A systematic review of the techniques used to value temporary health states

Roberts, Tracy; Jackson, Louise; Kinghorn, Philip; Ogwulu Chriscasimir, Chidubem

DOI:
10.1016/j.jval.2017.03.009

Document Version
Peer reviewed version

Citation for published version (Harvard):
Roberts, T, Jackson, L, Kinghorn, P & Ogwulu Chriscasimir, C 2017, 'A systematic review of the techniques used to value temporary health states', *Value in Health*, vol. 20, no. 8, pp. 1180-1197. https://doi.org/10.1016/j.jval.2017.03.009

A systematic review of the techniques used to value temporary health states

Funding source: No funding received

Keywords: Health State Utility Values, Temporary Health States, Systematic Review, Valuation technique

Word count: (4362); Tables 2; Figures 1; Box 2; References 81

Highlights:

- Relatively little attention is given in the health economics literature to the valuation of temporary health states.
- Complications associated with the valuation of temporary health states have previously been the reason given for the use of cost-effectiveness analysis rather than cost-utility analysis (preferred by many decision-making bodies) for evaluation in contexts such as sexual health.
- Although many studies reviewed suggested that chained standard gamble and time trade-off would be the ideal approaches for temporary health state valuation, they were often not used due to fears around the cognitive and time burden on respondents.
- Where chained approaches were used they, they were found to be feasible, less susceptible to bias and able to detect minimal changes in health states.
- Future research should focus on anchoring for chained standard gamble and time trade-off.

ABSTRACT

Objective: A broad literature on health state utility values (HSUVs) exists, but compared with chronic health states, issues surrounding the valuation of temporary health states have been poorly explored. This review aimed to assess the methods used by previous studies to value health states that are considered temporary, in order to determine the strengths and limitations associated with various approaches and inform future study designs.

Methods: A systematic review was undertaken to explore the methods used, assess how the valuation was conducted for diseases that might lead to health states deemed as temporary and identify the challenges encountered in the valuation of temporary health states.

Results: Of the 36 relevant studies, 22 were explicit that the health state being valued was temporary. Most of the studies used more than one technique (often incorporating both conventional and adapted approaches). In using adapted techniques, the primary challenge was identifying an appropriate intermediate 'anchor' health state and the possibility of negative utilities.

Conclusions: There is no agreement on the most methodologically robust approach to valuing temporary health states. Valuation is complex and important issues relating to the validity, practicality and reliability of the techniques used, were not adequately covered by most of the studies identified.

Keywords: Health State Utility Values, Temporary Health States, Systematic Review

INTRODUCTION

Economic evaluations are conducted to provide evidence on the cost-effectiveness of interventions and inform decisions on the allocation of scarce healthcare resources [1]. In many countries, including the United Kingdom (UK), decision-making bodies require interventions to be evaluated using quality-adjusted-life-years (QALYs) [2,3]. QALYs combine changes in quality-of-life (QoL) – reflecting individuals' relative preferences for health states (HS) and life-expectancy [4,5]. HS valuation studies assign numerical values known as health state utility values (HSUVs) to HS descriptions and they are essential to allow economic evaluations to be undertaken [6,7].

The methodology for valuing chronic HS has been discussed extensively [8-12]. A broad literature exists on HSUV estimates for a diverse range of chronic conditions, including obstructive lung diseases [13], diabetes [14,15], chronic mental illnesses [16-18] and musculoskeletal disorders [19,20], that can be incorporated into economic evaluations [6]. However, for some clinical conditions such as pregnancy-related complications [21-24] and sexually transmitted infections (STIs) [25-27], preference-based measurements for HSUVs are less widely researched or are perceived as more challenging [25,28,29], possibly because these conditions involve temporary health states (THS) [30]. It has been noted, for example, that in the economic literature on STIs, most studies report results in terms of cost per major outcome averted, because no robust values for QALYs exist [31-33]; those who have highlighted this trend cite the temporary nature of the HS as one of the key complicating factors.

In contrast to chronic HS whose durations are individuals' life-expectancies [34], THS are states described as lasting for a specified length of time (weeks, months or years), followed by a return to full health [30,34]. It has been argued that conventional approaches to valuing outcomes or measuring QoL such as Time trade-off (TTO), Standard gamble (SG) and Visual analogue scale (VAS) are not appropriate for such HS due to their underlying assumptions [35]. There are techniques designed specifically for THS valuation, which are *adapted* from the conventional methods (Box 1). These include chained approaches for TTO and SG, waiting time trade-off (WTO), and sleep trade-off. There is currently no 'gold standard' [30] from this range of possible approaches, and THS valuation has been

discussed less comprehensively in the literature [25,30,36]. Studies that have assessed THS have not always been explicit about the fact that the HS being valued is of temporary nature [26,27].

This paper reports findings from a systematic review of studies that valued diseases/conditions where the resulting HS can be defined as temporary. Two principal questions were explored, 1) what methodological approach was undertaken in the valuation of the HS, and 2) how was the tension between the temporary and chronic health components of the HS resolved or addressed? In this paper, THS is defined as any health condition which causes some impairment to QoL, lasting for a duration of one year or less, after which there is return to normal health [30]. For the purposes of this review we identified studies that either explicitly described the HS under consideration as temporary or where the reviewers assessed that the nature of the diseases or HS being considered is of a temporary nature. Thus even if the authors of the papers used a time period in the valuation exercise of longer than one year, it was still included in the review if the reviewers assessed the HS under consideration to be temporary based on the nature of the disease/HS.

By answering these questions, the review will help to identify an appropriate approach for the valuation of THS. The outcomes will inform the literature on the methods that should be used for estimating HSUVs and improve the robustness of the data used in economic evaluations involving THS.

METHODS

The review followed the guidelines of the UK Centre for Review and Dissemination (CRD) [43] and is reported following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [44].

Inclusion criteria

Papers published in English were included if the disease involved a THS, and the study included the valuation of THS. There was no design and year restriction in order to capture as many studies as possible and explore a range of methods. Studies were excluded if they were editorials, reviews or economic evaluations based on secondary data, or if they reported on HS valuations that did not include any HS which could be viewed as temporary.

Search Strategy

The following seven databases were searched: MEDLINE, Web of Science, EMBASE, Applied Social Sciences Index and Abstracts (ASSIA), ProQuest (minus ASSIA), NHS Economic Evaluation Database and Health Management Information Consortium (HMIC) (Appendix A). The reference lists of key papers were hand searched to identify other relevant studies. Where full texts of papers were unavailable, the study authors were contacted to request a copy.

Information from the database search was managed using RefWorks [45]. Paper selection was conducted by all authors in two stages based on methods described elsewhere [25,46]. In stage 1, studies were screened and categorised on the basis of title, keywords and abstracts into seven groups (A to G) (Appendix B). For stage 2, the full texts of potentially relevant studies were read and classified further according to the methods adopted (Appendix C).

Data extraction

A data extraction form was used to extract data on study background, valuation method, participants and other relevant details. The information was tabulated to facilitate detailed data comparison.

Analysis

A narrative review was undertaken to summarize the evidence [43] and to compare results and methods across studies and the likely reasons for any inconsistencies. A formal quality appraisal was not carried out as there are no agreed criteria to appraise studies of this nature. However, to inform the analysis, we developed a framework based on an adapted version of a checklist reported by Brazier et al. [47,48]. Brazier et al.'s original checklist reviewed the application of the psychometric criteria of validity, practicality, and reliability of measures of health-related quality of life (HRQoL). Three of the criteria in the checklist were modified slightly to fit this study's focus on THS valuation techniques, and an additional item was included to capture the issues associated with THS valuation (*Box 2*).

RESULTS

Systematic search

The database search identified 9,186 citations of which 3,449 were duplicates. A flow diagram of studies selected, excluded or retained is shown in Figure 1. Stage 1 of the categorisation (Appendix B) process led to the selection of 69 papers for further review. In the second stage of the process (Appendix C) 67 studies classified as being in groups A, B and C were read in full and further categorised into six groups (1-6). Studies in groups A1, A2, B1, B2, and C1 (Appendices B and C) were selected for inclusion in the narrative synthesis as they focused on the valuation of THS and relevant data were extracted from these 36 studies.

Summary of selected studies

The characteristics of the 36 studies are summarised and are presented in Tables 1 and 2. Sixteen studies used data collected from respondents in the United States [26,27,39,41,49-60]. Eight studies were UK based [61-68], six were conducted in The Netherlands [69-74] and three in Canada [75-77]. Two of the studies were carried out in two or more countries [78,79], while one study was conducted in Spain [80]. The studies' aims were diverse and included: the elicitation of HSUVs for specific conditions; assessing the feasibility and reliability of particular valuation techniques and comparing utilities among different populations. The HS covered by the papers were mostly related to treatments for various forms of cancers while two studies assessed pelvic inflammatory diseases (PID) [26,27].

Valuation technique

Twenty-two studies were explicit about the fact that the HS they were concerned with are THS . The remainder were not explicit but were judged to be concerned with the valuation of THS by the authors of this review (Table 2). Ten studies evaluated both chronic and temporary HS within a single study [26,27,50,53,57-59,61,62,66]. A variety of approaches were taken to accommodate this. Four papers used the same duration for all HS and applied the same technique justified on the basis of consistency [26,27,59,66]. Three papers used the same technique but adjusted the duration [58,61,62]. The remaining three used different techniques for different HS [50,53,57].

Thirty-one studies used more than one technique to value THS while only five used a single technique [39,57,61,62,64]. Overall, 12 papers used conventional methods such as SG and TTO to value THS [26,27,51,55,57,59-62,68,75,80]. Twenty-four papers used a variety of adapted methods which were specifically designed for THS. Twelve of these papers used the chained approach with TTO [63,73,79], SG [52,54,64,67,76,77] or both [65,70,72]. Eight studies used some modification of the conventional methods including SG [78] and TTO [53,56,58,66,69,71,74]. Three of the studies used WTO [39,49,50], while one used sleep trade-off [41] (Table 1).

Thirty studies specified the duration of the HS under consideration to the respondents. This ranged from two hours, for a study related to the side effects of chemotherapy in breast cancer patients [57], to 50 years for a study assessing utilities for acute PID [27] (Table 2). Three studies did not state the HS duration they used [51,52.74].

Justification for choice of method for THS

Among the 12 studies that used conventional methods (typically applied to chronic HS), to value THS, only four justified their choice of approach (Table 2). Three of the studies used VAS and/or TTO because they felt it was easier [66], less tedious [80] and involved less cognitive burden [57] than the chained appoach. The fourth study, which used both VAS and SG stated that the methods were used because the chained SG is cumbersome [75] and I could mpose an additional response burden on participants. Badia et al. [80] argued that chained methods are more complicated and would increase study costs.

Among the 24 studies that used adapted methods, 20 papers justified their choice. Five of the papers that used modified TTO and/or SG, cited the temporary nature of the HS being assessed [56,58,69,72] and the inappropriateness of classic conventional methods for THS valuation [74] as reasons for this. Brazier et al [66] used a modification of TTO [81], because they felt that the chained approach was likely to generate biased estimates. For the remaining 13 papers in which the chained approach, WTO or Sleep trade-off were used, all cited the temporary or mild nature of the HS being assessed as reasons for adopting the technique.

The second prevalent reason for using adapted techniques was to avoid the utilization of ‘dead’ as an anchor (five studies). In one case, the authors [76] felt that some THS might be tolerable enough such

that the HS ‘dead’ would not be an acceptable alternative, while others-[64,67] believed that the mild and relatively healthy nature of the HS meant that ‘dead’ was an unfeasible end-point. Moreover, they felt that respondents would be reluctant to gamble between such mild HS and ‘dead’ [67]. Another study [77] went further and stated that they avoided the use of ‘dead’ because they aimed to enable decision-making that would be acceptable to those who valued life above all else.

Three of the papers set out to assess the feasibility of the chained approach for THS valuation and hence focused on this method [54,70,73]. Johnston et al. [63] used the chained TTO because they felt that it was easier for participants than SG.

Challenges associated with THS valuation

Fourteen papers discussed the potential challenges associated with THS valuation (Table 2).

Chained approaches

Seven papers discussed the challenges encountered with the chained approach. A prevailing problem involved selecting a suitable intermediate (worst temporary) HS to use as an anchor in the first step of the chained approach. Johnston et al., [63] used chained TTO and found that despite carefully describing the intermediate HS, some respondents valued them as better than being in good health. Jansen et al., [70] used chained TTO and SG and also reported that some patients preferred the intermediate HS to some THS. This unexpected occurrence made the intermediate HS unsuitable for use in utility calculation for those respondents. In choosing an intermediate HS, Matza et al. [67] also reported that this was difficult due to differences in respondents’ risk aversion and their perception of the most severe THS.

The second challenge associated with THS valuation was the possibility of negative utilities being assigned to some HS by respondents. For example, Jansen et al., [70] reported a zero valuation for the intermediate HS by 59% of the respondents for TTO and 36% of respondents for SG. This was echoed by Locadia et al. [73] who noted that if negatives values were allowed, some respondents’ HSUVs would have been less than zero.

Some papers reported challenges in relation to the state of being ‘dead’ which was used in the second step of the chained approach [63,67,73]. One study, [63] found it difficult to value a THS on a scale that

included death in the short term, as respondents' values were highly sensitive to the inclusion of death in the valuation question. Another study [73] found a downward bias (values converging at lower utilities) among respondents using a conventional TTO method and during the second stage of the chained TTO when they were told that the THS was followed by the dead state.

Deciding on the absolute and relative durations to use for THS was challenging for one study [76]. The choice of duration is important as some papers reported that a relatively short duration could result in an upward bias (values converging at lower utilities) [67,73].

Modified techniques

Two studies [56,66] modified the conventional method in terms of duration, and both stated that their respondents found it difficult to understand the TTO tasks. One of the studies [66] felt that the self-completed format of their modified TTO was largely to blame for these difficulties.

Waiting Time Trade-Off

The three studies that utilised WTO to value THS highlighted several challenges with this method [39,49,50]. Firstly, for THS to be calculated, HSUVs for the condition of interest, or its absence or successful treatment are required [39]. Thus, the WTO studies, either did a literature search [50] or a primary study on patients using conventional TTO [39,49] to get these values. Secondly, the authors reported that WTO is suitable only for HS with a very short duration, as the utility values tend towards zero and negativity as the duration lengthens. Thirdly, the technique is difficult to apply in acute events and requires the presence of the disease symptoms for which the patient is being tested [49].

Conventional approaches

Three of the papers that used conventional methods discussed the challenges they encountered. Badia et al. [80] reported participants' reluctance to trade time in a worse HS, for time in a better HS, leading to many HS being assigned identical levels once aggregated and thereby invalidating the method. Smith et al., [26] used the conventional approach for chronic and THS and reported that the values generated for THS were not be ideal and that a chained approach would have been better. Finally, Dion et al., [75] used

VAS for THS valuation and SG for other HS and reported that, unexpectedly, some respondents found VAS more difficult than SG.

Health state development

The studies used a combination of approaches to develop HS descriptions. Most common was the use of expert opinion (23 studies), followed by literature reviews (15 studies) (Table 1). Only 12 studies used the experiences of patients, and five of them elicited patients' experiences through focus group discussions (FGD) and in-depth interviews. Three studies [61,62,80] selected HS from a subset of existing EQ-5D HS descriptors while one study [64] used previous guidelines. Nine of the papers were not explicit about how they developed HS descriptions for their studies.

Validity, practicality, and reliability of studies

The majority of the studies included in the review did not cover issues relating to the validity, practicality and reliability (Box 2) of the valuation techniques used in their reporting (Appendix D).

Validity

Among the studies that used more than one technique to value HS 22 included both choice based and non-choice based methods. For the studies that used just one technique, three [39,64,66] used a choice based technique and two [57,62] utilised VAS, a non-choice based technique. 15 papers commented on the respondents' understanding of the task, but only seven studies provided evidence to justify the points made. Five studies reported that they had evaluated respondents' understanding by assessing the difficulty level of the exercise [41,65,70,71,80], task acceptability [77], or the number of outliers during their data analysis [60]. The respondents' backgrounds were reported as representative of the population group being assessed in nearly all of the studies (30 out of 36 studies). Although 29 studies presented a full description of the HS scenarios used, the remaining studies did not describe them in detail. The theoretical validity of the valuation techniques was discussed explicitly by just nine papers [49,51,58,63,64,71,72,77,80]. Only five papers that assessed empirical validity, with one [80] examining the number and proportion of inconsistencies between the original rank order respondents assigned to the HS and the ordering derived from the same HS from their valuations.

Practicality

Twenty-six studies appraised the practicality of the techniques used in terms of the duration of the task, the response rate, and the completion rate. The time it took respondents to complete the task was reported by 11 papers. The response rate was reported by 17 of the selected papers. The data captured was assessed by 19 studies using the completion rate and/or the level of missing data. Fifteen studies presented details of the completion rate while two papers presented only the missing values [62,70] and another two [78,80] presented values for both the completion rate and the missing data.

Reliability

Among the six papers that reported on reliability, two [75,77] assessed the test-retest reliability, one [73] used correlation coefficient and two [41,71] stated that the techniques they had used were reliable but were not explicit about how they had reached this conclusion. One study used the TTO technique which they modified for duration and reported that they found the technique unreliable although no further details were included [56].

Discussion

Summary of main findings

The review identified and assessed studies that conducted valuations of THS in order to inform future economic evaluations involving such HS. Twenty-two studies highlighted themselves that some of the HS they were concerned with were temporary. The remaining 14 studies were deemed by the current authors to include the valuation of THS [Table 2]. Some studies that used conventional valuation methods included long durations for the valuation exercise [26,27,57,59,60,68]. Three acknowledged this as a limitation [26,27,59] and three justified this in that respondents would not trade off against death if they used a short duration in the valuation exercise [59,68,75]. However, four studies despite being explicit about the temporary nature of the conditions, and used adapted techniques, used a long duration [66,69,71,79] as they felt it represented a realistic life expectancy.

In the literature on HS utilities valuation, trading-off time is not an integral component of the VAS technique but it is integral to the TTO and SG techniques. It was clear that the duration of the HS used with SG and TTO in the studies was not specified to respondents in the case of the VAS, except in a

small number of cases where authors explicitly. stated that this *was* the case [63,75]. A majority of the papers merely used the VAS as a warm-up task to familiarise the respondents with the HS descriptions.

Thirty-one papers used more than one technique to conduct the valuation. The majority of studies (24 of 36) opted to use an adaptation of conventional methods of HS valuation due to the temporary nature of the HS of interest. The prevailing challenge in using the adapted techniques was the identification of an appropriate intermediate 'anchor' HS. Where the study included consideration of both temporary and chronic HS, a variety of approaches was adopted and there was no consensus amongst authors [26,27,50,57-59]. Although a number of studies used more than one method for valuation measurement, the prevalent technique, used by nearly all papers was VAS; this was typically used as a warm up or HS familiarisation exercise. and to identify the most appropriate HS to use as an anchor. The VAS is the least grounded in economic theory and does not involve any element of risk or uncertainty; hence it generates values and not utilities. [57].

The commonest technique used by the papers for THS valuation was the chained approach. The majority of papers argued that either chained TTO or chained SG was the most appropriate technique for the valuation of THS. However, even though there was popular agreement on its appropriateness, 15 of the papers did not use it because of its perceived complexity and cost: it involves extra steps compared to the conventional methods However, other studies showed that the chained methods are feasible, less susceptible to bias and can detect minimal changes in HS [65,70]. They are seen as psychologically less threatening to respondents because the state of being dead is replaced by an intermediate HS [70]; this makes it more suitable for THS valuation where a return to normal health is common. There was a lack of consensus amongst study authors over the more appropriate of the two techniques, which reflects wider discussions in the literature [47]. Where both approaches were used in the same study [65,70], there was no systematic difference between them. It has been suggested that the technique chosen should reflect the situation which respondents are likely to encounter in reality, either a situation of risk (for SG) or certainty (for TTO) [65]. Furthermore, it has been proposed that when trading length of life against QoL, TTO is more appropriate, while in a situation where there is a risk of impending death SG will be more suitable [82].

Papers using chained approaches for valuation reported some challenges with the most common relating to the choice of the anchor HS. The intermediate HS plays a central role in the chained approach. In selecting an intermediate HS, it is important that the HS can be broadly applied and compared across different scenarios. . Finally, very few studies reported on aspects such as validity and reliability (Appendix D)

Previous research

A previous study reviewed the techniques used for THS valuations for cost utility analysis (CUA) [30]. However, the study focused on the general advantages and disadvantages of each technique, rather than systematically reviewing the methods which had actually been used in practice. A study by Brazier et al.[47] examined the conventional methods used for the valuation of HS, however, the study included all HS and did not specifically focus on the issues relating to the valuation of THS.

The current study adds to the existing literature by systematically investigating which methods have been used to value THS and exploring wider aspects such the justification for choosing a method and the challenges encountered. These issues have not been addressed in previous studies.

Strengths and weaknesses

The main strength of this review is that it is the first that has explicitly assessed the different techniques used for THS valuation in the literature in terms of the technique used, the justification of the choice of technique and the challenges highlighted by the study authors. Another strength is the extensive and systematic search for relevant studies. The inclusion of a broad range of study types without any restrictions on the year of publication allowed a deep assessment of the approaches that have been used previously to value THS. The use of an adapted checklist [48] to inform the analysis is also a strength. This is the first time such a criteria has been used for adapted methods of HS valuation.

However, as is expected in any study, there were also limitations associated with the review. The scarcity of selected papers that reported on the feasibility, validity and reliability of their techniques made it difficult to assess the robustness of the approaches used. There was not sufficient evidence in the papers to draw conclusions regarding validity and reliability of adapted methods. However much of the discussion

about the use of adapated methods relates to their feasibility and the papers' findings show that these methods are feasible in a wide range of population groups.A second weakness relates to study selection; some studies did not explicitly state that the condition they were assessing included THS, thus a judgement needed to be made by the authors of this review. This might have led to some papers being missed. Finally, a pragmatic decision was made about search terms used. We scoped, consulted experts and refined terms and some terms like 'acute' were considered but ultimately deemed unhelpful to the review.

Implication for current practice and future research

This review has highlighted the methodological challenges and lack of attention given to the valuation of THS. Nonetheless, it has also identified a combination of approaches that have been successfully used for the valuation of such outcomes.

There is a need for more research on the appropriate use of THS techniques on the reliabilty and validity of the techniques. The area is still sparsely explored with most of the literature published more than 15 years ago, presumably under the assumption that the discipline has matured to reach equilibrium given the requirement of decision making bodies to use QALYs [5]. But the clinical areas that are currently devoid of appropriate HSUVs primarily because of the presence of THS need attention in order to appropriately inform robust resource allocation decisions [25]. Criteria for the quality appraisal of methods adapted specifically for THS measurement also needs attention.

Conclusions

This review has shown that in measuring HS utilities for different conditions the majority of studies employed more than one technique. There is no clear agreement on the most methodologically robust approach to valuing THS. Valuation is complex in terms of anchoring, and there is currently mixed and incomplete evidence on the level of comprehension by respondents. Finally, important issues relating to the validity, practicality and reliability of the techniques used, were not adequately covered by most of the studies identified.

References

1. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart, D. *Methods for the Economic Evaluation of Health Care Programmes* (3rd ed). Oxford: Oxford University Press, 2005.

2. Claxton K, Walker S, Palmer S, Sculpher M. Appropriate Perspectives for Health Care Decisions. CHE Research Paper 54 http://www.york.ac.uk/inst/che/pdf/rp54.pdf. York: Centre for Health Economics, University of York 2010.

3. National Institute of Care Excellence (NICE). *Guide to the methods of technology appraisal 2013.* Available from: https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf [Accessed June 5th, 2015].

4. Weinstein MC, Torrance G and McGuire A. QALYs: The Basics. *Value in Health* 2009; 12 (1).

5. NICE. *Guide to the processes of technology appraisal 2014*. Available from: http://www.nice.org.uk/article/pmg19/resources/non-guidance-guide-to-the-processes-of-technology-appraisal-pdf [Accessed October 28th, 2015].

6. Tengs T, Wallace A. One thousand health-related quality-of-life estimates. *Med Care* 2000; 38:583-637.

7. Tolley T. *What are health utilities? What is.....? Series. 2009*. Available from: http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Health-util.pdf [Accessed August 29th, 2015].

8. Bass EB, Steinberg EP, Pitt HA et al. Comparison of the rating sale and the standard gamble in measuring patient preferences for outcomes of gallstone disease*. Medical Decision Making* 1994; 14: 307-341.

9. Cook J, Richardson J and Street A. A cost-utility analysis of treatment options for gallstone disease: Methodological issues and results. *Health Economics* 1994; 3: 157-168.

10. Dolan P, Gudex C, Kind P and Williams A. Valuing health states: A comparison of methods. *Journal of Health Economics* 1996; 5: 209-231

11. Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decision Making.* 2006; 26(2):173-81.

12. Brazier J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics* 2008; 26:769-779.

13. Bereza BG, Nielsen AT, Valgardsson S, Hemels MEH, Einarson TR. Patient preferences in severe COPD and asthma: a comprehensive literature review. *Int J Chron Obstruct Pulmon Dis* 2015; 10: 739–744.

14. Poku E, Brazier J, Carlton J, Ferreira A. Health state utilities in patients with diabetic retinopathy, diabetic macular oedema, and age-related macular degeneration: a systematic review*. BMC Ophthalmology* 2013:13:74.

15. Kennedy-Martin T, Paczkowski R, Rayner S. Utility values in diabetic kidney disease: a literature review. *Curr Med Res Opin* 2015; 31(7):1271-82.

16. Alonso J, Angermeyer MC, Bernert S. et al. Disability and quality of life impact of mental health disorders in Europe: results from the European Study of the Epidemiology of Mental Health Disorders (ESEMeD) project. *Acta Psychiat Scand.* 2004; 109(Suppl 420):38–46.

17. Brazier J. Measuring and valuing mental health for use in economic evaluation. *J Health Serv Res Po*. 2008; 13 (Suppl 3):70–75.

18. Roberts J, Lenton P, Keetharuth AD, Brazier J. Quality of life impact of mental health conditions in England: results from the adult psychiatric morbidity surveys. *Health and Quality of Life Outcomes*. 2014; 12:6.

19. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring Health-related Quality of Life in Rheumatoid Arthritis: Validity, Responsiveness, and Reliability of EuroQol (EQ-5D) *British Journal of Rheumatology* 1997;36:551–559.

20. Brunner H, Maker D, Grundland B. et al. Preference-based measurement of health-related quality of life (HRQL) in children with chronic musculoskeletal disorders (MSKDs) *Med Decis Making* 2003; 23(4):314-22.

21. Mogos MF, August EM, Salinas-Miranda AA, Sultan DH, Salihu HM. A Systematic Review of Quality of Life Measures in Pregnant and Postpartum Mothers. *Applied research in quality of life*. 2013; 8(2):219-250.

22. Hill PD, Aldag JC. Maternal perceived quality of life following childbirth. Journal of Obstetric, Gynecologic, and Neonatal Nursing. 2007; 36(4):328–334.

23. Bijlenga D. Health-related quality of life after induction of labor versus expectant monitoring in gestational hypertension or preeclampsia at term. Hypertension in Pregnancy. 2011a; 30(3):260.

24. Huang K, Tao F, Liu L, Wu X. Does delivery mode affect women's postpartum quality of life in rural China? J Clin Nurs. 2011;25

25. Jackson L, Auguste P, Low N, Roberts T. Valuing the Health States Associated with Chlamydia trachomatis Infections and Their Sequelae: A Systematic Review of Economic Evaluations and Primary Studies. *Value in Health* 2014; 17; 116 – 130.

26. Smith KJ, Tsevat J, Ness RB, Wiesenfeld HC, Robberts MS. Quality of Life Utilities for Pelvic Inflammatory Disease Health States. *Sexually Transmitted Diseases* 2008;35 (3): 307–311.

27. Trent M, Lehmann HP, Qian Q et al. Adolescent and parental utilities for the health states associated with pelvic inflammatory disease. *Sex Transm Infect* 2011; 87:583e587. doi: 10.1136/sextrans-2011-050187.

28. Ungar W. Challenges in health state valuation in the paediatric economic evaluation. *Pharmacoeconomics* 2011;29:641-652.

29. Heazall AEP, Siassakos D, Blencowe H, et al. Stillbirths: economic and psychosocial consequences. *The Lancet* 2016; 87 (10018) 604-616.

30. Wright DR, Wittenberg E, Shannon Swan J, Miksad RA, Prosser LA. Methods for Measuring Temporary Health States for Cost-Utility Analyses. *Pharmacoeconomics* 2009; 27 (9) 713-723.

31. Roberts TE, Robinson S, Barton P, Bryan S, Low N. Screening for *Chlamydia trachomatis*: a systematic review of the economic evaluations and modelling. *Sexually Transmitted Infections*, 2006; *82*(3):93–200.

32. Turner K, Adams E, Grant A, et al. Costs and cost-effectiveness of different strategies for chlamydia screening and partner notification: an economic and mathematical modelling study. *BMJ* 2011; 342:c7250.

33. Shepherd J, Kavanagh J, Picot J et al.The effectiveness, and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13-19: a systematic review and economic evaluation. *Health Technol Assess* 2010;14(7).

34. Torrance GW. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* 1986; 5: 1-30.

35. Krabbe PFM, Essink-Bot M, Bonsel GJ. The comparability and reliability of five health-state valuation methods. *Soc. Sci. Med* 1997; 45 (11)1641-1652.

36. Hall J, Gerard K, Salkeld G, Richardson J. A cost-utility analysis of mammography screening in Australia*. Social Science Medicine* 1992; 34: 993-1004.

37. Bass EB, Steinberg EP, Pitt HA et al. Comparison of the rating sale and the standard gamble in measuring patient preferences for outcomes of gallstone disease*. Medical Decision Making* 1994; 14, 307-341.

38. Rutten-van Molken M, Bakker C, van Doorslaer E, van der Linden S. Methodological issues of patient utility measurement: Experience from two clinical trials. *Medical Care* 1995; 33(9), 922-937.

39. Swan JS, Fryback DG, Lawrence WF et al. A time-tradeoff method for cost-effectiveness models applied to radiology. *Med Decis Making* 2000; 20 (1): 79 – 88.

40. Tosteson AN, Kneeland TS, Moncur MM, et al. Has the impact of hormone replacement therapy on health-related quality of life been undervalued? *Gend Based* 2000; 9: 119-130.

41. Merlino LA, Bagchi I, Taylor TN, et al. Preferences for fractures and other glucocorticoid-associated adverse effects among rheumatoid arthritis patients. *Med Decis Making* 2001; 21:122–32.

42. Tosteson AN, Kneeland TS, Neasa RF et al. Automated current health time trade-off assessments in women's health. *Value Health*2002;5(2): 122-32

43. Centre for Reviews and Dissemination (CRD), University of York*. Systematic Reviews: CRD's guidance for undertaking reviews in health care.* CRD, University of York 2009. Available from: http://www.york.ac.uk/inst/crd/pdf/Systematic Reviews.pdf [Accessed June 20th, 2015].

44. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred Reporting Items for Systematic Reviews and Meta-analyses: the PRISMA Statement. *PLOS Med* 2009; 6(7): e1000097.

45. Refworks Database 2015. Available from: http://www.refworks.com/refworks2/default.aspx?r=references|MainLayout::init [Accessed March 1st 2015].

46. Roberts T, Henderson J, Mugfod et al. Antenatal ultrasound screening for fetal abnormalities: a systematic review of studies of cost and cost-effectiveness. *BJOG* 2002;109: 44-56.

47. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technology Assessment* 1999; 3(9) Available from: http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0011/64829/FullReport-hta3090.pdf [Accessed 18th August 2015].

48. Brazier J, Deverill M. A checklist for judging preference-based measures of health-related quality of life: learning from psychometrics. *Health Economics* 1999; 8: 41–51.

49. Swan JS, Lawrence WF, Roy J. Process Utility in Breast Biopsy. *Med Decis Making* 2006; 26:347–359.

50. Fennessy FM, Yin Kong C, Tempany CM, Swan SJ. Quality-of-life assessment of Fibroid treatment options and outcomes. *Radiology* 2011; 259(3): 785-792.

51. Hess LM, Malone DC, Reed PG, Skrepnek G, Weihs K. Preferences of patients and oncologists for advanced ovarian cancer treatment-related health states. *Health outcomes research in Medicine*. 2010; 1: e51-e59.

52. Chan JL, Kabete MU, Oldread E et al. The use of preferences to measure the benefit of adjuvant radiation therapy for stage 1 seminoma. *Int. J. Radiation Oncology Biol. Phys* 2002; 53(4):934–941.

53. Sun CC, Bodurka DC, Donato ML et al. Patient preferences regarding the side effects of chemotherapy for ovarian cancer: Do they change over time? *Gynaecologic Oncology* 2002; 87:118-128.

54. Matza LS, Secnik K, Rentz AM et al. Assessment of health state utilities for Attention-deficit/ hyperactivity disorder in children using parent proxy report. *Quality of life Research* 2005; 14(3):735-747.

55. Revicki DA, Hanlon J, Martin S et al. Patient-based utilities for bipolar disorder-related health states. *Journal of Affective disorders* 2005; 87:203-210.

56. Shumway M, Chouljian TL, Battle CL. Measuring Preferences for Schizophrenia Outcomes with the Time Trade-off Method. *Journal of Behavioural Health Services & Research*, 2005; 32(1): 14-26.

57. Bonomi AE, Boudreau DM, Fishman PA et al. Quality of life valuations of mammography screening. *Quality of Life Research* 2008; 17(5): 801-814.

58. Gu N, Wolf C, Leopold S, Manner PA, Doctor JN. A Comparison of Physician and Patient Time Trade-Offs for Postoperative Hip Outcomes. *Value in Health* 2009;12(4).

59. Havrilesky LJ, Broadwater G, Davis DM et al. Determination of the quality of life-related utilities for health states relevant to ovarian cancer diagnosis and treatment. *Gynecologic Oncology* 2009;113:216-220.

60. Jewell EL, Smrtka M, Broadwater G et al. Preference-Based Utility Scores for Adverse Events Associated With the Treatment of Gynecologic Cancers. *International Journal of Gynecological Cancer.* 2013;23(6):1157-1165.

61. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Economics* 1995; 4: 289-299.

62. Dolan P. Modelling valuations for health states: the effect of duration. *Health Policy* 1996; 38:189-203.

63. Johnston K**,** Brown J, Gerard K et al*.* Valuing temporary and chronic health states associated with breast screening. *Soc Sci Med* 1998; 47:213–22.

64. Robinson A, Thomson R, Parkin D, Sudlow M, Eccles M. How patients with atrial fibrillation value different health outcomes: a standard gamble study. *J Health Serve Res Policy* 2001; 6(2):92-98.

65. McNamee P, Glendinning S, Shenfine J et al. Chained Time Trade-Off and Standard Gamble Methods: Applications in Oesophageal Cancer. *The European Journal of Health Economics* 2004; 5(1): 81-86.

66. Brazier JE, Roberts J, Platts M, Zoellner YF. Estimating a preference-based index for a menopause specific health quality of life questionnaire. *Health and Quality of Life Outcomes* 2005; 3:13 doi: 10.1186/477-7525-3-13.

67. Matza LS, Boye KS, Yurgin N et al. Utilities and disutilities for type 2 diabetes treatment-related attributes. *Quality of Life Research* 2007; 16(7): 1251-1265.

68. Swinburn P, Lloyd A, Nathan P et al. Elicitation of health state utilities in metastatic renal cell carcinoma. *Current Medical Research and Opinion* 2010; 26(5).

69. Stiggelbout AM, Kiebert GM, Kievit J et al. Utility assessment in cancer patients: Adjustment of Time tradeoff scores for the utility of life years and comparison with Standard Gamble scores. *Med Decis Making* 1994; 14:82-90.

70. Jansen ST, Stiggelbout A, 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; 18:391-399.

71. Unic I, Stalmeier PFM, Verhoff LCG, Van Daal WAJ. Assessment of the Time-tradeoff values for Prophylactic Mastectomy of Women with a Suspected Genetic Predisposition to Breast Cancer. *Med Decis Making* 1998; 8:268-277.

72. Jansen ST, Stiggelbout A, Wakker PP et al. Unstable Preferences: A Shift in Valuation or an Effect of the Elicitation Procedure? *Med Decis Making* 2000; 20:62-71.

73. Locadia M, Stalmeier PFM, Oort FJ et al. A Comparison of 3 Valuation Methods for Temporary Health States in Patients Treated with Oral Anticoagulants. *Med Deci Making* 2004a; 24:625-633.

74. Locadia M, Bossuyt PMM, Stalmeier PFM et al. Treatment of venous thromboembolism with vitamin K antagonists: patients' health state valuations and treatment preferences. *Thromb Haemost* 2004b; 92(6): 1336-41.

75. Dion M, Tousignant P, Bourbeau J, Menzies D, Schwartzman K. Measurement of Health Preferences among Patients with Tuberculous Infection and Disease*. Med Decis Making* 2002; 22(Suppl): S102–S114.

76. Grunberg SM, Weeks J, Fischer Magnan W et al. Determination of Utility Scores for Control of Chemotherapy-Induced Nausea or Vomiting—CALGB 309801. *J Support Oncol* 2009; 5: W17–W22

77. Hogg K, Shaw J, Coyle D et al. Validity of standard gamble estimated quality of life in acute venous thrombosis. *Thrombosis Research* 2014; 134:819-825

78. Revicki DA, Wood M. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications *Journal of Affective Disorders* 1998; 48: 25–36.

79. Szende A, Schaefer C, Goss TF et al. Valuation of transfusion-free living in MDS: results of health utility interviews with patients *Health and Quality of Life Outcomes* 2009;7:81 doi:10.1186/1477-7525-7-81.

80. Badia X, Herdman M, Ohinmaa A. Feasibility and validity of the VAS and TTO for eliciting general population values for temporary health states: A comparative study. Health services & Outcomes Research Methodology 2001; 2:51-65.

81. Gudex C. Time trade-off user manual: Props and self completion methods. Centre for health Economics. University of York. York 1994.

82. de Haes JCM, Stiggelbout AM. Assessment values, utilities and preferences in cancer patients. *Cancer Treat Rev* 1996; 22:13-26