Sanghera, S., & Coast, J. (2019). Measuring quality-adjusted life-years when health fluctuates. *Value in Health*. https://doi.org/10.1016/j.jval.2019.09.2753

Title: Measuring QALYs when health fluctuates

Running title: Measuring QALYs when health fluctuates

Authors:

Sabina Sanghera (PhD, MSc)[1], Joanna Coast (PhD, MSc)[1]

[1]Health Economics Bristol, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 1NU, UK

Corresponding Author:

Dr Sabina Sanghera
Health Economics Bristol (HEB),
Population Health Sciences, Bristol Medical School
University of Bristol
1-5 Whiteladies Road
Bristol
BS8 1NU
Telephone: 0117 428 3124
Email: sabina.sanghera@bristol.ac.uk

Funding:

This manuscript is independent research supported by the National Institute for Health Research (**NIHR Post Doctoral Fellowship, Dr Sabina Sanghera, PDF-2017-10-008).** The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.

Key words:

Recall, timing of assessment, quality of life, QALYs, fluctuating health, economic evaluation

Acknowledgements:

The authors would like to thank participants at the Health Economists' Study Group meeting in Bristol, June 2018, for their comments on an earlier version of the manuscript, and particularly Paula Lorgelly for discussing the paper.

Highlights:

- The challenges in measuring quality of life when health fluctuates for use in economic evaluation have received little attention, with only a few studies focusing on particular issues within specific studies/conditions.
- This paper highlights the potential issues in measuring quality of life and calculating QALYs more broadly when health fluctuates and the influence on economic evaluation
- There is potential in the current practice of economic evaluations of conditions with fluctuating health states to distort treatment decisions away from the optimal allocation

Abstract

Recurrent fluctuations in health states can occur due to long-term conditions with episodic symptoms or through side effects of cycles of treatment. Fluctuations and associated duration of symptoms can be predictable (e.g. side effects of chemotherapy treatment) or unpredictable (e.g. relapse in multiple sclerosis). Such recurrent fluctuations in health states can have an important impact on a person's health-related quality of life. When symptoms vary by time of day, day of the week, or during the month, it is challenging to obtain reliable health-related quality of life estimates for use in assessing cost-effectiveness of interventions.

The adequacy of the quality of life estimate will be impacted by: (1) the standard recall period associated with the chosen measure (e.g. 'health today' for EQ-5D, 'past month' for SF-6D) and the way that respondents understand and make judgements about these recall periods, (2) the chosen timepoints for assessing health-related quality of life in relation to the fluctuations in health, and (3) the assumptions used to interpolate between measurement time points and thus calculate the QALYs.

These issues have not received sufficient methodological attention and instead remain poorly accounted for in economic analyses. There is potential for these issues to considerably distort treatment decisions away from the optimal allocation. This paper brings together evidence from health economics, psychology and behavioural economics to explore these challenges in depth; presents the solutions that have been applied to date; and details a methodological research agenda for measuring QALYs in recurrent fluctuating health states.

1.0 Introduction

Many international institutes recommend that quality-adjusted life years (QALYs) are used as a composite measure of length and health-related quality of life in economic evaluations.[1-6] To capture health-related quality of life, generic instruments such as SF-6D (derived from SF-12 or SF-36),[7,8] EQ-5D[9] and the HUI[10,11] are recommended by the Second Panel on Cost-effectiveness in Health and Medicine.[12] EQ-5D and SF-6D are commonly formally recommended across several countries, such as UK, Canada, Australia, New Zealand and Ireland[(1,2,5,6,13)]; the National Institute for Health and Care Excellence (NICE) in the UK recommends both SF-6D and EQ-5D, but prefers EQ-5D.[1] Health-related quality of life is then multiplied by the length of time in the given health state to estimate the QALY summary measure of health outcome.

Despite being recommended for use, these health-related quality of life measures may be problematic in economic evaluations for conditions or treatments that cause fluctuations in health-related quality of life. Fluctuations can occur in any person, but they are particularly prevalent in: (1) long-term conditions that have episodic symptoms; (2) responses to a trigger (e.g. stress increasing the likelihood of a seizure in epilepsy) or allergen, or (3) side effects of treatment. These fluctuations can be predictable (e.g. side effects of chemotherapy treatment) or unpredictable (e.g. relapse in multiple sclerosis). In some instances, minor or short-term changes in mental or physical health may be less likely to influence healthcare decision-making, but recurrent fluctuations that are regularly experienced can cause considerable variation in health and could affect decisions about treatment and care. The remainder of this paper is concerned with recurrent fluctuations.

Due to their nature (Table 1), conditions or treatments with recurrent fluctuating states are difficult to measure. Standard measurement and analytic approaches applied to 'typical', relatively steady conditions, are not suitable when fluctuations are recurring. This is due to: (1) the recall periods, (2) the timing of assessment, (3) the analytical assumptions used to estimate the QALY, and (4) the valuation methods employed to generate the health-related quality of life values. The focus of this paper is on the first three of these, which are broadly concerned with measurement rather than valuation. Valuation methods will be discussed in a separate paper.

These issues have received insufficient methodological attention and are poorly accounted for in economic evaluation. Both decision model-based economic evaluations and regression based within-

trial economic evaluations are subject to these concerns. If inappropriate methods are used in either of these applications of research, health-related quality of life estimates could be poor and economic recommendations to decision-makers misleading. In addition to current approaches to healthcare, these problems are of even greater concern in personalised medicine, where a 'one size fits all' approach to treatment and care of patients is no longer considered to be appropriate and the adequate estimation of health-related quality of life is essential for treatment to be appropriately tailored towards individual's needs.

This paper brings together evidence from health economics, psychology and behavioural economics to explore these challenges in depth, it outlines the relevance to both trial-based and model-based economic evaluations, describes the solutions that have been applied to date, and presents a methodological research agenda for measuring and calculating QALYs in recurrent fluctuating health states.

2.0 Measuring QALYs in recurrent fluctuating health states

Typically, within trial economic evaluations are based exclusively on resource use and health-related quality of life data that are prospectively collected during a randomised controlled trial. Whilst model-based economic evaluations can also be carried out alongside trials, data from other sources, such as observational studies, can be incorporated. For example, policy decisions made by NICE in the UK and other international institutes are generally based on the cost-effectiveness results of a decision model. The recommendations from the UK NICE guidelines manual, which provide a similar message to other international institutes, state that "A cost-effectiveness analysis could be modelled around a single well conducted randomised controlled trial, or by using decision-analytic techniques with probability, cost and health outcome data from a variety of published sources" ([1]The Guidelines Manual, 2012, pp 106). Any economic evaluation relies on good quality data collection that captures important changes in health. In both within trial economic evaluations and decision model-based economic evaluation, the timing of assessment of any outcome and the recall period used can introduce bias in the measurement of recurrent fluctuating health states by ignoring the temporal patterns of health-related quality of life in these states.

2.1 Timing of assessment

Data in observational studies can be collected through longitudinal cohort studies, take a snapshot of the population at a certain point in time through cross-sectional surveys, or can be analysed retrospectively through case control studies. In clinical trials, data collection is either event driven (e.g. questionnaires completed during routine GP visits or following an event) or time driven (e.g. questionnaires completed periodically at set time periods).[14] In any of these cases, for both predictable and unpredictable conditions, the resultant assessments of health-related quality of life may not comprehensively reflect patients' experiences and could lead to an overestimation or underestimation of outcomes.[15] This is because the patient may or may not be experiencing episodes of ill-health at the point of assessment or at each measurement time point; indeed, at the point of assessment the patient may be at the worst point of the fluctuation, best point of fluctuation, or at some point in-between (Figure 1). When fluctuations are unpredictable and random and if the sample size is large enough, on average, the quality of life assessment could be argued to be unbiased. But, even in this case there will be great variation in scores, which results in imprecise estimates. Finally, given symptoms are fluctuating, it is unclear what patients are taking into consideration when completing questionnaires, so it should not be assumed that a large sample size will overcome the matters described.

Concerns with the current implementation of approaches to measure health-related quality of life have been raised in respect of several recurrent fluctuating health states.[16] Particular attention has been paid to this issue in chemotherapy trials. Recent trials have collected EQ-5D at: baseline, 3 months, 6 months, 12 months[17] or 6 weeks, 16 weeks and 12 months[18]; or every 3 weeks[19] when chemotherapy is administered. Several studies in cancer chemotherapy have shown that the timing of assessment of cancer-specific measures can significantly influence health-related quality of life results.[15,20-24] The most severe effects of chemotherapy tend to occur during the first week of treatment, with partial or complete recovery by the day when chemotherapy is next administered.[15] Measurement of health-related quality of life is commonly taken on the day of treatment for logistical reasons and hence underestimates symptom burden, as treatment side effects will be missed.[15] The literature in non-recurrent acute events or illnesses (for example, hepatitis A and total knee arthroplasty) shows that the timing of assessment can significantly influence cost-effectiveness results when using EQ-5D or SF-6D.[25,26] even when fluctuations are not recurrent and short-term.

2.2 Recall periods

When health is fluctuating, the health-related quality of life information gathered in questionnaires, such as EQ-5D and SF-12, can be influenced by the recall period used and can unduly influence the data that are used in either economic evaluation type. Problems with recall periods have been discussed in relation to healthcare usage,[27,28] but much less attention has been paid to the recall periods of quality of life measures used in health economics. The recall periods of commonly used measures (see Table 2) vary from the immediate (such as 'health today' or 'current status') to longer time periods (such as 'health over the past four weeks'). Some measures do allow a choice of recall periods (for example, HUI, SF-36), but overtime one recall period has been commonly used in studies and appear to have become the standard timeframe for each measure. The alternative timeframes seem to be less commonly used.[29]

Assuming respondents adhere to the recall period, measures with shorter recall periods (such as EQ-5D's 'health today'), could result in changes in health-related quality of life being missed if the patient is not experiencing symptoms in that short time period, or indeed, if the measure is completed earlier in the day rather than later. In questionnaires with longer recall periods (such as SF-36/SF-12's 'past 4 weeks'), patients may focus on the worst health state, construct an average, or focus on recovery.[30] If the respondent experiences a change in health state during the recall period it is not clear what is reported.[30]

Longer recall periods have also been found to lead to a reduction of accuracy in recall due to memory problems as well as perceptions of events.[31] With respect to pain, findings suggest that patients' memories, and their response to questionnaires with retrospective recall periods, is determined by the worst and end part of an episode or day — known as the peak-end rule. [32,33] However, anticipatory emotions associated with knowing that a change in state is due to occur and the speed at which the change in state occurs are also thought to influence responses to retrospective questionnaires — both of which are relevant factors for recurrent fluctuating health states that are not currently taken into consideration.[34] Other factors that are not considered when health is fluctuating, but can influence recalled responses, are mood or person's state at the time of completion, the amount of time that has elapsed since the event, and the number of changes in the state that have occurred.[35,36] Each of these

concepts could be affecting responses, but there is not currently a clear enough understanding of how patients complete questionnaires to know what the responses reflect.

3.0 Calculating QALYs and influence on economic evaluations

Together with the measurement concerns outlined, the analytical approach used to estimate the QALY, the assumptions used and the approaches to handle missing data when health fluctuates can impact cost-effectiveness results regardless of the economic evaluation type employed. There are also some additional aspects of model-based economic evaluations that are impacted when health fluctuates that are discussed.

3.1 Assumptions of the QALY

The assumptions used to interpolate between measurement time points and thus to calculate the QALY can be problematic in estimating values for recurrent fluctuating health states. Commonly, a linear relationship between measurement time points is assumed in calculating the area under the curve to obtain a QALY value.[37] A linear relationship implies that changes in health-related quality of life show a constant rate of change between measurement time points (Figure 2). So, if the timing of assessment is inappropriate, the estimate of the QALY value is likely to be inaccurate when using linear interpolation. The changes in health-related quality of life may be over- or under-estimated in the QALY calculation. Although it has not been explored in recurrent fluctuating states, in the previous example of a one-off acute illness (hepatitis A) a difference in results was observed when three alternative analytical choices were explored to estimate the QALY for two different measure, SF-6D and EQ-5D.[25] The analytic choices included patients having: (1) constant health-related quality of life for the duration of the illness, when only one health-related quality of life estimate is available, which is represented by the area of a rectangle; (2) linear improvement in health-related quality of life; or (3) marginal improvement every day, assuming an exponential change. The impact of the assumption on the results may be more pronounced in recurrent fluctuating health states because the ‘event’ is recurring, compounding

discrepancies between estimated and experienced health-related quality of life, and emphasising the need to carefully consider the assumptions made to calculate the QALY and to justify the approach.

3.2 Missing data

Missing data occurs in most circumstances and health-related quality of life data are often imputed in economic evaluations. When health is fluctuating and health-related quality of life data are missing, assumptions of Missing At Random (MAR) or Missing Completely At Random (MCAR) are less likely to hold than for missing data in 'steady' conditions. Missing Not At Random (MNAR) may be more likely to occur as data may be missing at certain timepoints because the patient was experiencing bad days. Simple imputation methods, which can produce bias in any circumstance, such as last observation carried forward or next observation carried backward can be even more prone to bias if the number of missing observations is high or sequential in fluctuating states. More comprehensive methods to handle missing data, such as multiple imputation, are commonly used but can also introduce bias when health fluctuates. In multiple imputation, values are randomly and repeatedly drawn from an assigned distribution based on the existing observed data and the average used to impute the missing value.[38] Here, if patients are more commonly asked to respond at 'good' points during the cycle than at 'bad' points (as in chemotherapy trials) the imputation will largely be based on 'good' scores biasing the results. If fluctuations are predictable, information about the time point of the cycle should be incorporated into the regression model used to predict values for missing data to avoid biasing results. However, suitable approaches to multiple imputation for unpredictable fluctuations that can occur at any time have not yet been recommended.

3.3 Decision-modelling and trial based economic evaluations

In addition to the problems raised so far for model-based and within-trial economic evaluations, there are some additional aspects of model-based economic evaluations that need to be considered when health fluctuates.

Such concerns specific to decision-model based economic evaluations, include whether suitable health-related quality of life data are used to reflect fluctuations, how fluctuations are accounted for in the

model structure, and whether appropriate assumptions are used to assign health-related quality of life data to model health states. In addition, the health-related quality of life impact of the anticipatory feelings of the next exacerbation should also be appropriately considered by the analyst when assigning health-related quality of life values to mutually exclusive states (e.g. remission and relapse). Efforts are made by institutes and researchers, such as in the NICE guidelines manual and the guide to methods of technology appraisal[1] and the Phillips checklist for decision models,[39] to encourage researchers to justify structural assumptions and model inputs and to state methods used to appraise data sources and explore uncertainty, but these guides are understandably broad so they can be applied to many circumstances.

Box 1 details the areas for concern in both within-trial and decision-model based economic evaluations when health fluctuates.

Box 1. Checklist of issues to consider when conducting a within-trial or model-based economic evaluation when health fluctuates

1. **Does the condition or treatment cause fluctuating health?**
    i. Are the changes predictable or unpredictable?

2. **Are the outcomes measured appropriately (observational study, trials, meta-analysis)?**
    a. Timing of assessment/ when was quality of life collected
        i. Could the timing of assessment influence the results given the condition or treatment pathway?
        ii. Are the outcomes always assessed at the same time point?
        iii. Are fluctuations, exacerbations, good and bad days captured?
        iv. Could the timing of assessment influence the completion of the questionnaire?
    b. ***Recall measure/ how was quality of life collected***
        i. Could changes in health occur within the recall period of the questionnaire?
        ii. Could the recall period cause fluctuations to be missed?
        iii. Could there be any reason, related to the fluctuations, that mean respondents are not completing the questionnaire as intended?

3. **Could the methods of analysis influence the results?**
    a. ***Linear interpolation of the QALY***
        i. Could important changes occur in health in between measurement time points?
        ii. Will linear interpolation ignore changes between measurement time points?
    b. Missing data
        i. Is MCAR or MAR a sensible assumption — could missingness be related to the respondent experiencing symptoms?
        ii. Are appropriate methods used to impute data that account for fluctuations? Is information on time point of missingness used to inform the imputation approach?

4. **Does the decision-model account for fluctuations?**
    a. Model assumptions
        i. How are fluctuations accounted for in the model?
        ii. How are quality of life data applied to the model? Are appropriate assumptions used to fit data to health states?
    b. Model structure
        i. Does the model structure allow for changing health/exacerbations/ good and bad days?
        ii. Should the states be considered to be mutually exclusive? Are anticipatory emotions important and how have they been accounted for in the model?

c. Additional data source considerations

i. Are the quality of life data suitable for the health state?

ii. Are the quality of life data used across the states consistent? Due to the scores used, could the impact of fluctuations be inappropriately captured in multiple states?

d. Cycle length

i. Are changes expected within the time period of the cycle length?

ii. Is the time spent in each health state appropriate for fluctuations?

4.0 Solutions to date

Some approaches have been used to overcome the challenges of measuring QALYs for particular conditions where health fluctuates and obtain more accurate values, but these have focused on the particular issues within the specific trials/studies, and have not attempted to develop more general solutions. Approaches used here include those based on recall (by changing the timeframe) and those based on timing of assessment (administering the questionnaire more frequently).

4.1 Recall

The availability of alternative recall periods for some of the standard measures, such as SF-36 and HUI, suggests that the need for different recall periods in certain settings has been identified. However, a search of the literature shows that the alternative recall periods of past week (SF-12 and HUI), or past 2 weeks and 4 weeks for the HUI have not been widely used when health fluctuates.

One approach that has, however, been used is to ask for responses by referring to different recall periods. In migraine, the temporary nature of a migraine and the changes that occur within an attack are difficult to capture, such that one study asked respondents to complete EQ-5D both with reference to their most recent attack and for their health outside of the attack in an attempt to obtain more accurate values.[40] The authors suggested that these values can then be used in model-based economic

evaluations of migraine.[40] Similarly, in Parkinson's disease it is argued that the recall period of outcome measures (EQ-5D and a Parkinson's specific measure) are unlikely to capture the variation in symptoms and functioning that occur within each day between treatment doses, as treatment effects begin to wear off.[41] Consequently, patients were asked to complete EQ-5D-5L twice: according to their treatment 'on-time', when treatment is effective and treatment 'off-time', when it is less effective. Using the UK value set,[42] the index scores between the two states differed by 0.23, which led the authors to conclude that the current use of EQ-5D cannot capture variation in symptoms and that both on and off-time must be captured in estimating health-related quality of life for Parkinson's disease.

In other instances, researchers have used problematic recall periods as one reason for developing a preference-based measure from an existing condition-specific measure, as in the development of NEW-QoL-6D for epilepsy.[43] Here, one justification for the new measure was that the recall period of 'health today', used in EQ-5D, was seen as problematic. Though the appropriateness of this approach and the value of obtaining condition-specific QALYs is widely debated, condition-specific preference-based measures have commonly been developed in conditions that fluctuate. The reasoning for developing these measures is unlikely to be exclusively due to concerns with recall periods of standard measures, but it is interesting to note that condition-specific measures use a variety of recall periods For example, HAQ-DI for rheumatoid arthritis (past week), MSIS for multiple sclerosis (past 2 weeks), DLQI for psoriasis (past week), CFQ-R for cystic fibrosis (2 weeks), a combination of MSQ and HIT-6 for migraine (48 hours and past 4 weeks), and CCQ for COPD (past week or 24 hours) all have different recall periods to those used in generic measures which is likely to have been changed to increase the likelihood of capturing fluctuations in the condition.

Similarly, mapping algorithms have been developed from condition-specific measures to generic preference-based measure, such as EQ-5D, in many conditions that fluctuate.[44] This approach can be used to obtain QALYs when a preference-based measure has not been administered. The validity of mapping from condition-specific measures to generic preference-based measures, however, has not been assessed specifically with respect to the timeframe when the recall periods of the mapped questionnaires are different.

4.2 Timing of assessment

Rather than altering the recall period, other studies in conditions with fluctuating symptoms have altered the time at which the assessment is made. One early example of altering timing of assessment exists using a measure that is not used to generate QALYs. In relapsing-remitting multiple sclerosis, a neurological condition that affects the nervous system, people with the condition experience prolonged periods of remission and unpredictable relapses, or exacerbations of their condition. To ensure health-related quality of life was assessed comprehensively amongst patients in remission, one study demonstrated the feasibility of daily diary completion of the Euroqol Visual Analogue Scale (EQ-VAS) and symptom information for 6 weeks.[45]

4.3 Calculation of the QALY

Optimal timings of data collection and interpolation assumptions have not been proposed for fluctuating health states. But, recommendations have been made in one study on total knee arthroplasty, where a surgical intervention caused non-recurrent short-term fluctuations in health-related quality of life. Health-related quality of life increased rapidly in the first 3 months until it stabilised at 6 months. The influence of both various interpolations of the QALY and the longer lengths of time between timing of assessment were explored. For interpolation, an assumption of linear interpolation between timepoints was compared with interpolation through an assumed immediate (vertical) improvement in health-related quality of life at the previous measurement time point followed by a horizontal line to the next timepoint. The addition of data from more measurement timepoints was also explored and the extent of the errors in QALYs gained was compared across both interpolation methods. The authors recommend that additional specific time points of data collection for SF-12, along with linear interpolation, minimised the error in the QALY estimation by most closely reflecting a 'true' estimate which had more frequent measurement time points.[26] Similar research is required in recurrent fluctuating conditions.

5.0 A research agenda for measuring QALY benefits in recurrent fluctuating health states

This final section of the paper focuses on a research agenda in relation to these measurement issues. As indicated at the start of the paper, further work will consider valuation issues. Inevitably, as this paper focuses on measurement, it is linked to the situation where the area under the curve approach is used for QALY computation. Alternative methods of valuation may be needed for situations where these under the curve approaches are rejected[46]. Nevertheless, the focus of the research agenda here remains the measurement issues.

5.1. Timing of assessment

The most accurate estimates of QALYs would be expected when interval measurements are taken at each point where there is a change in health-related quality of life and are used alongside information on the duration of each state to generate relevant health-related quality of life curves. In some conditions, these points may be predictable, but in others they cannot be anticipated in advance.

Research that is needed to explore potential solutions includes: (1) asking patients if they are currently experiencing symptoms when completing the questionnaire, particularly if fluctuations are unpredictable — if the sample size is sufficient, this approach will provide some sense of the treatment effect on health-related quality of life with the possibility of exploring different ranges in a sensitivity analysis to see the impact on results; or (2) validating quality of life scores by identifying other studies that have specifically elicited quality of life during a fluctuation, if available, these values could be tested in a sensitivity analysis.

Other areas for research related to both timing of assessment and approaches to questionnaire completion could lead to the development of ways of aligning patient and proxy responses, if patients are able to complete measures at some points in their illness and not at others. Additionally, as patients are fluctuating between health states instead of remaining in one health state, the extent of adaptation over periods of assessment could also be explored to determine if patients' responses overtime are more or less subject to response shift when health is constantly changing.

5.2. Recall

With respect to recall periods, research should include understanding how patients with recurrent fluctuating symptoms complete questionnaires with standard recall periods: what patients are taking into consideration depending on the timing of assessment and how much the recall period influences responses. The process and appropriateness of tailoring measure recall periods to each condition could also be tested, followed by research to identify the extent to which changes in recall affect results as well as ease of completion for patients. For example, it may be easier for patients to complete a questionnaire if they are asked to specifically refer back to a particular episode rather than using a standard recall.

5.3. Calculation of the QALY

Further, given the influence of timing of assessment and analytical assumptions to calculate the QALYs in acute events, there is also a need to identify and assess new approaches against a 'best estimate' for measurement and make recommendations on the optimal approaches for timing of assessment, recall and interpolation assumptions for the QALY for recurrent fluctuating health states.

Research should also be carried out to determine the extent to which the approaches used to impute data for predictable and unpredictable recurrent fluctuations impacts on the results. The incorporation of secondary data on the nature and pattern of health-related quality of life fluctuations into imputation models could be explored to better predict health-related quality of life scores.

Finally, with respect to model-based economic evaluations, the checklist of issues presented in Box 1 should be consulted to ensure that suitable data are used in the model and section 4 of the checklist should be used to ensure the model structure and assumptions account for fluctuating health. Such research should lead to more robust recommendations on the measurement of outcomes in recurrent fluctuating health states by identifying the extent of the problem and developing a systematic approach to account for the impact of recurrent fluctuating health states in economic evaluation. As discussed at the outset of this paper, this research can also contribute to methods of capturing outcomes in economic evaluations of personalised medicine and could lead to treatment decisions that have been made based on robust outcome data. To fully explore the value of the proposed solutions the research should be carried out in different contexts to determine the extent to which this theoretical problem is taking place in practice.

References

1. NICE. Guide to the methods of technology appraisal. London: National Institute for Health and Care Research, 2013.
2. CADTH. Guideslines for the Economic Evaluation of Health Technologies: Canada. 4th ed ed. Ottawa: CADTH, 2017.
3. PBAC. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee. Australia: Australian Government Department of Health, 2016.
4. DACEHTA. Health Technology Assessment Handbook. Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health, 2007.
5. MSAC. Technical Guidelines for preparing assessment reports for the Medical Services Advisory Committee – Medical Service Type: Therapeutic. Australia: Australian Government: Department of Health, 2016.
6. HIQA. Guidelines for the economic evaluation of health technologies in Ireland. Dublin: HIQA, 2018.
7. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002; 21: 271-92.
8. Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol. 1998; 51: 1115-28.
9. Brooks R. EuroQol: the current state of play. Health Policy. 1996; 37: 53-72.
10. 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: 113-28.
11. Torrance GW, Feeny DH, Furlong WJ, et al. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Med Care. 1996; 34: 702-22.
12. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016; 316: 1093-103.
13. PHARMAC. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis. New Zealand: PHARMAC, 2017.
14. Fairclough DL. Patient reported outcomes as endpoints in medical research. Stat Methods Med Res. 2004; 13: 115-38.
15. Giesinger JM, Wintner LM, Zabernigg A, et al. Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. BMC Cancer. 2014; 14: 758.
16. Sculpher M. A cost-utility analysis of abdominal hysterectomy versus transcervical endometrial resection for the surgical treatment of menorrhagia. Int J Technol Assess. 1998; 14: 302-19.
17. Shiroiwa T, Fukuda T, Shimozuma K, et al. Long-term health status as measured by EQ-5D among patients with metastatic breast cancer: comparison of first-line oral S-1 and taxane therapies in the randomized phase III SELECT BC trial. Qual Life Res. 2016.
18. Arnold DT, Hooper CE, Morley A, et al. The effect of chemotherapy on health-related quality of life in mesothelioma: results from the SWAMP trial. Brit J Cancer. 2015; 112: 1183-89.
19. Andronis L, Goranitis I, Pirrie S, et al. Cost-effectiveness of zoledronic acid and strontium-89 as bone protecting treatments in addition to chemotherapy in patients with metastatic castrate-refractory prostate cancer: results from the TRAPEZE trial (ISRCTN 12808747). BJU Int. 2016.
20. Hakamies-Blomqvist L, Luoma ML, Sjostrom J, et al. Timing of quality of life (QoL) assessments as a source of error in oncological trials. J Adv Nurs. 2001; 35: 709-16.

21. Klee MC, King MT, Machin D, et al. A clinical model for quality of life assessment in cancer patients receiving chemotherapy. Ann Oncol. 2000; 11: 23-30.
22. Pater J, Osoba D, Zee B, et al. Effects of altering the time of administration and the time frame of quality of life assessments in clinical trials: an example using the EORTC QLQ-C30 in a large anti-emetic trial. Qual Life Res. 1998; 7: 273-8.
23. Hurny C, Bernhard J, Coates A, et al. Timing of baseline quality of life assessment in an international adjuvant breast cancer trial: its effect on patient self-estimation. The International Breast Cancer Study Group. Ann Oncol. 1994; 5: 65-74.
24. Ediebah DE, Coens C, Maringwa JT, et al. Effect of completion-time windows in the analysis of health-related quality of life outcomes in cancer patients. Ann Oncol. 2013; 24: 231-7.
25. Luyten J, Marais C, Hens N, et al. Imputing QALYs from single time point health state descriptions on the EQ-5D and the SF-6D: a comparison of methods for hepatitis a patients. Value Health. 2011; 14: 282-90.
26. Schilling C, Dowsey M, Clarke P, et al. Using Patient Reported Outcomes for Economic Evaluation: Getting the Timing Right. Value Health. 2016; 19: 949-950
27. Clarke PM, Fiebig DG, Gerdtham UG. Optimal recall length in survey design. J Health Econ. 2008; 27: 1275-84.
28. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. J Health Econ. 2014; 35: 34-46.
29. Keller SD, Bayliss MS, Ware JE, Jr., et al. Comparison of responses to SF-36 Health Survey questions with one-week and four-week recall periods. Health Serv Res. 1997; 32: 367-84.
30. Bansback N, Sun H, Guh DP, et al. Impact of the recall period on measuring health utilities for acute events. Health Econ. 2008; 17: 1413-9.
31. Stull DE, Leidy NK, Parasuraman B, et al. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. Curr Med Res Opin. 2009; 25: 929-42.
32. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain. 1996; 66: 3-8.
33. Stone AA, Broderick JE, Kaell AT, et al. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics? J Pain. 2000; 1: 212-17.
34. Fredrickson BL, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. J Pers Soc Psychol. 1993; 65: 45-55.
35. Bolger N, Davis A, Rafaeli E. Diary methods: Capturing life as it is lived. Annu Rev Psychol. 2003; 54: 579-616.
36. Baddeley A, Eysenck MW, Anderson MC. Memory. Hove: Psychology Press, 2009.
37. Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Econ. 2004; 12:1203-10
38. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj-Brit Med J. 2009; 339.
39. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoeconomics. 2006; 24: 355-71
40. Stafford MR, Hareendran A, Ng-Mak DS, et al. EQ-5D-derived utility values for different levels of migraine severity from a UK sample of migraineurs. Health Qual Life Outcomes. 2012; 10: 65.
41. Kerr C, Lloyd EJ, Kosmas CE, et al. Health-related quality of life in Parkinson's: impact of 'off' time and stated treatment preferences. Quality of Life Research. 2016; 25: 1505-15.
42. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997; 35: 1095-108
43. Mulhern B, Rowen D, Jacoby A, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. Epilepsy Behav. 2012; 24: 36-43.

44. Dakin H, Burns R, Yang Y, 2016. HERC database of mapping studies, Version 5.0 (Last updated: 16th May 2016). Available at: http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies.
45. Parkin D, Rice N, Jacoby A, et al. Use of a visual analogue scale in a daily patient diary: modelling cross-sectional time-series data on health-related quality of life. Social Science & Medicine. 2004; 59: 351-60.
46. Janssen MF, Birnie E, Bonsel G. Feasibility and reliability of the annual profile method for deriving QALYs for short-term health conditions. Medical Decision Making. 2008; 28: 500-510