	0.83	0.83	1.00	0.14	2.312E-02	0.40	1.00
Q5	0.80	0.80	0.80	0.13	2.162E-02	0.40	1.00
Q6	0.76	0.78	0.80	0.16	2.666E-02	0.40	1.00
Q7	0.96	1.00	1.00	9.738E-02	1.580E-02	0.50	1.00
Q8	0.83	0.85	1.00	0.13	2.164E-02	0.40	1.00
Q9	0.78	0.80	0.80	0.14	2.323E-02	0.40	1.00
Q10	0.75	0.75	0.70	0.15	2.454E-02	0.40	1.00
Q11	0.59	0.60	0.50[a]	0.15	2.474E-02	0.20	0.85
Q12	0.53	0.60	0.50[a]	0.20	3.163E-02	0.10	0.80
Q13	0.75	0.75	0.70	0.15	2.454E-02	0.40	1.00
Q14	0.76	0.75	1.00	0.18	2.840E-02	0.50	1.00
Q15	0.72	0.70	0.50	0.17	2.780E-02	0.40	1.00
Q16	0.74	0.73	1.00	0.18	2.896E-02	0.40	1.00
Q17	0.60	0.60	0.60	0.19	2.979E-02	0.10	1.00
Q22	0.18	0.18	0.00	0.17	2.750E-02	0.00	0.50

n=38
[a]Multiple modes exist. The smallest value is shown.

The mean and median TTO values are almost (and in many cases are) identical, which is a marked difference to the WTP values as presented above, indicating a less skewed distribution. They also appear to be somewhat dispersed in that the range of values covers, in all cases, more that 0.5 on the 0-1 scale. The values also appear to provide a similar pattern concerning the similarities between questions 1 to 6. However, they appear to be much more closely related across questions 13 to 17.

In terms of precision and dispersion, again the boxplot in figure 3 illustrates this graphically.

Figure 3: **Boxplot TTO Values**



Here we can see that values are quite dispersed within each question, although mean values are remarkably similar across questions.

4.5 Sensitivity of TTO and WTP to Changes in Health State

A priori, it was hypothesized that TTO values would not be as sensitive to small changes in health status as WTP, but that both would perform well when considering large changes in health status. This hypothesis was tested by examining the data for statistically significant differences across questions asked, for both TTO and WTP values. This was achieved with two forms of comparison. First, assessment of differences in levels (scores) between items and dimensions at the same level of quality of life (along the A to D classification). Second, assessment of scores across different levels within the same item, such as differences between levels B and D within an item for pain.

4.5.1 TTO Results

As outlined in the methods section, the “Wilcoxon Signed Ranks Test” was used to assess the degree to which differences between variables were statistically significant. These analyses were undertaken in two areas, and the results for each are outlined.

4.5.1.1 Different Items, and Dimensions, Of The Same 'Health Status' Level

Here an assessment was made of the significance of differences in values given for different single *items* within a dimension, and for dimensions globally (ie all items) of the *same* nominal 'level' of health status. Questions 1 though 17 were used to assess this. Questions 1, 2 and 3 represent dimensions of illness, independent living and social relationships of a quality of life level 'C', questions 4, 5 and 6 represent dimensions of physical activity, psychological well-being and pain at a level of 'B'. Questions 7 to 12 represent the same items, within the same dimensions, but at a level of 'D' (where a level of A is equivalent to 'good health', and 'D' is the 'worst' state).

Questions 13 to 17 are combination health states representing **all** items within a dimension, and are set in this case to a level of 'D'. Note that pain is not included as it is incorporated in the dimension of psychological well-being. Finally, for a check of consistency, assessment was made for significant differences between the value provided for question 22, which is a composite 'all worst' health state, taking the status presented in level D of every item within all dimensions, against the 'all worst' within each dimensions separately (questions 13 to 17).

The Wilcoxon tests demonstrated that in all the above relationships the TTO values were significantly different at a level of P=0.001 (virtually all being significant to P=0.00001). The **exceptions** to this were as follows.

Table 7: Sensitivity of TTO Values Across Items

Relationship	Significance Level (P=X)
Q3 vs Q4 (Soc Rel 'C' vs Physical 'B')	1.000
Q3 vs Q5 (Soc Rel 'C' vs Psych 'B')	0.005
Q4 vs Q5 (Physical 'B' vs Psych 'B')	0.005
Q5 vs Q6 (Psych 'B' vs Pain 'B')	0.007
Q9 vs Q10 (Soc Rel 'D' vs Physical 'D')	0.020
Q13 vs Q14 (Illness vs Indep Liv)	0.98
Q13 vs Q15 (Illness vs Soc Rel)	0.15
Q13 vs Q16 (Illness vs Physical)	0.53
Q14 vs Q15 (Indep Liv vs Soc Rel)	0.005
Q14 vs Q16 (Indep Liv vs Physical)	0.22
Q15 vs Q16 (Soc Rel vs Physical)	0.103

Some clear patterns emerge from this.

First, the assessment of difference between questions 1 to 6. Interestingly, all three questions at level 'C' (questions 1 to 3) were significantly different from each other, showing a marked 'preference' for not only the level of health status, but also which dimension of health it represents. However, questions 3 and 4 (Soc Rel level C and Physical level B) are identical (p=1), implying that Soc Rel is not as important a dimension as the others at level C. In fact, questions 3, 4 and 5 are all not significantly different (at p=0.001) from each other, although they would be at a significance level of p=0.005. A more intuitive, although less statistically robust, interpretation would be that, apart from questions 3 and 4, all other questions are significantly different at this level, implying that the TTO method as used in this study provided responses which valued each item within a different dimension as significantly different to those in other dimensions at a different level of health status (eg 1 versus 5) and, crucially, at the same level of health status (eg 1 versus 2).

Second, assessment across questions 7 to 12, which represent items in different dimensions all assessed at a level of 'D'. Here again, at a significance level of p=0.001, only questions 9 and 10 (Soc Rel and Physical) would not be significantly different. This is consistent with the results above concerning these two variables.

Third, in terms of different dimensions of the same level (questions 13 to 17), there were fewer significant differences, although question 17 (Psych, including pain) was significantly different from any other dimension (question). Questions 13 to 16 did not really demonstrate significant differences, with 13 and 14 virtually identical (p=0.98), and other relationships not significantly different at p=0.1 (apart from 14 and 15). It would appear from this that, at a level of 'all worst' within each dimension, TTO values are not significantly differentiating between dimensions, apart from Psych (although this included pain which also was significantly different in the previous tests).

4.5.1.2 Different Levels Of The Same Item

The second area of assessment concerned different levels within the same dimension. This was derived by assessing question 1 against question 7, which represents a difference within an item, within the dimension of illness, of level 'C' (question 1) and level 'D' (question 7). This was repeated for questions 2 vs 8, 3 vs 9 (both 'C' vs 'D'), and 4 vs 10, 5 vs 11 and 6 vs 12 (all 'B' vs 'D'). The results are presented below.

Table 8: Sensitivity of TTO Values Within Items

Relationship	Significance Level (P=)
Q1 vs Q7 (Illness 'C' vs Illness 'D')	0.0033
Q2 vs Q8 (Indep Liv 'C' vs Indep Liv 'D')	0.0022
Q3 vs Q9 (Soc Rel 'C' vs Soc Rel 'D')	0.0015

At a level of p=0.005 all questions would have been shown to be significantly different from each other, and this is perhaps the most intuitive interpretation of these results, as 0.0015 is only just beyond the level of significance of p=0.001, and the other results are not too dissimilar. However, what is of interest is that the relationship which certainly ***are*** significantly different at 0.001, are those between questions 4, 5 and 6 and 10, 11 and 12, which represent levels of 'B' versus 'D' for each dimension. This is what would be expected and adds weight to the 'face' validity of these results.

In addition, as predicted, and providing some test of 'face validity', question 22 was found to be highly significantly different from all other questions.

Overall TTO, as applied in this study, appears to perform well in differentiating significantly between different *levels* of health within each dimension, but not so well in differentiating between different *dimensions* of health at the same nominal 'level' of health status.

4.5.2 WTP Results

The WTP results were assessed in the same manner as the TTO values above, both in terms of dollar values and as a proportion of income, although ultimately the results differ little and only in terms of marginal differences in the levels of significance of various relationships (usually slightly higher for those expressed as a proportion of income).

4.5.2.1 Different items, and Dimensions, Of The Same 'Health Status' Level

Again, the Wilcoxon tests demonstrated that in all comparative relationships the WTP values were significantly different at a level of P=0.001 (and again virtually all were significant to P=0.00001), with exceptions as follows.

Table 9: Sensitivity of WTP Values Across Items

Relationship	Significance Level (P=X)
Q1 vs Q4 (Illness 'C' vs Physical 'C')	0.600
Q1 vs Q5 (Illness 'C' vs Psych 'B')	0.700
Q1 vs Q6 (Illness 'C' vs Pain 'B')	0.006
Q2 vs Q4 (Indep Liv 'C' vs Physical 'B')	0.004
Q2 vs Q5 (Indep Liv 'C' vs Psych 'B')	0.002
Q2 vs Q6 (Indep Liv 'C' vs Pain 'B')	0.122
Q4 vs Q5 (Physical 'B' vs Psych 'B')	0.485
Q4 vs Q6 (Physical 'B' vs Pain 'B')	0.050
Q5 vs Q6 (Psych 'B' vs Pain 'B')	0.036
Q9 vs Q10 (Soc Rel vs Physical)	0.25
Q11 vs Q12 (Psych vs Pain)	0.004
Q15 vs Q16 (Soc Rel vs Physical)	0.706

Again, some clear patterns emerge from this.

First, in the assessment of difference between questions 1 to 6 there are a great many fewer significant differences than for TTO. Questions 1 and 2 are both insignificantly different from questions 4, 5 and 6, meaning that only question 3, representing level C of Soc Rel is significantly different from all other variables within this set. Interestingly, and somewhat counterintuitive, the level 'C' items (1, 2 and 3) are all significantly different from each other, yet not from other dimensions which are represented by level 'B' health status. Furthermore, questions 4, 5 and 6 are not significantly different from each other, and the p values for these relationships are far higher than those for TTO, representing stronger associations between them. Intuitively, from this data one might suggest that WTP is not adequately differentiating between the different items here at different levels.

Second, assessment across questions 7 to 12 shows that, at a significance level of p=0.001, questions 9 and 10 (Soc Rel and Physical), and 11 and 12 (Psych and Pain) would not be significantly different. Questions 9 and 10 showed this for TTO also, although at a lower p value. Questions 11 and 12 would be significantly different if p=0.005 was adopted as the cut of value, and again this might be intuitively appealing.

Third, in terms of different dimensions of the same level (questions 13 to 17), there was a substantial difference compared with the results for TTO. Here only one relationship was found not to be significantly different, questions 15 and 16 (Physical and Psych well-being). This relationship was one of those found not to be significant for TTO also, although the p value for WTP results is higher once again.

4.5.2.2 Different Levels Of The Same Item

Perhaps the most significant result of the WTP assessment, compared with TTO, is that **all** differences between different levels within the same dimension were found to be highly significant. That is, question 1 versus 7, 2 vs 8, through to 6 vs 12 were all significantly different from each other.

In addition, as predicted, and providing some test of 'face validity', question 22 was found to be significantly different from all other questions.

Overall, WTP, compared with TTO as applied in this study, appears to perform far *less* well in differentiating significantly between different *dimensions* of health at the same nominal 'level' of health status, but performs better at differentiating between different *levels* of health within each dimension.

4.6 Correlation Between TTO and WTP

As well as assessing the relative sensitivity of TTO and WTP to changes in health status, the data was also examined for any systematic relationship between TTO and WTP values for each question. It might be assumed, *a priori*, that since both are seeking to place a valuation on a change in health status, they might demonstrate some systematic relationship, leading to, for example, a similar ranking of the health states, or a trend such that the worse a health state is valued in terms of TTO (ie more someone willing to give up time to avoid it) then the higher the WTP would be for that health state (ie the higher the WTP dollars to avoid it), and thus that WTP would show an upward trend as TTO values decline. Both these forms of assessment are presented here.

4.6.1 Ranking of Health States by TTO and WTP

Tables 10 and 11 provide an assessment of the ranking of states according to mean and median TTO, WTP and WTP as a proportion of income.

Table 10: Ranking of Questions by TTO, WTP$ and WTP%

QUESTION	RANKING*					
	TTO		WTP$		WTP%	
	Mean	Median	Mean	Median	Mean	Median
1	1	2	2	1	1	1
2	4	3	6	7	6	6
3	5	5	8	7	8	8
4	5	5	3	2	3	3
5	8	7	1	2	2	2
6	12	9	5	4	4	4
7	2	1	4	4	5	4
8	5	4	9	9	9	9
9	9	7	11	11	10	10
10	11	10	12	12	12	11
11	15	15	13	14	13	13
12	16	15	15	13	16	16
13	3	10	7	6	7	7
14	9	10	10	10	11	12
15	13	14	16	14	14	14
16	14	13	14	14	15	15
17	16	15	17	17	17	17
22	18	18	18	18	18	18

* where a rank of 1 indicates highest quality of life/lowest WTP to avoid.

Table 11: Rankings Presented as Questions Corresponding to Rank Number

Rank	TTO		WTP$		WTP%	
	Mean	Median	Mean	Median	Mean	Median
1	1	7	5	1	1	1
2	7	1	1	5	5	5
3	13	2	4	4	4	4
4	2	8	7	7	6	6
5	8	3	6	6	7	7
6	3	4	2	13	2	2
7	4	5	13	3	13	13
8	5	9	3	2	3	3
9	9	6	8	8	8	8
10	14	14	14	14	9	9
11	10	10	9	9	14	10
12	6	13	10	10	10	14
13	15	16	11	12	11	11
14	16	15	16	11	15	15
15	11	11	12	16	16	16
16	17	17	15	15	12	12
17	12	12	17	17	17	17
18	22	22	22	22	22	22

There are a few interesting features which emerge from this.

First, only WTP% has any substantial degree of correlation between rankings based on mean and median values - all except those ranked 11 and 12, which are reversed (questions 14 and 10 for mean, and 10 and 14 for median). In contrast, TTO values have the lowest correlation, with only those questions ranked 10, 11, and 15 through 18 being the same. All other rankings are differ. WTP$ has 11 out of 18 rankings the same, and thus performs somewhere between the other two in terms of equality of ranking based on mean or median values.

Second, there is a poor correlation of rankings between TTO and either WTP measure. Apart from the worst state, question 22, which is consistent across all, only 2 rankings from TTO are consistent across WTP$ (both mean figures), and 3 across WTP% (2 median figures, 1 mean).

However, allowing for differences of 1 ranking, the rankings are reasonably similar in terms of which questions are contained within the *quartiles* of the ranking (taking the lowest rank, 18, away as that is consistent across all techniques). WTP$ and WTP% are, as might be expected, the

closest in rankings, with very similar questions contained within each quartile of the rankings (as indicated in the split in the questions in table 11 above). TTO, in contrast, is still not well correlated, with only 2 out of 4 questions ever matching across quartiles with either WTP$ or WTP% (based either on mean or median values), except for the bottom quartile when 2 of 5 questions match for mean WTP$, and 5 of 5 for WTP% mean and median. It would appear, then, that rankings are more consistent across the bottom 25% of the health states surveyed, than the top 75%. The implication of this is that there is a great deal more stability in the valuation of those states considered to be the 'worst', and more ambiguity in the relative rating of states considered to be of a 'reasonable' quality of life.

Finally, an implication of this is that the dimensions of health rated as 'important', or not important, differ according to assessment by TTO or WTP. For example, for TTO illness is a state considered to be of reasonable importance, one which people will not trade many life years to avoid, whereas with WTP ($ and %) psychological well-being is valued less importantly and people willing to give up little to avoid it.

4.6.2 Scatterplot of TTO Versus WTP

One means of visually assessing correlation and trends within data is to plot the raw data in a scatterplot. Following the *a priori* assumption as outlined earlier in this section, one would expect such a plot to present a trend for WTP to increase as TTO values fall. That is, for respondents to present a higher WTP to avoid a state which has a lower quality of life as expressed by TTO (which is, after all, simply expressing a greater willing to pay in terms of time given up to avoid the health state in question).

This relationship is assessed for both WTP$ and WTP% in the charts below. These plots contain all TTO and WTP values for each question.

Figure 4: **Scatterplot for TTO vs WTP$**

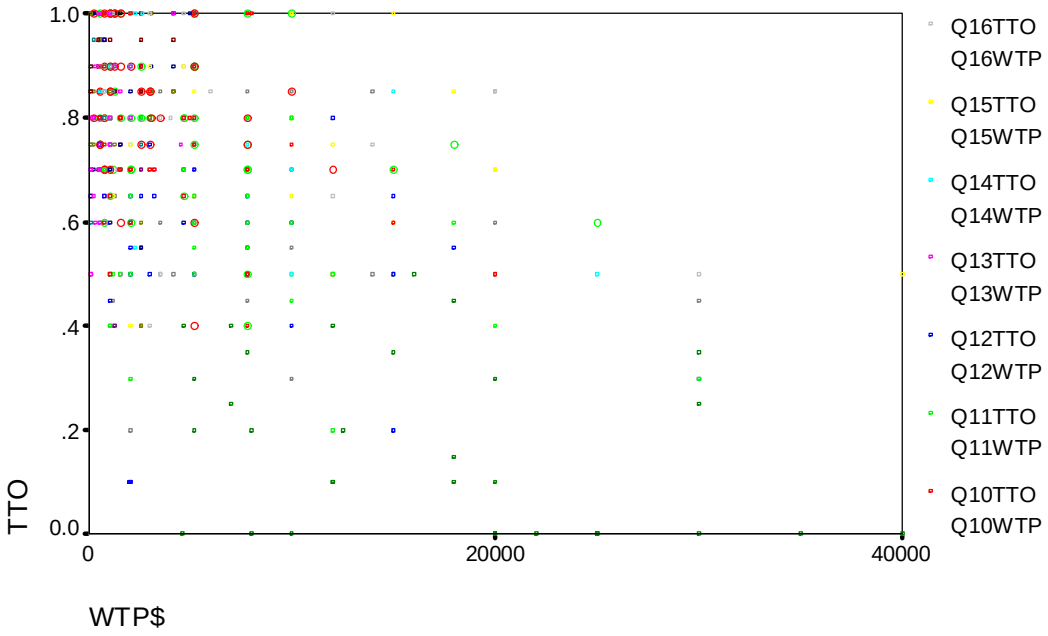
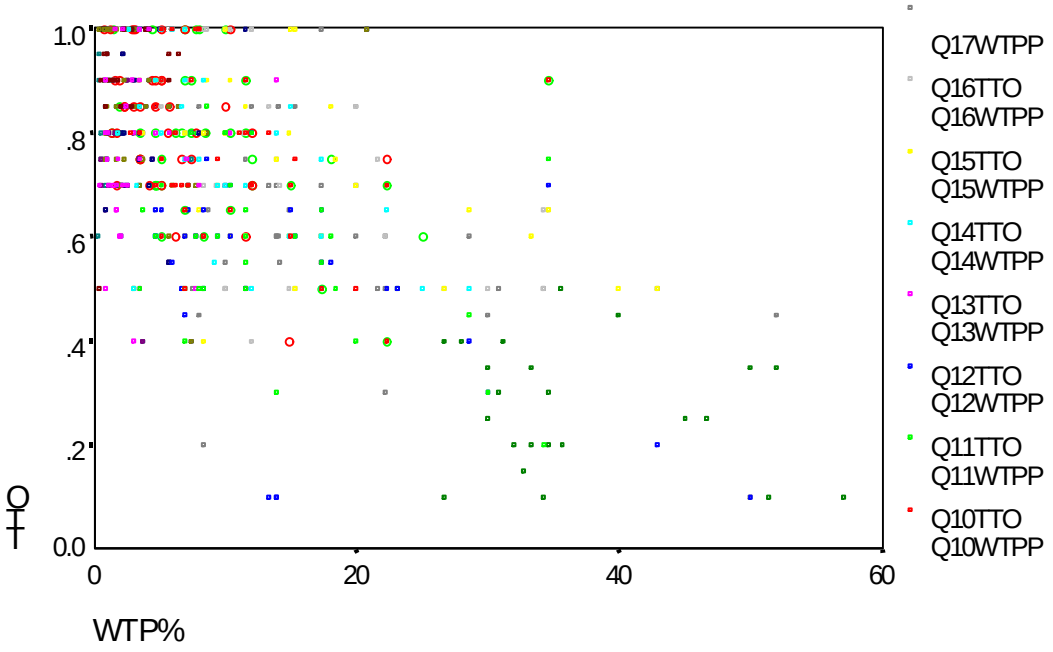


Figure 5: **Scatterplot TTO Versus WTP%**



Both charts provide a similar pattern of most values clustered within the top left hand corner, of high TTO values and low WTP values - consistent with the a prior hypothesis. Figure 3 shows no indication of any trend, with values tightly clustered in the upper left quadrant, although figure 4 shows some consistency with the hypothesised trend from upper left to lower right, but again the majority of values are clustered in the upper left quadrant. However, although there is no real trend consistent with that hypothesised, there is none which is **contrary** to it either (ie the data does not trend in the opposite direction).

From these plots, it appears as if the WTP may take a variety of values for any given TTO value provided, and there is little evidence of a significant trend for WTP values to increase as TTO values fall. From this there appears to be little relationship or convergence between the two preference-based measures of TTO and WTP.

This was further tested using Pearson correlation's. The results of this are presented in table 12.

Table 12: Pearson Correlation's Between TTO Versus WTP $ and WTP %

Question	WTP$ Value	WTP% value
1	+0.1	+0.03
2	-0.28	-0.16
3	+0.33	+0.07
4	+0.03	-0.19
5	-0.14	-0.28
6	+0.05	-0.11
7	+0.95	+0.02
8	-0.27	-0.44**
9	-0.23	-0.29
10	-0.27	-0.27
11	-0.35*	-0.41*
12	-0.36*	-0.49**
13	-0.18	-0.21
14	-0.37*	-0.49*
15	-0.16	-0.37*
16	-0.20	-0.45**
17	-0.39*	-0.42*
22	-0.24	-0.59**

* Significant at p=0.05

** Significant at p=0.005

The strongest associations here are between TTO and WTP%, which again may reflect the imposition of a 'tighter' scale and adjustment for differences in income making the instrument more similar to the TTO approach. In terms of questions, none of the first section, questions 1 to 6, appear correlated on either scale. Within the second section, questions 11 and 12 are correlated on both WTP scales with TTO, as are questions 14 and 17. It would appear from this, that there is greater correlation as health states deteriorate, than for more minor changes in health status, represented by the early questions. Although the majority of correlation's have predicted signs (negative, representing a higher TTO but lower WTP to avoid that state), there are some anomalies to this. The highest (r=-0.59), and most significant (p=0.001), correlation is for WTP% and TTO.

5 Discussion

The research presented in this paper examines the relationship between WTP and TTO as techniques for valuing changes in health status, using a single sample of respondents over a short time period. Both are similar in the extent to which they are survey's designed to ask people what they would be willing to pay from a capital stock (of time in case of the TTO, and money in the case of WTP) to purchase an increase in health status. However, in agreement with Bala et al (1998), it is clear from the data presented that "these value contracts are not similar and cannot be interpreted as equivalent methods". Interestingly, this conclusion is drawn even though each assessed the issue of comparability of techniques somewhat differently.

However, there are two important caveats to note. First, these results can only be considered indicative, rather than definitive, as they are based on a small sample. Second, the WTP values reflect only a limited range of benefit - that of health status. It is argued that a strength of WTP is the ability to encompass a wider range of benefits, in terms of option and externality value, although it is this authors contention that either technique may do this - it depends on how the scenario is specified (Olsen & Smith, 1999). Nevertheless, by indicating a lack of comparability based on a narrow definition of benefit, it is likely that the divergence would be more marked if these values were taken into account on the WTP side.

Notwithstanding these caveats, however, the directions in which the data point are interesting, novel and worthy of continued research. These include:

First, as was shown in figure 1, the boxplot of WTP$ values, individual WTP values became more dispersed as the health state worsened. The point was made in the discussion there that one possible explanation for this could be the increasing impact of the budget constraint (ATP). That is, there is a broad agreement between individuals concerning the 'value' of a minor change in health state. At this level the value is so low that there is no *effective*, or *relevant*, budget constraint, and thus all are free to provide this value when asked. However, as health state changes become more considerable, the budget constraint becomes an effective barrier to some individuals increasing their WTP in line with others thus forcing the divergence of valuations. That is, it may be that individuals, with the same income, would give approximately the same WTP response, and it is only the difference in income that causes the wider dispersion of values. There is therefore a range of health state changes in which income is irrelevant as an influencing factor, and a range where it becomes more relevant. Depending upon the good itself, this has important implications for the consideration of ATP as an important factor in the use of WTP as a methodology of obtaining values and is worth further research.

Second, neither TTO nor WTP differentiates very well between broad dimensions of health. This is likely to be due to either the dimensions used or that it is the level of health status which is important, and the dimension within which that level occurs is irrelevant. It is not, unfortunately, possible from this data to establish which is the case, and again warrants further research.

Third, it was found that WTP better differentiates between levels of health status within a dimension than TTO, implying that it is a more sensitive measure in the valuation of small changes in health status.

Fourth, as mentioned in section 4.6.1, the rankings provided by TTO and WTP for the health states valued are closer for those states considered to be the 'worst' quarter, with less stability in ranking for the remainder. This suggests that any comparability between techniques would have to be at the levels of health which are commonly viewed as being the worst, where the WTP in time or money is substantial to avoid them.

Finally, of those not responding due to ethical reasons, nine did not respond to TTO versus three non-responses to WTP, implying that there is in fact less ethical objection to WTP than TTO!

Overall, this study indicates that, with the possible of exception of very poor health states, **these two techniques should not be considered comparable**. Further, although WTP seems to perform better in being more sensitive to changes in levels of health status, the dispersion of values due to an income constraint needs to be addressed, and the development of means around this are worth pursuing. It might be, for example, that phrasing in terms of taxation or insurance, as has been advocated, may mean that the income constrain remains reasonable ineffective as a barrier at such low levels of actually WTP.

REFERENCES

Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien B 1998, 'Valuing outcomes in healthcare: A comparison of willingness to pay and quality-adjusted life years', *Journal of Clinical Epidemiology*, vol 51, no 8, pp 667-676.

Boadway RW 1974, 'The welfare foundations of cost-benefit analysis', *Economic Journal*, vol 84, vol 926-939.

Boadway RW, Bruce N 1984, *Welfare Economics*, Oxford, Basil Blackwell.

Drummond M 1989, 'Output measurement for resource allocation decisions in health care', *Oxford Review of Economic Policy,* vol 5, no 1, pp 59-74.

Drummond MF, O'Brien B, Stoddart G, Torrance G 1997, *Methods for the Economic Evaluation of Health Care Programmes*, New York, Oxford University Press (2nd Edition).

Hawthorne G, Richardson J 1995, An Australian MAU/QALY instrument: rationale and preliminary results, CHPE Working Paper 49.

Hawthorne G, Osborne R, McNeil H, Richardson J 1996, 'The Australian Multi-attribute Utility (AMAU): Construction and initial evaluation, CHPE Working Paper 56.

Hawthorne G, Richardson J, Osborne R, McNeil H 1997a, The Australian quality of life (AQoL) instrument: Initial validation, CHPE Working Paper 66.

Hawthorne G, Richardson J, Osborne R, McNeil H 1997b, The assessment of quality of life (AQoL) instrument: Construction, initial validation and utility scaling, CHPE Working Paper 76.

Hoaglin DC, Mosteller F, Tukey J, 1983, *Understanding Robust and Exploratory Data Analysis,* John Wiley & Sons Inc, New York. Chapter 3: Emerson JD, Strenio J, Boxplots and Batch Comparison.

Johansson PO 1995, *Evaluating Health Risks: An Economic Approach*, Cambridge University Press.

Johannesson M 1996, *Theory and Methods of Economic Evaluation of Health Care*, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Jones-Lee MW 1989, *The Economics of Safety and Physical Risk,* Blackwell, Oxford.

Jones-Lee MW, Hammerton M, Philips PR 1985, 'The value of safety: Results of a national sample survey', *The Economic Journal*, vol 95, pp 49-72.

Mehrez A, Gafni A 1989, 'Quality-adjusted life years, utility theory and healthy-year equivalents', *Medical Decision-Making*, vol 9, no 2, pp 142-149.

Mehrez A, Gafni A 1992, Preference-based outcome measures for economic evaluation of drug interventions: Quality adjusted life years (QALYs) versus healthy years equivalents (HYEs), *PharmacoEconomics*, vol 1, no 5, pp 338-345.

O'Brien BJ, Viramontes JL 1994, 'Willingness to pay: a valid and reliable measure of health state preference?', *Medical Decision Making*, vol 14, pp 289-297.

O'Brien B, Gafni A 1996, 'When do the 'dollars' make sense? Toward a conceptual framework for contingent valuation studies in health care', *Medical Decision-Making*, vol 16, pp 288-299.

Olsen JA, Smith RD 1999, From theory to practice: a review of the theoretical arguments for WTP, and its application in 54 surveys, CHPE Working Paper 83.

Olsen JA, Smith RD, Harris A, 1999, Economic theory and the monetary valuation of health care: An overview of the issues as applied to the economic evaluation of health care programs, CHPE Working Paper No. 82.

Pauly MV 1995, 'Valuing health care benefits in money terms', in *Valuing Health Care*", ed FA Sloan, Cambridge University Press, Cambridge.

Richardson J 1994, 'Cost utility analysis: what should be measured', *Social Science and Medicine*, vol 39, pp 7-21.

Smith RD 1997, Contingent valuation: Indiscretion in the adoption of discrete choice question formats?, Working paper 74, Centre for Health Program Evaluation.

Smith RD, Dobson M 1993, ‘Measuring utility values for QALYs: Two methodological issues’, *Health Economics*, vol 2, no 4, pp 349-355.

Smith RD, Jan S, Shiell A 1993, *What are Australians willing to pay for road safety?*, Discussion Paper No. 21, Centre for Health Economics Research and Evaluation, University of Sydney, Sydney.

Smith RD, Hall J, Gurney HG, Harnett PR 1993, A cost-utility approach to the use of 5-fluorouracil and levamisole as adjuvant chemotherapy for Dukes’ C colonic carcinoma’, *Medical Journal of Australia*, vol 158, no 5, pp 319-322.

Sugden R & Williams A 1978, *The Principles of Practical Cost-Benefit Analysis*, Oxford, Oxford University Press.

Thompson MS 1986, 'Willingness to pay and accept risks to cure chronic disease, *American Journal of Public Health*, vol 76, pp 392-396.

Torrance GW 1986, 'Measurement of health state utilities for economic appraisal', *Journal of Health Economics*, vol 5, pp 1-30.

APPENDIX 1

WTP Studies in Health Care: 1985-1998

Appel LJ, Steinberg EP, Powe NR, Anderson GF, Dwyer SA, Faden RR 1990, 'Risk reduction from low osmolality contrast media: What do patients think it is worth?', *Medical Care*, vol 28, no 4, pp 324-337.

Bala MV, Wood LL, Zarkin GA et al 1998, 'Valuing outcomes in healthcare: A comparison of willingness to pay and quality-adjusted life years', *Journal of Clinical Epidemiology*, (forthcoming).

Berwick DM, Weinstein MC 1985, 'What do patients value? Willingness to pay for ultrasound in normal pregnancy', *Medical Care*, vol 23, no 7, pp 881-893.

Blumenschein K, Johannesson M 1998, 'The relationship between quality of life instruments, health state utilities and willingness to pay in patients with asthma', *Annals of Allergy, Asthma and Immunology* (in press).

Chestnut LG, Keller LR, Lambert WE, Rowe RD 1988, 'Measuring heart patients' willingness to pay for changes in angina symptoms', *Medical Decision Making*, 1996; vol 16, no 1, pp 65-77.

Culbertson VL, Arthur TG, Rhodes PJ, Rhodes RS 1988, 'Consumer preferences for verbal and written medication information', *Drug Intell Clin Pharm*, vol 22, no 5, pp 390-396.

Donaldson C 1990, 'Willingness to pay for publicly provided goods: A possible measure of benefit?', *Journal of Health Economics*, vol 9, no 1, pp 103-118.

Donaldson C, Shackley P, Abdalla M, Miedzybrodzka Z 1995, ‘Willingness to pay for antenatal carrier screening for cystic fibrosis’, *Health Economics*, vol 4, no 6, pp 439-452.

Donaldson C, Mapp T, Ryan M, Curtin K 1996, ‘Estimating the economic benefits of avoiding food-borne risk: is 'willingness to pay' feasible?’, *Epidemiol Infect*, vol 116, no 3, pp 285-294.

Donaldson C, Shackley P 1997, ‘Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy’, *Social Science & Medicine*, vol 44, no 5, pp 699-707.

Donaldson C, Shackley P, Abdalla M 1997, ‘Using willingness to pay to value close substitutes: carrier screening for Cystic Fibrosis revisited’, *Health Economics*, vol 6, no 2, pp 145-159.

Donaldson C, Thomas R, Torgerson DJ 1997, ‘Validity of open-ended and payment scale approaches to eliciting willingness to pay. *Applied-Economics*, vol 29, no 1, pp 79-84.

Donaldson C, Mapp T, Farrar S, Walker A, MacPhee S 1997, ‘Assessing community values in health care: is the willingness to pay method feasible? *Health Care Analysis*, vol 5, pp 7-29.

Donaldson C, Hundley V and Mapp T 1998, ‘Willingness to pay: a method for measuring preferences for maternity care?’, *Birth*, vol 25, pp 33-40.

Eastaugh SR 1991, ‘Valuation of the benefits of risk-free blood. Willingness to pay for hemoglobin solutions’, *Int J Technol Assess Health Care*, vol 7, no 1, pp 51-57.

Ellison P, Marr J 1994, ‘Willingness to pay for unfunded health services in a family practice clinic’, *Leadership in Health Services*, vol 3, pp 14-17.

Fisher M, Marks A, Trieller K, Brody R 1985, ‘Are adolescents able and willing to pay the fee for confidential health care?’, *Journal of Pediatrics*, vol 170, pp 480-483.

Gafni A, Feder A 1987, ‘Willingness to pay in an equitable society: The case of the Kibbutz’, *International Journal of Social Economics*, vol 14, no 1, pp 16-21.

Gibb S, Donaldson C and Henshaw R 1998, ‘Assessing strength of preference for abortion method using willingness to pay’, *Journal of Advanced Nursing*, pp 30-36.

Golan EH, Shechter M 1993, ‘Contingent valuation of supplemental health care in Israel’, *Medical Decision Making*, vol 13, pp 302-310.

Gore PR, Madhavan S 1994, ‘Consumers' preference and willingness to pay for pharmacist counselling for non-prescription medicines’, *J Clin Pharm Ther*, vol 19, no 1, pp 17-25.

Granberg M, Wikland M, Nilsson L, Hamberger L 1995, ‘Couples' willingness to pay for IVF/ET’, *Acta Obstet Gynecol Scand*, vol 74, no 3, pp 199-202.

Hassan EO, el-Nahal N, el-Hussein M 1994, 'Acceptability of the once-a-month injectable contraceptives Cyclofem and Mesigyna in Egypt', *Contraception*, vol 49, no 5, pp 469-488.

Johannesson M, Jonsson B, Borgquist L 1991, 'Willingness to Pay for Antihypertensive Therapy - Results of a Swedish Pilot Study', *Journal of Health Economics*, vol 10, no 4, pp 461-474.

Johannesson M, Aberg H, Agreus L, Borgquist L, Jonsson B 1991, 'Cost-benefit analysis of non-pharmacological treatment of hypertension', *J Intern Med*, vol 230, no 4, pp 307-312.

Johannesson M, Fagerberg B 1992, 'A health-economic comparison of diet and drug treatment in obese men with mild hypertension', *Journal of Hypertension*, vol 10, no 9, pp 1063-1070.

Johannesson M, et al 1993, 'Willingness to pay for lipid lowering: A health production function approach', *Applied-Economics*, vol 25, no 8, pp 1023-1031.

Johannesson M, et al 1993, 'Willingness to pay for Antihypertensive therapy -- further results', *Journal of Health Economics*, vol 12, no 1, pp 95-108.

Johannesson M, Johansson PO 1997, 'Quality of life and the WTP for an increased life expectancy at an advanced age', *Journal of Public Economics*, vol 65, no 2. pp 219-228.

Johannesson M, O'Conor RM, Kobelt-Nguyen G, Mattiason A 1997, 'Willingness to pay for reduced incontinence symptoms', *British Journal of Urology*, vol 80, pp 557-562.

Kartman B, Andersson F, Johannesson M 1996, 'Willingness to pay for reductions in angina pectoris attacks', *Medical Decision Making*, vol 16, pp 248-253.

Kartman B, Stalhammar NO, Johannesson M 1996, 'Valuation of health changes with the contingent valuation method: a test of scope and question order effects', *Health Economics*, vol 5, no 6, pp 531-541.

Krupnick AJ, Cropper ML 1992, 'The effect of information on health risk valuations', *Journal of Risk and Uncertainty*, vol 5, pp 29-48.

Lee SJ, Neumann PJ, Churchill WH, Malachowski ME, Weinstein MC, Johannesson M 1997, 'Patients' willingness to pay for autologous blood donation', *Health Policy*, vol 40, pp 1-12.

Lee SJ, Liljas B, Neumann PJ, Weinstein MC, Johannesson M, 'The impact of risk information on patients' willingness to pay for autologous blood donation', *Medical Care* (in press).

Lindholm L, Rosen M, Hellsten G 1994, 'Are people willing to pay for a community-based preventive program', *Int J Technol Assess Health Care*, vol 10, no 2, pp 317-324.

Nathan RG, Bont GM, Minz RB 1994, 'Patient interest in receiving audiotapes of information presented by their physicians', *Arch Fam Med*, vol 3, pp 509-513.

Neumann PJ, Johannesson M 1994, 'The willingness to pay for in vitro fertilization: A pilot study using contingent valuation', *Medical Care*, vol 32, no 7, pp 686-699.

O'Brien BJ, et al 1995, 'Assessing the economic value of a new antidepressant: A willingness-to-pay approach', *PharmacoEconomics*, vol 8, no 1, pp 34-45.

O'Brien BJ, Goeree R, Gafni A, Torrance GW et al 1997, 'Assessing the value of a new pharmaceutical: A feasibility study of contingent valuation in managed care', *Medical Care*, (forthcoming).

Olsen JA, Donaldson C 1998, 'Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health care programs', *Social Science and Medicine*, vol 46, no 1, pp 1-12.

Osmond MH, Klassen TP, Quinn JV 1995, 'Economic comparison of a tissue adhesive and suturing in the repair of pediatric facial lacerations', *Journal of Pediatrics*, vol 126. pp 892-895.

Pennie RA, O'Connor AM, Garvock MJ, Drake ER 1991, 'Factors influencing the acceptance of Hepatitis B vaccine by students in health disciplines in Ottawa', *Canadian Journal of Public Health*, vol 82, pp 12-15.

Phillips K, et al 1997, 'Willingness to pay for Poison Control Centres', *Journal of Health Economics,* vol 16, no 3, pp 343-357.

Ramsey SD, Sullivan SD, Psaty BM, Patrick DL 1997, 'Willingness to pay for antihypertensive care: evidence from a staff-model HMO', *Social Science and Medicine*, vol 44, no 12, pp 1911-1917.

Reutzel TJ, Furmaga E 1993, 'Willingness to pay for pharmacist services in a Veterans Administration Hospital', *Journal of Research in Pharmaceutical Economics*, vol 5, no 2, pp 89-114.

Ryan M 1996, 'Using willingness to pay to assess the benefits of assisted reproductive techniques', *Health Economics*, vol 5, no 6. pp 543-558.

Ryan M 1997, 'Should government fund assisted reproductive techniques? A Study Using willingness to pay', *Applied Economics*, vol 29, no 7, pp 841-849.

Ryan M, Ratcliffe J, Tucker J 1997, 'Using willingness to pay to value alternative models of antenatal care', *Social Science & Medicine*, vol 44, no 3, pp 371-380.

Slothuus U 1997, Comparison of two contingent valuation methods: closed-ended questions with and with-out follow-up. Paper presented at the 18th Nordic Health Economists' Study Group Meeting, August 1997, Finland.

Stalhammer NO 1996, 'An empirical note on willingness to pay and starting-point bias', *Medical Decision Making*, vol 16, pp 242-247.

Thompson MS 1986, 'Willingness to pay and accept risks to cure chronic disease', *Am J Public Health*, vol 76, no 4, pp 392-396.

Weaver M, Ndamobissi R, Karnfield R, et al 1996, 'Willingness to pay for child survival: Results of a national survey in Central Africian Republic', *Social Science and Medicine*, vol 43, pp 985-998.

Zethraeus N 1998, 'Willingness to pay for hormone replacement therapy', *Health Economics*, vol 7, pp 31-38.

APPENDIX 2

WTP Questionnaire

1 VALUATION QUESTIONS

The questions will all follow the same format. You will be presented with a pair of health states. In each case I will ask *you to assume that you are in health state B,[Good Health]* and will ask you the **maximum amount** you would be willing-to-pay *to avoid going immediately into State A,* as presented. Please think of this willingness to pay as being directly paid out-of-pocket. This payment will be ongoing and have to be made regularly. If it stops you will go into State A immediately. You can think of this regular willingness to pay as either a weekly, fortnightly, monthly or yearly figure.

Also, for the purpose of this interview, please imagine that there will be no help with paying for this treatment from Medicare. If you want to avoid going into Health State A you will have to pay for treatment yourself. We want to know how much **you** would be willing to pay.

When giving your response, that is, what you are **willing** to pay, please bear in mind what you are **able to afford**. Given your own circumstances (assets and income), bear in mind that if you are willing to pay an amount to secure this improvement you will have less to spend on other things, such as entertainment, your house, car, holidays and so on.

If you are not willing to pay anything to change health states, then please say zero.

As always there aren't right or wrong answers. We're only interested in what you think.

[INTERVIWER: Get each set of health states out one by one, and offer the choice to respondents. Record the maximum willingness to pay for each one in the space provided and also record whether or not the prompt was used.]

2 PROMPT

If the respondent is unable to give an amount then this prompt may be used and the 'prompt used' circled.

PROMPT

Would it help if I suggested an amount to give you a start?

Would you be willing to pay $X [vary X across $25, $50, $75, $100]

If no, then read out decrements in $10 amounts steadily and quickly.

If yes, read out increments in $10 amounts steadily and quickly.

When respondent stops, ask

So you would be WTP $X pw to be in HS B rather than HS A?

Would you pay exactly X or slightly more or less. How much exactly?

WRITE IN SUM $.....

PROMPT USED

If any response is a zero, ask the respondent to clarify if they are simply not willing to pay anything, that they do not actually value the change in healthstate (ie are indifferent between states A and B), or there is another reason - such as income constraint, ethical objections to such valuation, feel that this payment may become policy, etc.