

Study Protocol: Comparison of Inconsistency between Time Trade Off and Discrete Choice Experiments in EQ-5D-3 L Health State Valuations

Sanjeeva Kularatna^{1,2}, Joshua Byrnes^{1,2}, Paul A. Scuffham^{1,2}

Abstract


Background: The EQ-5D-3L is the most widely used multi-attribute utility instrument for describing health states. A popular method for valuing the EuroQOL five dimension (EQ-5D)-3L health states is the time trade-off approach (TTO) where quality of life is traded against length of life. However, TTO valuations can provide logically inconsistent values. That is, where a respondent provides a utility value for one health state that is lower than the score they give for a logically worse health state. More recently there has been a tendency by researchers to use discrete choice experiments (DCE) as opposed to TTO in health state valuation exercises; however, DCEs often exclude dominant choices by design. The aim of this paper is to explore the differences in the rate of logical inconsistency health state valuations between TTO and DCE methodologies.

Methods: A representative sample of the Australian general population will be recruited from an online cohort. Of the 243 EQ-5D-3L health states, a number of health state sets, comprising of potentially logically inconsistent health state pairs, will be used for the valuation. Participants will be asked to value given health state sets using both TTO and DCE methods. Consequently, the proposed study is not a health state valuation exercise, but rather an evaluation of competing methods under controlled circumstances. Logical inconsistency will be assessed based on comparing quantitative health state valuations within the TTO and stated preferences from discrete choices within the DCE. The count of logical inconsistencies will be estimated at an individual level for both approaches and compared. The comparison of the two approaches will identify if there are significant differences between the number of logical inconsistencies produced from DCE and TTO methods.

Discussion: A difference in logical inconsistency is only one of many criteria for selecting the best approach for conducting health state valuations. It is recommended to examine the strengths and limitations of each methodological approach both theoretically and empirically.

This is novel research into important methodological concerns often overlooked up until now.

Keywords: logical inconsistency, time trade off, discrete choice experiments, utility, EQ-5D-3L

¹Centre for Applied Health Economics, School of Medicine, Griffith University, Queensland, Australia; ²Population and Social Health Research Programme, Griffith Health Institute, Griffith University, Queensland, Australia *Corresponding author  s.kularatna@griffith.edu.au

1. Background

Economic evaluations have an established role in reinforcing health care decision making in the developed world. Decision makers can make evidence based informed decisions based on health economic evaluations in prioritising health interventions and government subsidies. Many government advisory and decision maker institutions make use of the quality-adjusted life years (QALYs) model to make decisions.^{1,2} Use of QALYs as the outcome measure allows comparable benefits across different interventions and diseases. QALYs combine quantity and quality of life into a single metric. Specifically, quality adjustments between health states are accounted for with utility weights.³ These weights represent the preference of the population, for all given health states.⁴

Utility weights are the building blocks of the QALY model.^{5,6} Thus, how explicitly the utility weights represent the preference of the population the decision makers are presiding over is critical. The valuation procedure of health states to estimate utility weights in contemporary health economics is fairly well established.⁷ Pre-defined health states are valued by the general public by using preference elicitation techniques. Within preference elicitation, respondents typically trade quantity of life (or a risk of losing a quantity of life) for quality of life.

Health states are defined by multi attribute utility instruments (MAUI). MAUIs are standardised instruments that describe a health state with respect to different dimensions. Each dimension has different severity levels typically starting with “no problems”. There are two broad categories of MAUIs, generic or specific. Generic MAUIs are suitable for all patients focusing on core aspects of health-related quality of life. This is compared to disease specific MAUIs, which focus on providing greater detail of symptoms, side effects and aspects of functioning life as related to particular disease or class of diseases.

Whilst there exists a number of well-established generic MAUIs,⁸⁻¹¹ the EuroQOL five dimension (EQ-5D) questionnaire has become one of the most widely used and accepted MAUIs for defining health states for utility weight valuation. Moreover, the EQ-5D has the largest number of national value sets¹²⁻¹⁵ and has been specifically nominated as the preferred measure by National Institute for Health and Care Excellence in the United Kingdom.¹

Over the last two decades, starting from the Measurement and Valuation of Health study, EQ-5D valuations have used time trade-off (TTO) methods to elicit population preferences.¹⁵ In TTOS, time is directly traded off for quality. The TTO method has been preferred to other proposed methods, including standard gamble, on the grounds of completeness, logical consistency, construct validity and test-re-test reliability.¹⁶

Generally, an EQ-5D health state valuation using a TTO, a selected number of health states are directly valued by the respondents. Of the selected health states, one respondent could value 10 to 15 health states. Using the direct valuations, regression models are fitted to estimate the entire EQ-5D value set.

Within the sample of health states given to an individual to value, it is expected that worse health states are valued lower than better health states.¹⁷ However, individual preference, religious belief, education, employment, incomprehension and setting may elicit preferences from a respondent that value logically worse health states above better health states.^{17,18} This phenomenon is called logical inconsistency. It is debatable whether to include logically inconsistent valuations when estimating a national value set and the effect that their inclusion or exclusion has with respect to utility weight estimates and the impact on decision making.

There are a finite number of health state pairs that could lead to a logically inconsistent valuation from a respondent. The definition of logical inconsistency provided by Dolan *et al.*¹⁷ states that a health state can only be considered logically better than another health state if: a) at least one dimension is superior; and b) no dimension is inferior. For example, when health state X has some levels better and worse than health state Y, logical inconsistency cannot be predetermined. When state A is logically better on at least one dimension, and no worse on other dimension than state B, A should have a higher utility value than B. Otherwise if A is given a lower utility value than B from the same individual, the pair is defined as logically inconsistent. However, Lamers *et al.* argued that the difference between utility values should be more than 0.05 for them to be identified as logical inconsistency.¹⁹ Thus, if health states B have a utility value of higher than 0.05 over A, the A and B pair is considered logical inconsistency. The value of 0.05 is used as this is the level of sensitivity a utility value can achieve based on conventional TTO valuation. When a pair of health states are identified as being logically inconsistent, typically both health states (the better and worse state) are considered inconsistent.

More recently, discrete choice experiments (DCE) are becoming popular among health economists as an elicitation method for EQ-5D valuations.²⁰ Though long used in other fields, DCEs are gaining acceptance in health care research as a valid tool of measuring consumer preferences.²⁰ Its application in health state valuation is a relatively new and promising development. Within a DCE, a respondent is provided with two (or more) alternative health states described with respect to the domains of the MAUI and a period of time spent within each health state before dying. The respondent then selects which health state they would prefer. The process of presenting different alternative sets of health states is repeated for each respondent. Regression analysis is then used to estimate the marginal rate of substitution between time and the levels of severity for each domain used to describe the health state for the entire sample population.⁴ Unlike the TTO method a utility weight per health state is not directly elicited from each participant but rather modelled based on their selection of preferred health states.

Whilst logical inconsistency has been identified as a methodological quandary within TTO, logical inconsistency has not been explored within DCEs for health state valuations. Potentially, a DCE would be associated with lower levels of logical inconsistency because competing health states are presented to a participant side by side where as using the TTO method each health state is presented sequentially. In addition, dominant choices, that is, where a logically superior alternative is presented alongside a logically inferior alternative, are often removed from DCEs by design. Analysts remove such choice options based on grounds that this may decrease overall participant response to the DCE. Moreover, DCE and TTO may exert differing levels of cognitive burden upon participants which may influence participants' ability to avoid providing logically inconsistent responses. As such, it is important to examine the logical inconsistency associated with DCE preference exercises and moreover compare rates of logical inconsistency between TTO and DCE. Thus, the present study aims to examine the difference between logical inconsistency valuations using DCE and TTO preference elicitations for health states.

2. Methods

This is a prospective study using a representative sample of the Australian general population. Participants will complete both a TTO and DCE task to measure a sample of health states. The health states will be described using the EQ-5D-3L tool. All participants will value a selected set of health states which contain potentially logically inconsistent health state pairs. That is, to include within the set of health states, states which logically dominate other health states. Each participant will first rank the health states from best to worst health states and then conduct a TTO and DCE for these health states.

In order to overcome potential order bias, participants will be randomised to receive either the TTO or the DCE task first. The TTO and DCE methods will be compared against two measures. First, the ordering of health states as estimated by TTO and the DCE exercises will be compared to the rank ordering task. This will provide an overall identification of each method's ability to reflect the natural rank as provided by the participant, Secondly, rates of logical inconsistency derived from each methodology will then be compared between the TTO and DCE tasks.

Ethical clearance was obtained from Griffith University human research ethics committee (MED/05/14/HREC).

EQ-5D-3L

The EQ-5D-3L has been widely used to describe and value health states in a large number of countries including the United Kingdom,²¹ Germany,²² Australia,¹² United States,²³ Japan,²⁴ Argentina,¹³ Chile,²⁵ Thailand²⁶ and Sri Lanka.^{27,28} Moreover, the reliability and validity of the EQ-5D-3L as a MAUI has been widely established.²⁹⁻³¹ In total, the EQ-5D-3L has 243 health states making it relatively easy for valuation (for example the EQ-5D-5L has 3,125). A health state described using the EQ-5D-3L is relatively easy to understand for the participant as it has only five dimensions and three levels within each dimension. Using EQ-5D-3L, full health is described as 11111, stating there are no problems in mobility, personal care, usual activities, no pain/discomfort and no anxiety/depression. The worst imaginable health state according to the EQ-5D-3L is 33333. In 33333, a person is confined to bed, unable to wash or dress self, unable to perform usual activities, has extreme pain and discomfort and is extremely anxious or depressed.

Selection of health state for valuation

Out of the 243 EQ-5D-3L health states, a number of health state sets will be identified for the purpose of this study. Each set will consist of six health states that create 15 potentially logically inconsistent pairs as per the Dolan *et al.*¹⁷ definition. In addition to these six health states each participant will also value "the pits" (33333), the state of "immediate death" using the TTO approach.

Time Trade Off

The TTO approach involves asking subjects to consider the relative amounts of time they would be willing to sacrifice to avoid a certain poorer health state. For each health state, participants will first be given a choice between immediate death or living in the given health state for 10 years followed by immediate death,⁴⁰ to define which states are considered better/worse than death.

For states better than death, a value of 5 years will be offered in full health instead of 10 years in the given health state. According to each participant's response, participants will be offered further trade-offs using the outward titration approach.⁴⁰ In this approach, the time spent in full health will be increased or decreased in increments of 1 year. When respondents refuse a full year increase or accept a full year decrease, the midpoint of 6 months is then offered. As an example, if a respondent agreed to 3 years of full health instead of 10 years in the given health state and refused 2 years of full health, then 2.5 years of full health will be offered as the trade-off. In this study, as the focus is on logical inconsistency, time trade off calculations will be limited to 6-month increments. For states better than death, the TTO value is calculated as $x/10$, where x is the time the participant agrees to be spent in full health instead of 10 years in the given health state.

For states worse than death, an alternative of immediate death will be compared with a combination of living in the given health state and full health followed by death. The total duration will be 10 years. If the time spent in the given health state is y and time spent in full health is 10 years, the TTO value is calculated as $((y/10) - 1)$, ensuring negative values given by states worse than death will have a lower bound of -1 .

Next, logical inconsistency per participant will be identified. To achieve this, each dimension between each pair has to be compared according to the Dolan definition.¹⁷ That is the value given to state A should be lower than the value given to state B when state B is 'logically 'better on at least one dimension and no worse on other dimensions''.

Five dummy variables are created for the comparison between the two health states for each dimension. Each dummy variable equals 1 if health state two is worse, with respect to that particular dimension, and 0 if health state two is better. If the 5 dummy variables are not equivalent (i.e. 5 x 0, or 5 x 1) then the pair is deleted. Specifically, pairs with 5 dimensions with mixed 0 and 1 do not align with the logical inconsistency definition.

Real logical inconsistency observations will then be identified from the potential pairs of health states in which an logical inconsistency could exist. The logically better HS from each pair is first identified. The health state valuations for each health state within the pair are compared and if the logically better health state has a lower valuation logical inconsistency is confirmed. Following this, the recognised logical inconsistency observations are identified within the original dataset. A dummy variable with 1 for logical inconsistency observation is created for either HS associated with the inconsistent pair. For each respondent the numbers of inconsistent health states are then counted to identify those respondents with higher levels of inconsistent responses.

Discrete Choice Experiment

All pairs derived from the six health states, that comprise a health state set, will be presented to respondents as a discrete choice. Based on six health states per set, there will be 15 discrete choice experiments comprised of two health states (based on the combination formula (1); where n = number of health states; 6 and r = number included within the choice; 2). With each discrete choice experiment, the participant faces an opportunity to report a logically inconsistent choice. A logically inconsistent choice will be reported if a health state is selected as the preferred health state compared to a logically superior health state.

Sampling

Size

The precision of results with respect to the rate of logical inconsistency was considered to inform the required sample size. The number of possible health state pairs was calculated as 19,503 for the maximum plausible health states (198) used in EQ-5D valuations ($n = 198$, $r = 2$).

$$n!/(n-r)!(r!) \quad (1)$$

The sample size was calculated as 377 for 95% confidence, 5% confidence interval for a population of 19,503 total health states. However, the present study plans to collect data from a minimum of 1,000 participants to also include covariate analysis of routine demographic variables (age, sex, education).¹⁵

Selection and Recruitment

Participants will be recruited through an Australian online-cohort provider. Participants will be invited to complete the online tasks and in return will receive a financial reward managed by the online-cohort provider. Online cohorts are becoming a widely accepted source of recruiting survey participants. However, concerns have been raised with respect to the representativeness of this population to the general public. Firstly, online cohorts by their very nature have access to a computer (including tablets and smart phones) and the internet. Secondly, the implied hourly rate of reward for completing a survey is less than the average wage rate in Australia. As a result, open recruitment from such sources may result in a sample that is not representative of the Australian general public. Consequently, recruitment into this study will be restricted by predetermined quotas with respect to age, sex and income. Quotas developed with respect to the 2011 Australian national census based on these factors are provided in Table 1.

Table 1. Proportion Sample Distribution by Age, Sex and Income

	Income (\$AUD)	Age			
		<25	25-<45	45-<65	65+
Males	Negative/Nil	0.4%	0.3%	0.9%	0.2%
	\$1-\$299	1.2%	2.0%	4.3%	4.9%
	\$300-\$599	1.0%	2.9%	6.1%	5.8%
	\$600-\$999	0.8%	3.4%	4.8%	1.4%
	\$1,000-\$1,499	0.2%	2.0%	2.4%	0.3%
	\$1,500-\$1,999	0.0%	0.7%	1.1%	0.1%
	\$2,000+	0.0%	0.5%	0.8%	0.2%
	Not stated	0.0%	0.0%	0.0%	0.0%
	Total	4.0%	12%	20%	13%
Females	Negative/Nil	0.4%	0.5%	1.3%	0.2%
	\$1-\$299	1.2%	2.7%	5.2%	4.7%
	\$300-\$599	1.5%	5.0%	7.6%	7.3%
	\$600-\$999	0.5%	3.4%	4.1%	0.8%
	\$1,000-\$1,499	0.1%	1.2%	1.7%	0.2%
	\$1,500-\$1,999	0.0%	0.4%	0.8%	0.1%
	\$2,000+	0.0%	0.2%	0.3%	0.1%
	Not stated	0.0%	0.0%	0.0%	0.0%
	Total	4.0%	13%	21%	13%

Source: Derived from Australian Census 2011; viewed 5 June 2014; <http://www.abs.gov.au/websitedbs/censushome.nsf/>

Outcome Measures

The primary outcome measure is the number of logical inconsistency for each individual for both TTO and DCE valuations for the same set of health states. As a secondary outcome measure, the rank order of health states derived from the TTO, DCE and ranking task will be compared. Socio-economic characteristics will be explored as potential determinants of logical inconsistency using univariate and multivariate regression analyses.

Respondents who are identified as non-traders i.e. give the same value for all health states within the TTO analysis or those who select only the first or last option within the DCE are excluded from further analysis.

3. Discussion

Currently there are no published literatures comparing logical inconsistencies from TTOs with DCEs. People have varying preferences for different health states. These preferences can vary across individuals. However, if a person prefers a worse health state over a better health state it is questionable whether that preference should be considered for policy making decisions. The aim of this study is to ascertain the effects of valuation procedure on the logical inconsistencies using the same sample of respondents. The proposal put forwarded by the present study explores this question theoretically by making an individual value the same set of health states using both valuation methods at the same time point.

There are inherent differences between the two valuation methods. For example, the TTO method can produce individual valuations for each health state while DCEs estimate parameter values that explain utility and are then used to estimate health state values. However, we do not propose to measure the logical inconsistency in each valuation method in the same way. Identification of logical inconsistency from TTOs is essentially quantitative while from DCEs it is categorical. It does not make sense to use the same method of logical inconsistency identification for inherently different valuation methods. Our aim is to identify the number of logical inconsistency potentially made in each valuation method using unique technique for each. Moreover, the present study is not a health state valuation but only an exercise to find the variations of logical inconsistency in two different health state valuation methods on theoretical grounds. Our results can influence the selection of valuation method in future research. Therefore, the present method is a test under controlled circumstances to explore the use of valuation methods.

TTO and DCEs both have other methodological limitations and variations not explored in the present analysis. In TTOs, the trading off of a constant proportion of remaining life to improve the quality of life,³² assumption of liner utility for duration, bias of values downward, probability weighting, loss aversion and scale compatibility are limitations.³³ TTO values can be affected by discounting too.³⁴ Being less cognitively restrained in application is considered to be an advantage of DCEs. However, the ordinal nature of choice to make cardinal utilities raises theoretical arguments on the validity of DCE.³⁵ Moreover, the lack of uncertainty in TTO and DCE valuation could make these methods inferior to method such as standard gamble. Nevertheless, Buckingham *et al.*³⁴ argued that the valuation method should be judged by its ability to act as a proxy for utility rather than its capacity to model the situation being valued.

This proposed study examines the methodological variations of TTO and DCE on the grounds of logical inconsistency only. Whilst, this study cannot derive an optimal valuation method it provides substantial increase in the understanding of key methodologies used in health economics. Consequently, further advances in the valuation of health could be achieved thus ensuring greater optimality of resource allocation.

Declaration of Competing Interests

The authors declare that they have no competing interests.

References

- ¹ (NICE) NICE: **Technical guidance for manufacturers and sponsors on making a submission to a technology appraisal.** London: NICE; 2001.
- ² Harris AH, Hill SR, Chin G, Li JJ, Walkom E: **The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994-2004.** *Med Decis Making* 2008;**28**(5):713-22.
- ³ Drummond MF: **Methods for the economic evaluation of healthcare programmes.** Oxford Medical Publications 1999.
- ⁴ Norman R, Cronin P, Viney R, et al: **International comparisons in valuing EQ-5D health states: a review and analysis.** *Value Health* 2009;**12**(8):1194-200.
- ⁵ Scuffham PA, Whitty JA, Mitchell A, Viney R: **The use of QALY weights for QALY calculations: a review of industry submissions requesting listing on the Australian Pharmaceutical Benefits Scheme 2002-4.** *Pharmacoeconomics* 2008;**26**(4):297-310.
- ⁶ Pliskin JS, Shepard DS, Weinstein MC: **Utility functions for life years and health status.** *Operations Res* 1980;**28**(1):206-24.
- ⁷ Morimoto T, Fukui T: **Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals.** *J Epidemiol* 2002;**12**(2):160-78.
- ⁸ EuroQol Group: **EQ5D. EuroQol.** 2011. <http://www.euroqol.org/eq-5d/reference-search/reference-search.html>. Accessed September 15, 2013.
- ⁹ Brazier J, Roberts J, Deverill M: **The estimation of a preference-based measure of health from the SF-36.** *J Health Econ* 2002;**21**(2):271-92.
- ¹⁰ AQL: **Assessment of Quality of life.** 2011. <http://www.aqol.com.au/research-papers.html>. Accessed October 15, 2011.
- ¹¹ Feeny D, Furlong W, Torrance GW, et al: **Multiattribute and single-attribute utility functions for the health utilities index mark 3 system.** *Med Care* 2002;**40**(2):113-28.
- ¹² Viney R, Norman R, King M, et al: **Time trade-off derived EQ-5D weights for Australia.** *Value Health* 2011;**14**(6):928-36.
- ¹³ Augustovski FA, Irazola VE, Velazquez AP, Gibbons L, Craig BM: **Argentine valuation of the EQ-5D health states.** *Value Health* 2009;**12**(4):587-96.
- ¹⁴ Tsuchiya A: **Estimating an EQ-5D population value set: the case of Japan.** *Health Econ* 2002;**11**(4):341-53.
- ¹⁵ Dolan P, Gudex C, Kind P, Williams A: **The time trade-off method: results from a general population study.** *Health Econ* 1996;**5**(2):141-54.
- ¹⁶ Dolan P, Gudex C, Kind P, Williams A: **Valuing health states: a comparison of methods.** *J Health Econ* 1996;**15**(2):209-31.
- ¹⁷ Dolan P, Kind P: **Inconsistency and health state valuations.** *Soc Sci Med* 1996;**42**(4):609-15.
- ¹⁸ Wittrup KL, Jørgen T, Pedersen KM: **An assessment of inconsistencies in the valuation of hypothetical EuroQol (EQ-5D) health states.** 2008;University of Southern Denmark.
- ¹⁹ Lamers LM, Stalmeier PFM, Krabbe PFM, Busschbach JJV: **Inconsistencies in TTO and VAS values for EQ-5D health states.** *Med Decis Making* 2006;**26**(2):173-81.

- ²⁰ Norman R, Cronin P, Viney R: **A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states.** *Appl Health Econ Health Policy* 2013;**11**(3):287-98.
- ²¹ Dolan P: **Modeling valuations for EuroQol health states.** *Med Care* 1997;**35**(11):1095-108.
- ²² Bernert S, Fernandez A, Haro JM, et al: **Comparison of different valuation methods for population health status measured by the EQ-5D in three European countries.** *Value Health* 2009;**12**(5):750-8.
- ²³ Shaw JW, Johnson JA, Coons SJ: **US valuation of the EQ-5D Health States: development and testing of the D1 Valuation Model.** *Med Care* 2005;**43**(3):203-20.
- ²⁴ Tsuchiya A, et al: **Estimating an EQ-5D population value set: the case of Japan.** *Health Econ* 2002;**11**(4):341-53.
- ²⁵ Zarate V, Kind P, Valenzuela P, et al: **Social valuation of EQ-5D health states: the Chilean case.** *Value Health* 2011;**14**(8):1135-41.
- ²⁶ Tongsir S, Cairns J: **Estimating population-based values for EQ-5D health states in Thailand.** *Value Health* 2011;**14**(8):1142-5.
- ²⁷ Kularatna S, Whitty JA, Johnson NW, Scuffham PA: **Study protocol for valuing EQ-5D-3L and EORTC-8D health states in a representative population sample in Sri Lanka.** *Health Qual Life Outcomes.* 2013;**11**(1):149.
- ²⁸ Kularatna S, WA, Johnson NW, Scuffham PA: **Valuing EQ-5D Health States for Sri Lanka.** Submitted to *Quality of Life Research.*
- ²⁹ Augestad LA, Rand-Hendriksen K, Kristiansen IS, Stavem K: **Learning effects in time trade-off based valuation of EQ-5D health states.** *Value Health* 2012;**15**(2):340-5.
- ³⁰ Pickard AS, Wilke CT, Lin HW, Lloyd A: **Health utilities using the EQ-5D in studies of cancer.** *Pharmacoeconomics* 2007;**25**(5):365-84.
- ³¹ Lee CF, Ng R, Luo N, et al: **The English and Chinese versions of the five-level EuroQoL Group's five-dimension questionnaire (EQ-5D) were valid and reliable and provided comparable scores in Asian breast cancer patients.** *Support Care Cancer* 2013;**21**(1):201-9.
- ³² Bleichrodt H, Johannesson M: **Standard gamble, time trade-off and rating scale: Experimental results on the ranking properties of QALYs.** *J Health Econ* 1997;**16**(2):155-75.
- ³³ Jansen SJ, Stiggelbout AM, Wakker PP, et al: **Patients' utilities for cancer treatments: a study of the chained procedure for the standard gamble and time tradeoff.** *Med Decis Making* 1998;**18**(4):391-9.
- ³⁴ Buckingham K, Devlin N: **A theoretical framework for TTO valuations of health.** *Health Econ* 2006;**15**(10):1149-54.
- ³⁵ de Bekker-Grob EW, Ryan M, Gerard K: **Discrete choice experiments in health economics: a review of the literature.** *Health Econ* 2012;**21**(2):145-72.