UTS:CHERE

The Centre for Health Economics Research and Evaluation (CHERE) was established in 1991. CHERE is a centre of excellence in health economics and health services research. It is a joint Centre of the Faculties of Business and Nursing, Midwifery and Health at the University of Technology, Sydney, in collaboration with Central Sydney Area Health Service. It was established as a UTS Centre in February, 2002. The Centre aims to contribute to the development and application of health economics and health services research through research, teaching and policy support. CHERE's research program encompasses both the theory and application of health economics. The main theoretical research theme pursues valuing benefits, including understanding what individuals value from health and health care, how such values should be measured, and exploring the social values attached to these benefits. The applied research focuses on economic and the appraisal of new programs or new ways of delivering and/or funding services. CHERE's teaching includes introducing clinicians, health services managers, public health professionals and others to health economic principles. Training programs aim to develop practical skills in health economics and health services research. Policy support is provided at all levels of the health care system by undertaking commissioned projects, through the provision of formal and informal advice as well as participation in working parties and committees.

University of Technology, Sydney
City campus, Haymarket
PO Box 123 Broadway NSW 2007
Tel: +61 2 9514 4720
Fax: + 61 2 9514 4730
Email: mail@chere.uts.edu.au
www.chere.uts.edu.au

Health care policy evaluation: empirical analysis of the restrictions implied by Quality Adjusted Life Years

Rosalie Viney[1] and Elizabeth Savage[1]

CHERE WORKING PAPER 2006/10

1. Centre for Health Economics Research and Evaluation
Faculty of Business
University of Technology, Sydney

Version: June 2005

Abstract

This paper investigates the nature of the utility function for health care, defined over the probability of survival, survival duration, health state and cost of treatment. A discrete choice experiment, involving treatment choice for a hypothetical health condition is used to test restrictions on preferences in the QALY model. We find that preferences do not conform to expected utility, and there are significant interactions between health state and survival duration. Individual characteristics are significant, implying substantial differences in valuations of health states across the population. The results suggest the QALY approach distorts valuations of health outcomes.

Acknowledgements

This research was supported by a National Health and Medical Research Council Program Grant.

1 Introduction

Economic evaluation is used increasingly by health system decision-makers to determine allocation of health care resources between services and levels of subsidy. Since 1993, legislation has required that new pharmaceuticals to be listed on the Australian Pharmaceutical Benefits Scheme are assessed in terms of cost-effectiveness as well as safety and efficacy. A similar requirement has been introduced for new procedures and diagnostic tests to be funded on the Medicare Medical Benefits Schedule. In 1999 the United Kingdom government established the National Institute for Clinical Excellence which has responsibility for developing guidelines for health care services for health authorities based on clinical and economic evaluation and for undertaking technology appraisal based on effectiveness and cost-effectiveness. Denmark, Wales, Finland, Sweden, the Netherlands and the Canadian province of Ontario all have formal requirements for the use of economic evaluation in reimbursement decisions for pharmaceuticals and a number of other jurisdictions have some less formal arrangements for use of economic evaluation.

The theoretical underpinnings of economic evaluation in health care are based in welfare economics. In circumstances where free markets are not expected to result in socially optimal allocations, cost-benefit analysis provides information about the welfare outcomes of different policies. In cost-benefit analysis, market and shadow prices are used to provide revealed preference valuations of the costs and benefits of alternative courses of action, with the decision rule being to fund the project from the set of available alternatives which has the highest net social benefit at the margin. In general, policy decisions must be made in the context of a budget constraint, and there will both gainers and losers from any policy. The value of gains and losses must be estimated and aggregated using some social welfare function to evaluate the social welfare impacts of costs and benefits of policies, most commonly using money as the metric. This requires information on individuals' preferences (Slesnick 1998).

Public funding decisions require assessment of resource allocation across diverse treatments for a range of diseases with disparate impacts. In practice economic evaluation in health care has focused only on health and has not used monetary measures of welfare impacts. Cost-effectiveness analysis (CEA) is more commonly used than cost-benefit analysis because of practical difficulties with finding acceptable monetary valuations of survival and quality of life outcomes and ethical concerns with using these valuations has meant that in health economics (Drummond et al. 1997). In CEA, outcomes are measured in terms of a uni-dimensional "natural" unit (cases prevented, lives saved or life-years saved). If outcomes of interventions are multi-dimensional the rankings arising from CEA may differ from welfare rankings.

The outcomes that have been considered in health care are survival and quality of life, and cost-utility analysis (CUA), a more generalized form of CEA, has emerged as the dominant approach. In CUA survival and quality of life are combined into a single outcome measure, most commonly the Quality Adjusted Life Year (QALY) (Torrance 1986; Drummond et al. 1997). The time in a particular health state is weighted by a quality weight between zero (death) and one (full health) that reflects society's

willingness to trade-off between quality of life and survival. QALYs are designed to allow comparisons across interventions with disparate outcomes, across different health care conditions and population groups. QALYs are treated as a cardinal measure of welfare impacts of health interventions, allowing ranking in terms of cost per additional QALY gained.

The QALY approach requires an accurate description of the health outcomes associated with a disease state and/or intervention and a method for eliciting preferences for the health outcomes. A number of approaches for describing and eliciting preferences for health outcomes have been developed, all of which impose restrictive assumptions about how the attributes of health outcomes are combined, and how individuals trade-off between expected health outcomes and other goods. If these restrictions do not hold then QALYs will not be an accurate representation of preferences, and the resulting resource allocation decisions may not be welfare improving. It is important to determine whether these restrictions can be supported empirically, and hence, whether QALYs provide an index of utility.

The restrictions on individual utility imposed by the QALY model have been investigated and formally stated in a number of papers (Pliskin et al. 1980; Bleichrodt 1995; Johannesson 1995; Bleichrodt and Quiggin 1997; Bleichrodt et al. 1997; Miyamoto et al. 1998; Ried 1998; Bleichrodt and Quiggin 1999), but there has been relatively little empirical investigation of them. Revealed preference data are generally not of a suitable form to allow such investigation, and the complexity of health care contexts makes it difficult to design experimental studies that provide for general investigation of the QALY assumptions.

A range of approaches based on stated preference data have been used to test aspects of the QALY model, generally based on the use of standard gamble or time trade-off experiments in different populations (Bleichrodt and Johannesson 1997; Bleichrodt and Johannesson 1997; Bleichrodt et al. 2003; Spencer 2003). A feature of these experiments is that they have tended to be designed to test a specific assumption or subset of assumptions.

Over the past decade there has been growing interest in the application of discrete choice experiment (DCE) methods to in health economics. DCEs allow for analysis of preferences for complex multi-attribute goods when limited market data are available. The approach is readily applicable to the multi-dimensional nature of health outcomes where there are impacts on different aspects of quality of life and on survival, and enables quantification of the individual trade-offs between the different dimensions. For example, Hakim and Pathak (Hakim and Pathak 1999) use a DCE to derive utility weights for a health related quality of life instrument (the EQ-5D instrument) and Gyrd-Hansen (Gyrd-Hansen 2003) uses a modification of a DCE approach to estimate the value of a QALY.

DCEs are an alternative to methods that have previously been used to test aspects of the QALY model. They can provide a rich source of data to explore the underlying utility function for health and health care, particularly the trade-off between quality of life and survival. An appropriately designed DCE allows estimation of more flexible utility specifications, and more general tests, allowing a more general policy

evaluation framework. In this paper we use a DCE to explore the restrictions on preferences for health care imposed by the QALY model.

In Section 2 of the paper we provide an overview of the QALY model and its basis in consumer preference theory. Section 3 describes the experiment undertaken to test the QALY model. Section 4 describes the data collection and analysis, and in Section 5 of the paper we present the empirical results and discuss the implications for the use of cost-utility analysis for health care resource allocation.

2 A model of individual health care decision making

An individual is assumed to derive utility over her lifetime from her health state and from consumption of goods in each period:

$$U = U((h_1, c_1), \ldots, (h_T, c_T)) \tag{1}$$

The lifetime profile of health states and consumption is uncertain. A particular illness or disease is associated with a probability distribution over profiles, and a treatment can be conceptualized as modifying this probability distribution. Individuals making choices about treatments face a gamble defined over health and other outcomes. A treatment can be characterised as a lottery over time profiles of health states and other consumption:

$$P = \{((H_1, C_1); p_1), \ldots ((H_m, C_m); p_m)\} \tag{2}$$

where pi is the probability of experiencing the health and consumption profile:

$$(H_i, C_i) = (h_{i1}, c_{i1}), \ldots, (h_{iT}, c_{iT}) \quad , \; i = 1 \ldots m \tag{3}$$

The evaluation of alternative treatments is therefore a problem of determining the preference ordering over different lotteries each of the form of equation (2) above, and the associated utilities. For example, in choosing between treatment (TR) and no treatment (NT) for a particular disease, the individual will choose treatment if:

$$U(P^{TR}) > U(P^{NT}) \tag{4}$$

The restrictions on preferences imposed by the QALY model have been discussed by a number of authors, addressing different specifications of the utility function (Pliskin et al. 1980; Miyamoto and Eraker 1988; Johannesson et al. 1996; Bleichrodt and Quiggin 1997; Bleichrodt et al. 1997; Miyamoto et al. 1998; Dolan 2000). Bleichrodt and Quiggin (1999) define the conditions under which the QALY model is a valid index of utility defined over a profile of non-constant health states and consumption, as in equation (2) above.

The first step in deriving the QALY model from equation (4) is to assume that preferences conform to Von Neumann Morgenstern (VNM) expected utility, so that

utility over a time profile of health states can be represented as the probability weighted sum of each time profile, thus:

$$U = \sum_{i=1}^{m} p_i . \tilde{U}(H_i, C_i) \tag{5}$$

The second step is to specify that the utility function $\tilde{U}(H_i, C_i)$ is of additive independent form over the one-period utility functions:

$$\tilde{U}(H_i, C_i) = \sum_{t=1}^{T} u_t(h_{it}, c_{it}) \tag{6}$$

Bleichrodt and Quiggin (1997; Bleichrodt and Quiggin 1999) show that, when expected utility is imposed, additive independence holds if the preference relation over lotteries of consumption and health profiles satisfies marginality. A preference relation over a lottery satisfies marginality if the preference ordering depends only on the marginal probability distributions rather than the joint probability distribution. For example, suppose there are two time periods, and in each time period the health state can be good, h_t^g or bad, h_t^b, and the consumption level can be 10 or 1. Marginality implies that an individual is indifferent between Lottery A, with a 0.5 probability of the good outcome $(h_t^g, 10)$ in both periods and a 0.5 probability of the bad outcome $(h_t^b, 1)$ in both periods and Lottery B, with a 0.5 probability of the good outcome in the first period and the bad outcome in the second period, and a 0.5 probability of the bad outcome in the first period and the good outcome in the second period:

$$U^A\left[\left((h^g,10),(h^g,10);0.5\right),\left((h^b,1),(h^b,1);0.5\right)\right] = U^B\left[\left((h^g,10),(h^b,1);0.5\right),\left((h^b,1),(h^g,10);0.5\right)\right]$$

Marginality excludes all complementarity across time periods, such as a dislike for variation across time, or a preference for a lottery which avoids experiencing the bad outcome in both time periods.

The third step is to impose the restriction that one period utility functions are identical, which is achieved by assuming that the preference relation over lotteries conforms to symmetry. Symmetry holds if, for a given profile (H_i, C_i) the individual is indifferent to the time ordering of the health state-consumption combinations, that is, between profiles with the same health state-consumption combinations but in a different sequence. Symmetry cannot hold if the individual has a positive rate of time preference. With symmetry:

$$\tilde{U}(H_i, C_i) = u(h_{i1}, c_{i1}) + ... + u(h_{iT}, c_{iT}) = \sum_{t=1}^{T} u(h_{it}, c_{it}) \tag{7}$$

The fourth step is to impose independence between consumption and health state in the identical one-period utility functions. This is achieved by assuming the one-period

utility function satisfies standard gamble invariance. Standard gamble invariance holds if, for any levels of consumption c and c':

$$u(h,c) \geq p\,u(h',c) + (1-p)u(h'',c) \Leftrightarrow u(h,c') \geq p\,u(h',c') + (1-p)u(h'',c')$$

Imposing standard gamble invariance means that if consumption is held constant, the preference ordering over health state-consumption combinations depends only on the preference ordering over health states. Bleichrodt and Quiggin (1999) show that imposing standard gamble invariance means that $u(h_t, c_t)$ can be written as an affine transformation of $v(h_t)$:

$$u(h_t, c_t) = w_t(c_t) + z(c_t)v(h_t) \tag{8}$$

Because the utility function is unique up to a positive linear transform, equation (8) can be further simplified by choosing an appropriate transformation. If h is defined such that at death, h is equal to zero, and $v(h_t)$ is scaled such that *v(0)=0*, then

$$\begin{aligned} u(\text{death}, c_t) &= w_t(c_t) + z(c_t)v(0) \\ &= w_t(c_t) \end{aligned} \tag{9}$$

The final step is to impose an assumption known as the zero condition (Bleichrodt et al. 1997; Bleichrodt and Quiggin 1999), which states that the individual obtains no more utility from additional consumption when dead. The zero condition implies that the utility of consumption when life duration is zero is invariant to the level of consumption, hence for any real non-negative levels of consumption, c' and c'', when the health state is death

$$\begin{aligned} u(0, c_t') &= u(0, c_t'') \\ w_t(c_t') + z(c_t')v(0) &= w_t(c_t'') + z(c_t'')v(0) \\ w_t(c_t') &= w_t(c_t'') \end{aligned} \tag{10}$$

Therefore, the zero condition implies that $w_t(c_t)$ is a constant, and again, employing the fact the utility function is unique up to a positive linear transform, $w_t(c_t)$ can be set equal to zero. Thus, if preferences for lotteries defined over lifetime profiles of health care and consumption conform to the VNM expected utility axioms, marginality, symmetry, standard gamble invariance and the zero condition, the utility function for lotteries over profiles of health states and consumption can be written as the QALY model (Bleichrodt and Quiggin 1999):

$$U = \sum_{i=1}^{m} p_i \sum_{t=1}^{T} z(c_{it})v(h_{it}) \tag{11}$$

The function $v(h_{it})$ gives the QALY weight, scaled such that v(death) is zero and v(full health) is unity. If these restrictions hold, the QALY weight for a particular health state can be assumed to be independent of time, consumption, and the profile of

health states or consumption already experienced. This simplifies the problem of valuing the outcomes of health interventions to one of identifying the set of health states that will occur as a result of the disease/intervention, the probabilities that they occur and the likely time profile of health states, and then calculating appropriate measures of the QALY weight, v(h) of those health states.

This model does not explicitly allow for time preference. If discounting is introduced, the symmetry restriction must be replaced by trade-off consistency. Trade-off consistency holds if the individual's strength of preference for one outcome over another in a given time period is invariant across time periods. Bleichrodt and Quiggin (1999) show that by imposing trade-off consistency in addition to marginality, standard gamble invariance and the zero condition, under VNM expected utility,

$$U = \sum_{i=1}^{m} p_i \sum_{t=1}^{T} \lambda_t z(c_{it}) v(h_{it}) \tag{12}$$

This model allows the discounting factor, λ_t , to vary over time, and can be further simplified by assuming a constant discount rate. A constant discount rate is achieved by imposing a stationarity condition on preferences. Stationarity means that the preference ordering over two different profiles of health state-consumption combinations that share one common outcome is unaffected by the timing of the common outcome.

In summary , as noted by Dolan (2000), the "enhanced generalisability that comes from constructing an almost infinite number of profile scores (QALYs) from valuations of a finite number of composite states comes at a price, namely, a number of restrictive assumptions about individual preferences" (p.1731). Once these restrictions are imposed, relatively straightforward stated preference experiments can be used to value health outcomes. The most commonly used experiments are standard gamble and time trade-off experiments (Torrance 1986; Drummond et al. 1997).

3 A discrete choice experiment to test the restrictions of the QALY model

In DCEs, respondents are presented with a series of hypothetical choice sets, each with two or more alternatives, and asked to choose their preferred alternative from each choice set. The alternatives in the choice set are described in terms of attributes, which may be generic (common across alternatives) or alternative specific (belonging only to particular alternatives). The alternatives may be labeled, to allow for an underlying preference for a particular product or brand (for example, medical treatment, surgical treatment), or they may be unlabelled (for example, treatment A, treatment B). The attributes of the alternatives are varied over a plausible and policy relevant range, generally expressed as a set of discrete levels. Experimental design principles are used to determine the selection of the combination of levels of attributes for each alternative to be included in each choice set.

Discrete choice analysis derives from Random Utility Theory (Thurstone 1927; McFadden 1973; Manski 1977) and is described elsewhere (Louviere et al. 2000; Hall

et al. 2002; Viney et al. 2002). Utility is not directly observable, but can be estimated from observed choices. Given the choice between two alternatives *j* and *k*, the probability that individual *i* chooses alternative *j* is:

$$\begin{aligned} P_i(j) &= \Pr(U_{ij} \geq U_{ik}) \\ &= \Pr\left(V_{ij} - V_{ik} \geq \varepsilon_{ik} - \varepsilon_{ij}\right) \end{aligned} \tag{13}$$

where U_{ij} is the utility of choice *j* for individual *i*, V_{ij} is the systematic component of the utility and ε_{ij} is the random, or unexplained, component, which may be due to unobserved or unobservable attributes of the choice, unobserved taste variation or measurement error (McFadden 1973; Ben-Akiva and Lerman 1985). The systematic component of utility V_{ij} depends on attributes of the alternative, *j*, and on attributes of the individual, *i*, making the choice. The model estimated depends on assumptions made about the distribution of the random component, and the nature of the choice being modelled. It is commonly assumed that the ε_{ij} are independently and identically distributed, with a Gumbel distribution, leading to a conditional logit model (McFadden 1973) :

$$P_i(j) = \frac{e^{V_{ij}}}{e^{V_{ij}} + e^{V_{ik}}} = \frac{e^{\mu\beta'\mathbf{x}_{ij}}}{e^{\mu\beta'\mathbf{x}_{ij}} + e^{\mu\beta'\mathbf{x}_{ik}}} \tag{14}$$

The scale parameter μ is inversely proportional to the variance of the random component. It is not possible to estimate the β's separately from the scale parameter, and it is commonly assumed that $\mu = 1$.

We develop a discrete choice experiment that allowed for observation of choices over lotteries that vary in terms of the probability of outcomes, defined in terms of duration of survival, quality of life and consumption, allowing for tests of the restrictions imposed by the QALY model. The experiment is a discrete choice analogue of the standard gamble and time trade-off tasks typically used to elicit weights for QALYs, with cost of treatment included as a proxy for loss of consumption. Under VNM expected utility, the key restrictions of the QALY model to be tested are marginality, symmetry and standard gamble invariance. It is also possible to test whether preferences for health care conform to VNM expected utility. The zero condition is both relatively non-controversial and not readily empirically testable.

The experiment presents a series of forced choices between two labelled alternatives, "treatment" (TR) and "no treatment" (NT) for a set of hypothetical health conditions. In each choice set, respondents are endowed with a health condition and asked to choose between treatment and no treatment. A forced choice allows all possible outcomes to be specified explicitly, rather than in the case of an experiment with an opt-out alternative, where the perceived attributes of the opt-out alternative cannot be specified in the experiment and may vary across choice sets and across respondents.

Each alternative in a choice set represents a lottery defined over health outcomes and consumption. For the treatment alternative, there is a cost to the consumer of

undergoing treatment, and the treatment provides a probability of returning to full health for a specified survival duration, and a complementary probability of death. For the no treatment alternative there is no cost to the consumer, but the consumer has a probability of remaining in the endowed health condition for a specified survival duration, and a complementary probability of death. Thus, the outcomes of the lottery are a specified chronic health state, with a variable but known duration, or death.

Table 1 presents the attributes and levels used in the experiment. For the treatment alternative, the attributes are the probability, P_{tr}, that the treatment will be successful and return the individual to full health (for which there is a complementary probability that the treatment will be unsuccessful and result in death); the life expectancy, T_{tr}, the individual will experience if treatment is successful; and the cost of treatment, C. For the "no treatment" alternative, the attributes are the probability, $P_{nt,}$ of surviving beyond one month with the health condition (for which there is a complementary probability of dying within one month)[1]; the life expectancy, T_{nt}, for an individual who survives with the health condition; and the health related quality of life, H, associated with the condition. Thus, in the no treatment alternative the cost attribute is fixed at zero, and in the treatment alternative, the health state is fixed at full health. For reasons of plausibility, the probabilities of survival in the no treatment alternative vary over a larger range than in the treatment alternative.

Table 2 summarises the health states used in the experiment, which are drawn from the EQ-5D (Dolan 1997; Hakim and Pathak 1999). The health states were selected to cover the range of health states represented by EQ-5D, to be ordered, and to ensure that the levels on each dimension are reasonable when considered together, for example, so that "confined to bed" is not combined with "no problems with usual activities".

All possible choice sets are given by the factorial combination of the attributes of both alternatives, which results in 4096 choice sets. For each alternative there are 64 possible combinations of attributes. To ensure that all within-alternative interactions could be estimated, the experimental design incorporates the full factorial for each alternative. This requirement can be minimally met by combining each combination of attributes for the first alternative with one combination for the second alternative, but the choice of these combinations to construct the choice sets affects the efficiency of the design. Details of the experimental design and the data collection are provided in Viney, Savage and Louviere (Viney et al. forthcoming).

4 Data and model

A random sample of 347 respondents each completed 16 choice sets, providing 5552 observations. Figure 1 presents plots of the marginal frequencies for each of the attributes, that is the proportion of respondents who chose treatment when presented

[1] This attribute reflects the "immediate death" outcome in SG tasks for obtaining QALY weights, but the attribute was described in terms of dying within a month to make it more realistic, and to make it more equivalent to the probability of death from treatment (since it might be assumed that it could take up to one month between the decision to have treatment and the treatment taking place).

with each level of a particular attribute, with all other attributes varying. The marginal frequencies show the average impact of a change in the level of the attribute on the probability of choosing treatment. The marginal frequency plots suggest that respondents were most responsive to the probability of survival in each alternative and to the health state experienced in the no treatment alternative. There is relatively little difference between health states 3 and 4 in terms of the probability of choosing treatment, but for the two poorer health states, there is a large increase in the probability of choosing treatment. Respondents were least responsive to changes in the level of the cost attribute. Increasing the survival duration in the treatment state from 10 to 20 years had a large impact on the probability of choosing treatment, but for increases above 20 years there was a relatively small increase in the probability of choosing treatment. There is a similar but smaller effect for increases in the survival duration in the no treatment health state.

Of the 347 respondents, 46 had missing data for some socioeconomic characteristics, most commonly household income. These were excluded from the estimation to allow comparison of models with and without individual covariates. Preliminary estimation showed that the estimated coefficients for the experimental variables were robust to exclusion of these respondents. The mean age of respondents was 37.2 years (sd=11.45). Table 3 summarises the characteristics of the estimation sample (n=301).

To test the assumptions of the QALY model, a multiplicative error term is assumed:

$$U_{ij} = V_{ij}\left(P_j, HS_j, T_j, C_j; z_i\right)e^{\varepsilon_{ij}}, \quad j = tr, nt$$

In log form, the probability of individual i choosing the treatment alternative is given by:

$$\Pr\left(\ln U_{i,tr} \geq \ln U_{i,nt}\right) = \Pr\left(\ln V_{i,tr} - \ln V_{i,nt} \geq \varepsilon_{i,nt} - \varepsilon_{i,tr}\right) \tag{15}$$

If the zero condition is accepted as an uncontroversial restriction on preferences, the general form of the utility function, with expansion up to quadratic terms, is given by:

$$\left.\begin{aligned} &\ln U_{ij} = \ln g\left(P_j\right) + v_j\left(HS_j, T_j, C_j; z_i\right) + \varepsilon_j \qquad j = tr, nt \\ &\text{where} \\ &v_{tr}(.) = \alpha_{tr} + \beta_{T_{tr}} \ln T_{tr} + \beta_C C + \beta_{C,T_{tr}} C \ln T_{tr} + \beta_{C^2} C^2 + \sum_n \beta_n z_{n_i} \\ &v_{nt}(.) = \beta_{T_{nt}} \ln T_{nt} + \sum_k \beta_{HS_{nt,k}} HS_k + \sum_k \beta_{T_{nt},HS_{nt,k}} T_{nt} HS_k \end{aligned}\right\} \tag{16}$$

The health state under the treatment alternative (full health) is constant and is therefore included in the treatment constant. Cost is zero for the non-treatment alternative. The health states for the no treatment alternative are dummy effects coded for estimation. This general form allows for rank dependent expected utility (Quiggin 1993), in which probability enters via a probability weighting function g(Pj), for example,

$$\ln g(P_j) = \gamma \ln P_j \tag{17}$$

Expected utility is a special case of rank dependent expected utility function, with the probability weighting function equal to the probability.

Under the QALY restrictions, the expected utility of an alternative is given by the product of probability of surviving, the survival time and the QALY weight associated with the health state. In log form:

$$\ln U_j = \ln P_j + \ln T_j + \varphi_j(HS_j) + \kappa(C_j) + \varepsilon_j,\ j = tr, nt \tag{18}$$

If the QALY restrictions hold the coefficients on the variables that are common to the treatment and no treatment alternatives (time and probability) should be equal, consumption should be separable, there should be no interaction effects (zero coefficients on the interaction terms), and the coefficients on the survival time and probability terms in the utility function should both be equal to one. To test these restrictions, a number of nested models were estimated using logit and compared using likelihood ratio tests.

6 Results

In all models, an intercept is included only for the treatment alternative, indicating the average propensity to choose treatment over no treatment. Because the choices in the experiment were labelled as treatment and no treatment, a preference for treatment over no treatment can be interpreted either as a preference for the labelled alternative 'treatment' or as a preference for full health over other poorer health states. Individual specific covariates are interacted with a treatment dummy, and therefore the coefficient on a particular covariate can be interpreted as the impact of that characteristic on the choice of treatment over no treatment. All covariates except the age and income variables are dummy effects coded. The age and income variables are entered in log form.

Model comparisons are presented in Table 4. Table 5 presents parameter estimates for the main models of interest, based on the tests of restrictions, discussed below. Model 1 is the most general specification, corresponding to equation 16. The parameters on probability and survival duration differ for treatment and no treatment and are freely estimated, and interaction terms and individual specific covariates are included.

In Model 1, coefficients on the cost and cost interaction terms are not significant, and a test of equality of parameters indicates that the coefficients on the survival time under treatment and no treatment are not significantly different from each other. Model 2 tests the restriction that the coefficients on the cost terms are jointly equal to zero. The likelihood ratio tests reveal that this restriction cannot be rejected. Model 3 further imposes the restriction that the coefficient on the survival time term is generic across treatment and no treatment alternatives (although it does allow for interaction between survival time and health state). Again, this restriction cannot be rejected. As

shown in Table 4, all further restrictions on parameters imposed by the QALY model are rejected. Thus, Model 3 represents the preferred generalised QALY model.

In Model 3, the coefficient on the intercept, which captures the value of full health, is positive and significant, suggesting that, ceteris paribus, individuals have a preference for treatment over no treatment. The coefficients on covariates indicate how preference for treatment varies by individual characteristics. All covariates except sex have a significant impact on treatment indicating that individual characteristics are important in determining tastes for health care and health outcomes. In the standard QALY approach any heterogeneity in preferences for quality of life and survival is ignored; the average QALY weight for the population is used to value interventions. In this analysis, the individual's age, own health status, education, income and wealth are found to be important in determining preferences for treatment. The estimated coefficients for the covariates are robust to different model specifications. The effect of age is negative and significant, suggesting that the preference for treatment decreases with age. The respondent's rating of his/her own health also has a significant impact. Those in poor health are significantly less likely to choose treatment compared with those who rated their own health as fair, good or excellent. The effect of own health rating may reflect the fact that respondents with more experience of poor health states are more willing to tolerate remaining in poor health, or are relatively risk averse. However, those who had a chronic health condition displayed a preference for treatment over no treatment.

The coefficient on income is positive and significant. This suggests that those on higher incomes have a preference for treatment over no treatment. The income term may also capture the impact of the individual's budget constraint despite the cost attribute being insignificant. Similarly, both wealth, reflected by home ownership and holding private health insurance, are positively associated with a preference for treatment. The effect of education is not monotonic. Those whose highest completed level of education is secondary school and those whose highest completed level of education is a tertiary qualification prefer no treatment over treatment, while those for whom the highest completed level of education was a trade certificate or diploma preferred treatment. Family structure is also relevant to preferences for treatment over no treatment. Individuals with partners prefer no treatment and individuals with dependent children are less likely to choose treatment over no treatment. Individuals who had been hospitalized in the previous five years prefer no treatment over treatment. This may reflect negative experiences of hospital care, or of the outcomes of treatment.

In Model 3, the "Generalised QALY" model, the coefficients on the probability terms differ for the treatment and no treatment alternatives, suggesting utility is state dependent. The coefficient on the probability of survival for treatment is not significantly different from one (95% CI: 0.67-1.23), consistent with expected utility. However, for no treatment, the coefficient on probability term is significantly greater than one, suggesting a non-expected utility model may be appropriate when evaluating outcomes involving poor health. The coefficient on the generic survival time term is less than one, suggesting a positive rate of time preference.

The failure to reject the restriction of zero coefficients on the cost terms in the genearlised QALY model provides weak support for separability of consumption, although the lack of significance may suggest that the cost attribute was poorly understood, or that the levels did not result in respondents making trade-offs. This is particularly the case given that covariates capturing income and wealth are significant.

Model 3a differs from Model 3 in that it imposes expected utility. The likelihood ratio test rejects this restriction. Model 4 imposes the restriction of no interaction between health state and survival time. This restriction is also rejected (Model 4 vs Model 3). Tests of coefficient restrictions for Model 4 reject both the hypotheses that the generic coefficient on survival time is equal to unity and that the coefficients on the probability terms for the treatment and no treatment alternatives are equal. Model 5 imposes the restriction of a generic probability coefficient across the treatment and no treatment alternatives, and Model 6 further imposes the restriction that this coefficient is equal to unity, imposing expected utility, as in the QALY model. Model 6 differs from the QALY model in that it allows for time preference, and allows valuations of health states to vary by individual characteristics. Comparisons of Models 5 and 4 and Models 6 and 5 reject these restrictions.

Model 7 imposes all the QALY restrictions, but allows individual specific covariates. In Model 8, all the QALY restrictions are imposed. Comparison of Models 7 and 8 show that including covariates significantly improves the model as was suggested by the significance of the individual specific covariates in all the other models. Given the logarithmic specification, the exponent of the treatment intercept (including the covariates) can be interpreted as a valuation of full health, and the exponent of each of the health state terms can be treated as a valuation of that health state relative to full health, under the constraint that the QALY restrictions are imposed. Inclusion of the covariates therefore changes the valuation of full health relative to poorer health states. Overall the comparison of the models leads to a rejection of the restrictions imposed by the QALY model.

Table 6 illustrates the impact of imposing the QALY restrictions in valuing outcomes of interventions. It presents the relative health state valuations as calculated from the estimated QALY model and the generalised QALY model. Because of the experiment design, these health state valuations are not directly comparable with the published EQ-5D tariffs, as they are constrained to lie between one (full health) and zero (death), with no worse-than-death health states allowed, and if there is a preference for treatment over no treatment, the no treatment health states will be scaled to reflect this. It is possible however to compare the health state valuations across models. The QALY model weights are presented in the first column of Table 6. As the QALY model does not include individual covariates the estimated QALY weights do not vary across individuals. They are calculated as the ratio of the exponent of the estimated coefficient for each health state to the exponent of the constant term under the treatment alternative, which can be interpreted as the value of full health.

Because the health state valuations in the generalised QALY model depend on individual characteristics, the generalised QALY weights in Table 6 are calculated for a particular individual (female, aged 50, annual household income of $50,000, reported own health as good, highest education level is secondary school, in a couple

with dependent children, renting, with no chronic health conditions, no private health insurance, has been hospitalised in the last five years). Further, because the generalised QALY model does not impose expected utility, allows state dependence, and includes interactions between health states and survival durations, the weights vary depending on the probabilities of survival under each alternative and on the survival duration.

Results are presented across three probability levels: high probability of survival in both alternative (Ptr=Pnt=0.99); low probability of survival in both alternatives (Ptr=Pnt=0.5); and different probabilities of survival under each alternative (Ptr=0.99, Pnt=0.50). In each case, the survival duration under treatment and non-treatment are equal, but are varied across four levels, between 10 and 40 years. The estimated generalised QALY weights are calculated as the ratio of the value of each health state relative to full health, which now depend on probability and survival duration.

For the worst health states (h1 and h2) the valuations monotonically decrease with duration, but for health states that are closer to full health, the valuations monotonically increase with duration. Comparing the high and low probability of survival cases, the generalised QALY weights are larger for higher survival probabilities. The QALY weights for the no treatment health states are reduced when the treatment survival probability is high relative to the no treatment survival probability. This variability in health state valuations demonstrates that imposing the QALY model restrictions distorts valuations.

7 Conclusions

This paper presents analysis from a discrete choice experiment that suggests that the QALY model, which is the dominant model in economic evaluation of health care, may not be a valid representation of individual preferences for health care. In particular, the results show that health state valuations vary with individual characteristics, with probability of survival and non-monotonically across survival durations. The utility associated with very poor health states may decline as survival duration increases. This suggests that the use of the simple QALY model in health care resource allocation may introduce distortions. The extent of these distortions is largely an empirical question and will vary across interventions, particularly in terms of how the intervention affects probability of survival, survival duration and quality of life. Health care interventions involve risky outcomes for individuals, and cost-utility analysis fails to take account of the impact of risk on individual welfare because it implicitly assumes that there is no risk at the population level. Further, applying a simple discounting factor that does not take account of the variation in valuation of survival duration across health states is unlikely to reflect preferences. The empirical results in this paper confirm the theoretical conclusions that the QALY restrictions are unrealistic (Bleichrodt and Quiggin 1999).

This raises the question of why cost-effectiveness has become the dominant approach to economic evaluation in health care. This partly relates to reluctance to place monetary values on health gains because this is considered ethically objectionable,

and partly to the perception that allocation of health care resources should be concerned with maximisation of health gain alone. However, the use of a cost-effectiveness threshold as a decision making criterion explicitly monetises health gain. Further, as Bleichrodt and Quiggin (1999)note, an implication of the QALY model is that the utility of a given health gain will be greater at higher levels of consumption. Thus, the QALY model does not avoid the ethical concern that higher income individuals place a higher values on a given health gain.

Table 1 : Attributes and Levels

Treatment (tr)			No Treatment (nt)		
Probability of Survival (P_{tr})	Life Expectancy (T_{tr})	Cost (C)	Probability of Survival (P_{nt})	Life Expectancy (T_{nt})	Health State (HS)
54%	10 years	$80,000	39%	10 years	H1
69%	20 years	$60,000	59%	20 years	H2
84%	30 years	$40,000	79%	30 years	H3
99%	40 years	$20,000	99%	40 years	H4

Table 2: Health States used in the discrete choice experiment[1]

	H1	H2	H3	H4	Full Health
Mobility	3	2	2	2	1
Self Care	3	2	1	1	1
Usual Activities	3	3	2	1	1
Pain/Discomfort	3	2	2	2	1
Anxiety/Depression	3	3	2	1	1

1. 1 represents the best level ("no problems" or "not" for that dimension), 2 represents an intermediate level ("some problems" or "moderate") and 3 represents the worst level ("extreme" or "unable to").

Table 3: Sociodemographic characteristics: sample proportions (n=301)

Male	43.5%
Female	56.5%
Married/defacto	61.2%
Dependent children	45.5%
Self reported health	
Excellent	38.9%
Good	50.1%
Fair	10.0%
Poor	0.3%
Chronic health condition	14.3%
Hospitalisation in last 5 years	21.0%
Highest education level	
Completed primary or secondary	38.5%
Trade Cert/Diploma	29.6%
Bachelor or higher degree	31.9%
Household income	
≤ $20,000	13.3%
$20,001-30,000	12.0%
$30,001-40,000	14.0%
$40,001-50,000	12.3%.
$50,001-60,000	12.0%
$60,001-$70,000	9.0%
$70,001-$80,000	5.7%
$80,001-$100,000	7.3%
>$100,000	14.6%
Private Health Insurance (hospital)	56.1%
Own/buying home	60.1%

Table 4: Tests of the QALY Restrictions

Model	Description	Restrictions	Log likelihood	Test	Test statistic	5% Critical Value (df)	Reject/ Accept
1	General model (RDEU)		-4966.81				
2	Consumption separable (RDEU)	$\beta_C = \beta_{C^2} = \beta_{C.T} = 0$	-4967.39	LR:2 vs 1	1.16	8.81 (3)	Accept
3	Generic survival time (RDEU)	$\beta_{Ttr} = \beta_{Tnt} = \beta_T$	-4968.01	LR:3 vs 2	1.24	3.84 (1)	Accept
3a	Generic survival time (EU)	$\beta_{Ptr} = \beta_{Pnt} = 1$ $\beta_{Ttr} = \beta_{Tnt} = \beta_T$	-4975.79	LR:3a vs 3	15.56	5.99 (2)	Reject
4	No time-health state interaction (RDEU)	$\beta_{T.HS_j} = 0$	-4973.55	LR:4 vs 3	11.08	8.82 (3)	Reject
				$\beta_T = 1$	207.77	3.84 (1)	Reject
				$\beta_{Ptr} = \beta_{Pnt}$	6.09	3.84 (1)	Reject
5	Generic Probability coefficient (RDEU)	$\beta_{Ptr} = \beta_{Pnt} = \beta_P$	-4976.60	LR: 5 vs 4	6.10	3.84 (1)	Reject
				$\beta_P = 1$	8.83	3.84 (1)	Reject
6	QALYs with time preference and covariates (EU)	$\beta_{Ptr} = \beta_{Pnt} = 1$ $\beta_{T.HS_j} = 0$	-4981.06	LR: 6 vs 5	8.92	5.99 (2)	Reject
				LR: 6 vs 3a	10.54	7.81 (3)	Reject
7	QALYs with covariates (EU)	$\beta_P = \beta_T = 1$	-5084.34	LR: 7 vs 6	206.56	3.84 (1)	Reject
8	QALYs (EU)	$\beta_z = 0 \; \forall \; z$	-5434.59	LR: 8 vs 7	700.50	25.00 (15)	Reject

Table 5: Results of Logit Estimation for Selected Models

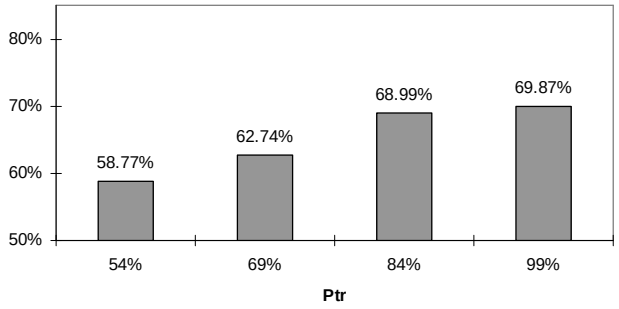
	Model 3 Generalised QALYs			Model 3a EU Generalised QALYs			Model 6 QALYs : Time Pref and Cov			Model 8 QALYs		
Variable	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
Intercept	3.886	1.046	0.000	2.058	0.749	0.006	2.037	0.749	0.007	1.351	0.045	0.000
Lnptr	0.955	0.141	0.000	-	-	-	-	-	-	-	-	-
Lnpnt	1.389	0.100	0.000	-	-	-	-	-	-	-	-	-
lntime	0.350	0.044	0.000	0.347	0.044	0.000	0.361	0.044	0.000	-	-	-
hsfx1	0.102	0.381	0.790	0.089	0.378	0.814	-0.754	0.063	0.000	-0.715	0.062	0.000
hsfx2	0.225	0.363	0.535	0.220	0.359	0.540	-0.184	0.058	0.001	-0.178	0.057	0.002
hsfx3	-0.049	0.344	0.887	-0.053	0.340	0.877	0.385	0.055	0.000	0.357	0.054	0.000
lntimehs1	-0.280	0.121	0.020	-0.271	0.120	0.024	-	-	-	-	-	-
lntimehs2	-0.133	0.115	0.247	-0.129	0.114	0.254	-	-	-	-	-	-
lntimehs3	0.141	0.109	0.194	0.139	0.108	0.195	-	-	-	-	-	-
tr_ovh_p	-2.086	0.599	0.000	-2.059	0.595	0.001	-2.070	0.596	0.001	-	-	-
tr_ovh_f	0.577	0.226	0.011	0.579	0.224	0.010	0.581	0.224	0.010	-	-	-
tr_ovh_g	0.988	0.212	0.000	0.970	0.210	0.000	0.972	0.211	0.000	-	-	-
tr_lnage	-1.941	0.182	0.000	-1.924	0.181	0.000	-1.917	0.181	0.000	-	-	-
tr_sexfx	0.047	0.048	0.320	0.047	0.047	0.321	0.046	0.047	0.326	-	-	-
tr_lninc	1.427	0.089	0.000	1.412	0.088	0.000	1.411	0.088	0.000	-	-	-
tr_ed_sc	-0.167	0.066	0.011	-0.164	0.065	0.012	-0.164	0.065	0.012	-	-	-
tr_ed_tdip	0.380	0.068	0.000	0.377	0.068	0.000	0.377	0.068	0.000	-	-	-
tr_couplfx	-0.220	0.060	0.000	-0.215	0.060	0.000	-0.215	0.060	0.000	-	-	-
tr_depkid	-0.269	0.051	0.000	-0.267	0.050	0.000	-0.267	0.050	0.000	-	-	-
tr_ownbuy	0.396	0.061	0.000	0.393	0.061	0.000	0.391	0.061	0.000	-	-	-
tr_phi	0.295	0.052	0.000	0.292	0.052	0.000	0.291	0.052	0.000	-	-	-
tr_chron	0.184	0.072	0.010	0.185	0.072	0.010	0.185	0.072	0.010	-	-	-
tr_hosp	-0.137	0.058	0.018	-0.140	0.058	0.015	-0.140	0.058	0.016	-	-	-
Log L	4968.01			-4975.7877			-4981.0601			-5434.5858		
Pseudo R^2	0.1835			0.1552			0.1543			0.0958		

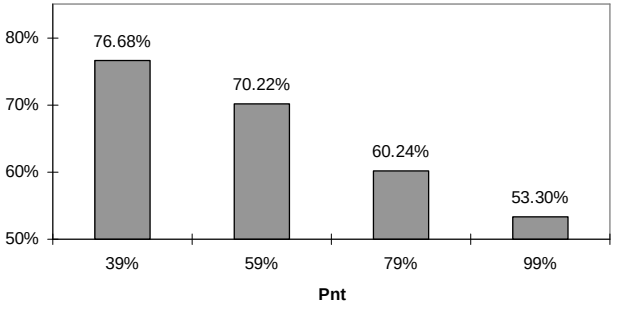
Table 6: Comparison of Health State Valuations under the QALY and Generalised QALY Models

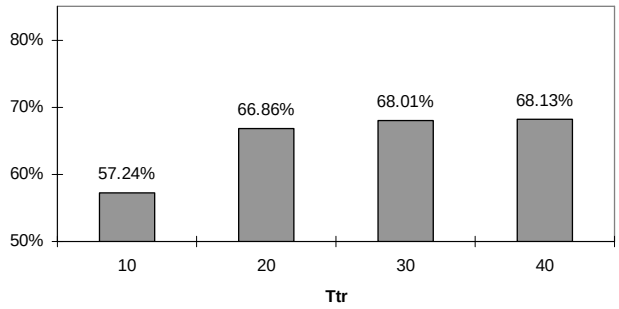
Health states	QALYs Model 8	Generalised QALYs Model 3											
		Ptr=Pnt=0.99				Ptr=Pnt=0.50				Ptr=0.99; Pnt=0.50			
		T=10	T=20	T=30	T=40	T=10	T=20	T=30	T=40	T=10	T=20	T=30	T=40
h1	0.127	0.183	0.151	0.134	0.124	0.136	0.112	0.100	0.092	0.071	0.058	0.052	0.048
h2	0.217	0.291	0.265	0.251	0.242	0.216	0.197	0.187	0.180	0.113	0.103	0.097	0.094
h3	0.370	0.416	0.458	0.485	0.506	0.309	0.341	0.361	0.376	0.161	0.177	0.188	0.196
h4	0.443	0.447	0.539	0.602	0.651	0.332	0.401	0.448	0.484	0.173	0.209	0.233	0.252

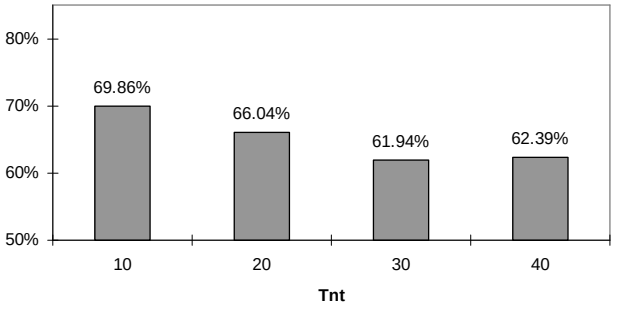
Figure 1: Plots of Marginal Frequencies

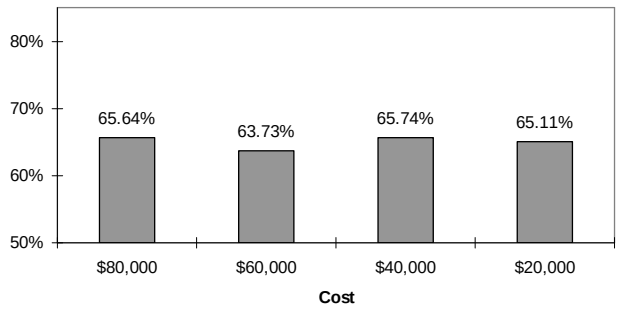
Treatment Attributes **No Treatment Attributes**

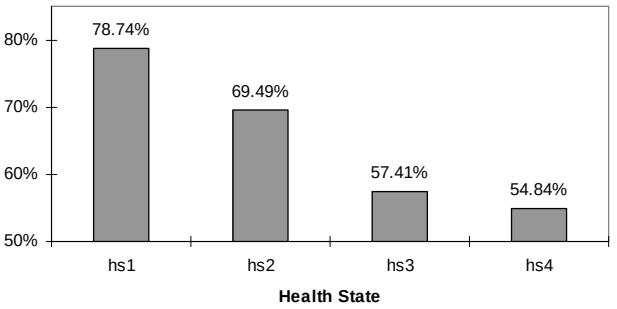












References

Ben-Akiva, M. and Lerman, S.R.,1985. Discrete choice analysis: theory and application to travel demand. MIT Press. Cambridge

Bleichrodt, H.,1995. QALYS and HYEs: Under what conditions are they equivalent? Journal of Health Economics 13, 17-37.

Bleichrodt, H. and Johannesson, M.,1997. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. Journal of Health Economics 16, 155-175.

Bleichrodt, H. and Johannesson, M.,1997. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. Medical Decision Making 17, 21-32.

Bleichrodt, H., Pinto, J.L. and Abellan-Perpinan, J.M.,2003. A consistency test of the time trade-off. Journal of Health Economics 22, 1037-1052.

Bleichrodt, H. and Quiggin, J.,1997. Characterizing QALYs under a General Rank Dependent Utility Model. Journal of Risk and Uncertainty 15, 151-165.

Bleichrodt, H. and Quiggin, J.,1999. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost benefit analysis? Journal of Health Economics 18, 681-708.

Bleichrodt, H., Wakker, P. and Johannesson, M.,1997. Characterizing QALYs by Risk Neutrality. Journal of Risk and Uncertainty 15, 107-114.

Dolan, P.,1997. Modeling valuations for EuroQol health states. Medical Care 35, 1095-1108.

Dolan, P.,2000. The measurement of health-related quality of life for use in resource allocation decisions in health care. Handbook of Health Economics Volume 1B. Culyer, A. and Newhouse, J. Amsterdam, Elsevier.

Drummond, M.F., O'Brien, B.J., Stoddart, G.L. and Torrance, G.W.,1997. Methods for the economic evaluation of health care programmes. Oxford University Press. Oxford; New York

Gyrd-Hansen, D.,2003. Willingness to pay for a QALY. Health Economics 12, 1049-1060.

Hakim, Z. and Pathak, D.S.,1999. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Economics 8, 103-116.

Hall, J., Kenny, P., King, M., Louviere, J., Viney, R. and Yeoh, A.,2002. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Economics 11, 457-465.

Johannesson, M.,1995. Quality-adjusted life-years versus healthy-years equivalents - A comment. Journal of Health Economics 14, 9-16.

Johannesson, M., Jonsson, B. and Karlsson, G.,1996. Outcome measurement in economic evaluation. Health Economics 5, 279-296.

Louviere, J.J., Hensher, D.A. and Swait, J.D.,2000. Stated choice methods: analysis and applications. Cambridge University Press. Cambridge, U.K.; New York

Manski, C.,1977. The Structure of Random Utility Models. Theory and Decision 8, 229-154.

McFadden, D.,1973. Conditional logit analysis of qualitative choice behavior. Frontiers of econometrics. Zarembka, P. New York, Academic Press**:** 105-142.

Miyamoto, J.M. and Eraker, S.A.,1988. A multiplicative model of the utility of survival duration and health quality. Journal of Experimental Psychology: General 117, 3-20.

Miyamoto, J.M., Wakker, P.P., Bleichrodt, H. and Peters, H.J.M.,1998. The Zero-Condition - a Simplifying Assumption in Qaly Measurement and Multiattribute Utility. Management Science 44, 839-849.

Pliskin, J., Shepard, D. and Weinstein, M.,1980. Utility functions for life years and health status. Operations Research 28, 206-224.

Quiggin, J.,1993. Generalized expected utility theory: the rank-dependent model. Kluwer Academic Publishers. Boston [Mass.]

Ried, W.,1998. QALYs versus HYEs - what's right and what's wrong. A review of the controversy. Journal of Health Economics 17, 607-625.

Slesnick, D.,1998. Empirical approaches to the measurement of welfare. Journal of Economic Literature 36, 2108-2165.

Spencer, A.,2003. The TTO method and procedural invariance. Health Economics 12, 655-668.

Thurstone, L.L.,1927. A Law of Comparative Judgement. Psychological Review 34, 283-286.

Torrance, G.,1986. Measurement of Health State Utilities for Economic Appraisal: A Review. Health Economics 5, 1-30.

Viney, R., Lancsar, E. and Louviere, J.,2002. Discrete choice experiments to measure consumer preferences for health and healthcare. Expert Reviews in Pharmacoeconomics and Outcomes Research. 2, 319-326.

Viney, R., Savage, E. and Louviere, J.,2005. Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Economics. 14, 349-362.