



The Psychometric Properties of a Self-Administered, Open-Source Module for Valuing Metastatic Epidural Spinal Cord Compression Utilities

Markian Pahuta¹ · Aaron Frombach² · Emile Hashem² · Stewart Spence² · Christina Sun² · Eugene K. Wai³ · Joel Werier³ · Carl van Walraven⁴ · Doug Coyle⁵

Published online: 3 September 2018
© The Author(s) 2018

Abstract

Introduction Web surveys are often used for utility valuation. Typically, custom utility valuation tools that have not undergone psychometric evaluation are used.

Objectives This study aimed to determine the psychometric properties of a metastatic epidural spinal cord compression (MESCC) module run on a customizable open-source, internet-based, self-directed utility valuation platform (Self-directed Online Assessment of Preferences [SOAP]).

Methods Individuals accompanying patients to the emergency department waiting room in Ottawa, Canada, were recruited. Participants made SOAP MESCC health state valuations in the waiting room and 48 h later at home. Validity, agreement reliability, and responsiveness were measured by logical consistency of responses, smallest detectable change, the interclass correlation coefficient, and Guyatt's responsiveness index, respectively.

Results Of 285 participants who completed utility valuations, only 113 (39.6%) completed the re-test. Of these 113 participants, 92 (81.4%) provided valid responses on the first test and 75 (66.4%) provided valid responses on the test and re-test. Agreement for all groups of health states was adequate, since their smallest detectable change was less than the minimal clinically important difference. The mean interclass correlation coefficients for all health states were > 0.8, indicating at least substantial reliability. Guyatt's responsiveness indices all exceeded 0.80, indicating a high level of responsiveness.

Conclusions To our knowledge, this is the first validated open-source, web-based, self-directed utility valuation module. We have demonstrated the SOAP MESCC module is valid, reproducible, and responsive for obtaining ex ante utilities. Considering the successful psychometric validation of the SOAP MESCC module, other investigators can consider developing modules for other diseases where direct utility valuation is needed.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s41669-018-0092-1>) contains supplementary material, which is available to authorized users.

✉ Markian Pahuta
mark@pahutamd.com

Aaron Frombach
aaf312@gmail.com

Emile Hashem
ehash070@uottawa.ca

Stewart Spence
sspen085@uottawa.ca

Christina Sun
christina.sun@uottawa.ca

Eugene K. Wai
ewai@toh.on.ca

Joel Werier
jwerier@toh.on.ca

Carl van Walraven
carlv@ohri.ca

Doug Coyle
doug.coyle@uottawa.ca

¹ Department of Orthopaedic Surgery, Henry Ford Health System, 2799 West Grand Blvd, CFP-6, Detroit, MI, USA

² The University of Ottawa, Ottawa, ON, Canada

³ Division of Orthopaedic Surgery, The Ottawa Hospital, Ottawa, ON, Canada

⁴ Department of Epidemiology and Community Medicine, Institute for Clinical Evaluative Sciences, Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, Canada

⁵ School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

Key Points for Decision Makers

The Self-directed Online Assessment of Preferences (SOAP) tool is freely available, open-source utility-elicitation software compatible with modern web browsers, including touch-screen mobile devices.

SOAP modules can easily be developed for other clinical scenarios.

The SOAP MESCC module is valid, reproducible, and responsive for *ex ante* utility elicitation.

1 Introduction

Quality-adjusted life-years (QALYs) are used to concurrently quantify morbidity and mortality within a single parameter [1]. For this reason, QALYs can facilitate the discussion of risks and benefits during patient counseling regarding treatment options [2]. To help make funding decisions, policy makers may also combine QALYs with cost estimates to calculate the incremental cost-effectiveness ratio [3]. QALYs are calculated using “utilities,” or health-related quality-of-life (HRQoL) weights, which are obtained by direct valuation or from generic health status measures [4].

The choice of utility valuation approach is driven by available data. Direct valuation is the classical approach in which individuals rate hypothetical health state descriptions using the time trade-off or standard gamble procedures [5]. These procedures can be used to measure utilities for very specific and uncommon health states. However, it can be cumbersome to develop valid health state descriptions for particular diseases. Alternatively, techniques have been developed to convert generic health status measures (e.g. EuroQol-5 Dimensions [EQ-5D], Short Form-6 Dimensions [SF-6D], or Health Utilities Index 3) to utilities [1]. Conversion of generic health state measures is advantageous because custom health state descriptions are not required. However, utilities can only be obtained for health states actually observed in a cohort of patients involved in the generic health survey.

Unfortunately, generic health scores have not been collected for many diseases, meaning direct valuation is necessary for measuring utilities. Best practices in economic evaluation are to recruit a sample of healthy individuals from the general population for utility valuation [6, 7]. Traditionally, general population utility valuation has been conducted using face-to-face interviews, phone interviews, or postal surveys [8]. These forms of survey administration are time intensive and costly, so web-based surveys

are increasingly being used [9–22]. Typically, these studies are conducted using proprietary software, which limits its application to other disease contexts. Furthermore, the psychometric properties of these propriety software programs have not been assessed [23].

It is important to determine whether web-based utility valuation has acceptable psychometric properties. If this mode of administration has acceptable psychometric properties, rather than building custom software for new utility valuation studies, it would be beneficial and efficient for investigators to be able to build disease-specific modules on a common platform that has been used to develop modules with acceptable psychometric properties. To meet this need, we developed a new open-source (non-proprietary), web-based, self-directed utility valuation platform useable on major computer systems (including touch-screen devices) called the Self-directed Online Assessment of Preferences (SOAP) (Appendix 1 and 2 in the Electronic Supplementary Material [ESM]). SOAP was designed with flexibility in mind and can accept new health state descriptions (modules) with minimal programming.

We decided to first create a SOAP module for metastatic epidural spinal cord compression (MESCC), a condition for which HRQoL data are limited. MESCC can be treated with surgery or radiotherapy, but few high-quality studies compare these interventions using generic health status measures for patients. However, surgery and radiotherapy outcomes could be compared using utilities obtained by direct valuation of hypothetical probe health state descriptions. The European Organisation for Research and Treatment of Cancer (EORTC) MESCC working group has developed an HRQoL questionnaire for MESCC [24]. Items from this questionnaire could be used to generate health state descriptions for a SOAP module.

The objective of this study was to determine whether the SOAP platform can be used to develop a valid, reproducible, and responsive module for MESCC. For this first application of the SOAP platform, we developed a MESCC module based on the work of the EORTC and measured psychometric properties in a general population sample.

2 Methods

2.1 Self-directed Online Assessment of Preferences (SOAP) Platform

Electronic utility valuation protocols are distinguished by the form of health state descriptions, assessment approach, navigation rules, and auxiliary functions [25]. A detailed description of these elements for the SOAP MESCC module are provided in Appendix 1 and 2 in the ESM.

2.2 Metastatic Epidural Spinal Cord Compression (MESCC) Module

EORTC phase I development of a MESCC questionnaire in Canada found that patients and healthcare providers felt that ambulation, urinary continence, pain, and independence were important HRQoL issues for MESCC. Since phase I development was restricted to HRQoL and did not specifically consider treatment effects and adverse events, we reviewed prospective studies on MESCC to identify reported outcomes and adverse events [26–29]. The EORTC items captured all treatment outcomes identified in our review. However, the review identified a large and disparate set of adverse effects. To develop a manageable decision analytic model, all adverse effects were grouped as an “other symptoms” attribute.

A tabular (point-form) presentation of health states was chosen as it is preferred by participants, is believed to decrease cognitive burden compared with the narrative format, and produces similar results to the narrative format [30, 31]. Therefore, we presented health states as a point-form list of five dysfunctional attributes: non-ambulatory (N), incontinent of urine (I), pain (P), dependent (D), and “other symptoms” (S). To reduce the number of potential health states, EORTC items were collapsed to indicate the presence (+) or absence (–) of the dysfunctional attribute, producing 32 discrete health states (Appendix 3 in the ESM).

When possible, the phrasing for presence or absence of dysfunctional attributes was created using the same EORTC items identified in the MESCC module development process (Table 1). Items were rephrased to the second person

and restructured as declarative sentences. Items describing feelings or worries were not utilized as we wanted to make the health states descriptions as objective as possible. The rationale for the specific attribute formulation was as follows:

1. Dependence (D). The two items identified by the MESCC working group were combined into one attribute to highlight the implications of loss of independence. The qualifiers “do” and “do not” were added to indicate complete function and dysfunction.
2. Lack of ambulation (N). The MESCC working group developed a new item that was used as the functional level. Again, two items were combined to highlight the implications of loss of mobility.
3. Incontinence of urine (I). The item identified by the MESCC working group with a qualifier was used as the functional level. An item from the EORTC bladder cancer module (BLM44) was used to highlight the implications of loss of bladder control.
4. Pain (P). As MESCC can only occur in the cervical spine to the thoracolumbar junction, pain was not differentiated by the terms “upper” and “lower” back as was identified by the MESCC working group. As most patients with spine metastasis will have some element of pain, the functional state had patients requiring pain medications. Use of pain medication served as a qualifier and was taken from the EORTC bone metastasis module (BM38).
5. Other symptoms (S). Again, to maintain efficiency, all adverse effects were characterized by several common adverse symptoms. These items were all taken from the core EORTC questionnaire.

Table 1 Health state attributes

Attributes	Symbols ^a	Items	EORTC codes ^b
Dependence	D–	You do not need help with eating, dressing, washing yourself or using the toilet. You are not dependent on others	C05, BM49
	D+	You need help with eating, dressing, washing yourself or using the toilet. You are dependent on others	C05, BM49
Lack of ambulation	N–	You walk without assistance	NA
	N+	You have lost mobility because of weakness in your legs. You need to stay in bed or a chair (or wheelchair) during the day	BM48, BN48, C04
Incontinence of urine	I–	You have no trouble controlling your bladder	BN50
	I+	You need to insert a tube in your bladder or you need to wear a diaper full time, to control your bladder	BN50, BLM44
Pain	P–	You have back pain relieved by pain medications	BM31, BM38
	P+	You have back pain not relieved by pain medications	BM31, BM38
Other symptoms	S–	You have no other uncomfortable symptoms	NA
	S+	You have one or more of: nausea, vomiting, shortness of breath, lack of appetite, diarrhea	C14, C15, C08, C17

EORTC European Organisation for Research and Treatment of Cancer, NA not applicable

^a + indicates the presence of a dysfunctional attribute, – indicates an absence

^b Codes refer to identifiers in the EORTC item bank

Valuations were obtained using the standard gamble method using a ping-pong search algorithm. In the standard gamble, success is typically framed as “perfect” health for an undetermined period of time. In this context, this can be inferred to be the absence of any dysfunctions. Therefore, the fully functional health state (D-, N-, I-, P-, S-) was chosen as the success anchor. To eliminate confusion around life expectancy, all scenarios were framed as having a certain life expectancy of 5 years; that is, for both the probe health scenario and success health scenario, participants were told their life expectancy would certainly be 5 years. This was the maximum survival reported in a randomized controlled trial on treatments for MESCC [27]. Probe health states were presented in a random order.

The MESCC module was pilot tested in a sample of 40 participants to assess acceptability and ease of use. Participants were asked to rate the SOAP MESCC module using a five-point Likert rating for the statement “[t]his website is easy to use”, and 92.5% of participants strongly agreed or agreed with the statement.

2.3 Subjects

To be compliant with best practice in economic evaluation, we sought to conduct a direct utility valuation study with a sample of the general population who have not experienced MESCC using the SOAP MESCC module (ex ante valuation) [6]. Prior to this general population direct valuation study, psychometric properties of the SOAP MESCC module had to be evaluated. To approximate a general population sample for this psychometric validation study, participants were recruited from the emergency department waiting rooms at The Ottawa Hospital, an academic hospital in Ottawa, Ontario, Canada. Only patients’ family members or friends (i.e. individuals accompanying patients) aged ≥ 18 years were eligible to participate. Participants were required to be able to read English and have access to the internet outside of the hospital. A minimum sample size of 50 participants has been recommended in published guidelines for reliability and responsiveness evaluations. [23]. To ensure robust results, we set the sample size for this study at 75.

2.4 Survey Procedures

Participants completed the first survey in the emergency department using a touch-screen device. Investigators did not assist participants in navigating or completing the survey. Each participant valued the health state D+N+ I+P+S+, one randomly selected singly dysfunctional health state, and another triply dysfunctional health state. Dysfunctional elements were nested to ensure a logical ordering of utilities

for the three health states. For example, if the singly dysfunctional health state was D-N-I+P-S-, the triply dysfunctional health state could include incontinence and two of dependence, lack of ambulation, pain, or other symptoms.

Investigators contacted participants via email and/or phone 2 days after the initial survey with information to access the retest. Participants completed the second survey using their personal device. For the retest, participants were presented with the same probe health states they completed in the emergency room, but states were presented in a new random order.

2.5 Statistical Analysis

“Validity” refers to whether a tool under investigation measures what it is supposed to measure [32]. Specifically, “construct validity” concerns whether results obtained using the tool under investigation are consistent with a priori hypotheses [32]. We hypothesized that utility valuations should follow the logical ordering of health states, with utility valuations following the relationship: singly \geq triply \geq fully dysfunctional. We considered singly = triply = fully a valid response because we could not exclude the possibility of a ceiling effect with one dysfunction. Participant responses were deemed “valid” if their utilities followed this order. The proportion of participants providing valid responses on the test and retest was computed.

“Reproducibility” concerns the stability of participants’ responses on repeated testing and can be characterized by agreement and reliability [32]. “Agreement” quantifies the absolute differences in participants’ repeated responses. We assessed agreement using the smallest detectable change [23]. We classified agreement as adequate if the smallest detectable change was less than the minimal clinically important difference (MCID) [23]. By anchoring to Eastern Cancer Oncology Group functional levels, an MCID of 0.05 for cancer utilities obtained by the three-level EQ-5D (EQ-5D-3L) has been proposed [33]. This MCID has also been used for direct utility valuation by the standard gamble and time trade-off of EQ-5D-3L health states [34]. The precision of the standard gamble algorithm used in our study was also 0.05. Therefore, we used an MCID of 0.05 in this study. Systematic differences between the test and retest sessions were quantified using the smallest detectable change calculation. “Reliability” concerns the fraction of pooled study variance across the repeated tests attributable to differences between participants (participant variance) and individual test–retest variability (noise) [32]. If responses are stable, the ratio of noise to participant variance should be small, and the ratio of participant variance to variance for the pooled results from test and retest should be high. Reliability accounting for systematic differences between the test and retest, stratified for the number of dysfunctions in the health state, was

quantified using the intraclass correlation coefficient (ICC) using the following categories: < 0.21, slight reliability; 0.21–0.40, fair reliability; 0.41–0.60, moderate reliability; 0.61–0.80, substantial reliability; > 0.80, almost perfect reliability [35]. An ICC \geq 0.70 was considered adequate [23].

“Responsiveness” reflects the ability of a tool to detect clinically important changes and can be quantified using Guyatt’s responsiveness index [36]. This index is proportional to the ratio of the MCID to the root mean squared error of the difference between the test and retest value. If test–retest variability is small relative to the MCID, the tool is deemed responsive because meaningful changes are of greater magnitude than test–retest fluctuation [37]. Values of 0.20, 0.50, and 0.80 were interpreted as small, moderate, and large levels of responsiveness, respectively [38].

Statistical analysis was performed using the statistical programming language R [39]. The distribution of age (Kruskal–Wallis test) and sex (Chi squared test) was compared between participants providing valid and invalid responses on the test and retest. We considered logically ordered responses to be valid, that is decreasing utilities assigned to the singly, triply, and fully dysfunctional states. Age was assessed using a one-way analysis of variance (ANOVA). Sex was assessed using the Chi squared test. Reproducibility, agreement, reliability, and responsiveness were only measured for participants providing valid responses on both the test and the retest. Since the SOAP tool is intended for measuring average utilities from the general public, average measures (rather than individual measures) of smallest detectable change, ICCs, and Guyatt’s responsiveness indices were calculated [40].

3 Results

Of 285 participants who completed utility valuations in the emergency department, only 113 (39.6%) completed the retest. Of these 113 participants, 92 (81.4%) provided valid responses on the first test, and 75 (66.4%) provided valid responses on the test and retest (Table 2). The response validity pattern was not associated with age ($p=0.2336$) or sex ($p=0.971$) (Table 2). Only data from the participants providing valid responses on both the test and the retest were

Table 2 Characteristics of participants stratified by response pattern

Characteristics	Valid → valid	Valid → invalid	Invalid → valid	Invalid → invalid	<i>p</i> value
Total	75 (66.4)	17 (15.0)	17 (15.0)	4 (3.6)	
Age in years	53.0 [33.5, 61.0]	43.0 [33.0, 48.0]	51.0 [38.0, 62.0]	43.0 [33.25, 52.0]	0.2336
Sex					0.971
Female	47 (62.7)	11 (64.7)	11 (58.8)	3 (75.0)	
Male	28 (37.3)	6 (35.3)	6 (41.2)	1 (25.0)	

Data are presented as *N* (%), except for age, which is presented as median [interquartile range]

Table 3 Agreement, reliability, and responsiveness measurements

Number of dysfunctions	Smallest detectable change	ICC (95% CI)	Guyatt’s responsiveness index
Single	0.036	0.879 (0.808–0.923)	1.45
Triple	0.035	0.898 (0.838–0.935)	1.47
Full	0.042	0.826 (0.724–0.890)	1.24

CI confidence interval, ICC intraclass correlation coefficient

used for reproducibility and responsiveness analysis. Seven respondents skipped at least one scenario during the test and were classified as providing invalid responses. Only one respondent skipped one question during the retest, and the responses were also classified as invalid.

Agreement for all groups of health states was adequate since their smallest detectable change was less than the MCID of 0.05 (Table 3). Mean ICCs were all > 0.8, indicating substantial reliability, and all ICCs were significantly greater than the pre-specified threshold of 0.7 (Table 3). Guyatt’s responsiveness indices all exceeded 0.80, indicating large responsiveness for the utility evaluation (Table 3) [38].

4 Discussion

Utility valuation studies are traditionally conducted using face-to-face interviews, phone interviews, or postal surveys. These modes of administration have undergone psychometric validation. Web surveys are increasingly used for utility valuation and usually use custom and proprietary valuation tools that have not been psychometrically validated. It would be beneficial and efficient for investigators to be able to build disease-specific modules on a common platform that has been used to develop modules with acceptable psychometric properties.

We developed a new platform called the SOAP (Appendix 1 and 2 in the ESM). For the first application of this platform, we developed a module for MESCC health states. The SOAP platform met published benchmarks for reproducibility (both agreement and reliability) and responsiveness for utility measurement. This study demonstrated that

the SOAP platform can be used to develop modules with acceptable psychometric properties.

In total, 81.4% of participants provided valid responses on the first test, and 66.4% of participants provided valid responses on both the test and the retest. These results should be considered in the context of other *ex ante* valuation studies reported in the literature. We classified a participant's responses as valid if their utility valuations decreased with increasing dysfunctional attributes in the health state. For example, if a participant valued the fully dysfunctional health state higher than the single dysfunctional health state, their responses were classified as invalid. This definition of validity is termed "logical consistency" and has been used in traditional general population *ex ante* utility valuation studies of EQ-5D-3L health states.

Logical consistency rates for face-to-face valuations have been reported for the UK and Netherlands [41, 42]. In the UK study, 12 pairs of health states per participant could be evaluated for logical consistency. The median rate of logical consistency, per participant, ranged from 83.8 to 91.7%. In the Dutch study, 87.6% of participants provided at least one pair of logically inconsistent valuations. Postal surveys conducted in the USA and New Zealand reported at least one logically inconsistent pairing in 88 and 79% of participants, respectively [43, 44]. With 81.4% of participants providing a valid response (28.6% providing a logically inconsistent response), the logical consistency rate for the SOAP MESCC module was similar to that of traditional population studies. Logical consistency has also been assessed for other self-administered general population *ex ante* utility valuation studies of EQ-5D-3L health states over the internet [19, 45, 46]. Each study reported a logical consistency rate < 70%.

Compared with the SOAP MESCC module, the face-to-face, postal, and web-based EQ-5D-3L utility valuation studies required greater cognitive effort because participants rated more health states (between five and ten) that were also more complex (five attributes and three levels of dysfunction). Furthermore, these studies did not provide error checking, whereas the SOAP MESCC module notified participants of a logical error if they rejected a lottery with 100% of success. Considering these differences, a logical consistency rate of 81.4% on the first test with the SOAP MESCC module is consistent with the literature.

Valuing MESCC health states using the classical standard gamble is problematic for two reasons. First, the classical standard gamble uses perfect health as a top anchor, which is an unrealistic outcome for metastatic cancer. Second, the classical standard gamble considers timeless (i.e. perpetual) health states, which is incongruent with the metastatic cancer disease process. To make the standard gamble more realistic, we characterised perfect health as the absence of dysfunctions and restricted all health states (including the top anchor) to a survival period of 5 years. These modifications

may affect the interpretation of our results relative to classic utility assessment.

Utilities are typically estimated for specific health states and are used to weight the time in such health states. Consequently, a utility value for a specific state is typically considered "timeless," that is, utilities are usually assumed not to change with time spent in a health state [47]. As a reflection of this, the duration of time spent in a probe health state is not specified in the classical standard gamble [5]. For MESCC health states, we were concerned that the most severe health states would connote poor survival and therefore confound the measurement of HRQoL using the standard gamble with quantity of life. To alleviate this difficulty, we explicitly stated a 5-year duration for each health state, which was the longest survival observed in a randomized controlled trial of treatments for MESCC [27]. This approach has also been used in other utility valuation studies for cancer health states [48]. This modification to health state descriptions should not affect results because the standard gamble (and all other utility-elicitation methods) relies on the "utility independence" assumption [49]. Under this assumption, if a health state has a utility of x , the utility of this health state for 5 years should still be x . Unfortunately, a systematic review concluded that individuals tend not to satisfy the utility independence assumption with no consistent pattern of violation [50]. We are unaware of any algorithm to convert utilities for fixed period of time to "timeless" utilities. Consequently, the utilities measured in this study may not be directly comparable to utilities obtained using the classical standard gamble.

A strength of our study is that we built on the work conducted by the EORTC MESCC working group to ensure the attributes in the MESCC module were appropriate and representative of the MESCC disease process. A limitation of our study is that we did not assess criterion validity by comparing utilities obtained by SOAP MESCC and a "gold standard" [32]. This could be done by having patients with MESCC value their own health using the SOAP MESCC module and comparing these utility valuations with those derived from a generic health questionnaire. We did not have the resources to conduct such a study. Furthermore, measures of logical validity, reproducibility, and responsiveness are more relevant than MESCC criterion validity to investigators considering developing modules for new diseases.

To our knowledge, this is the first validated open-source, web-based, self-directed utility valuation module. For the first application of the SOAP platform, we developed a module for MESCC health states. We have demonstrated the SOAP MESCC module is valid, reproducible, and responsive for obtaining *ex ante* utilities. Considering the successful psychometric validation of the SOAP MESCC module, other investigators can consider developing modules for other diseases where direct utility valuation is needed.

Author Contributions Study conception and design: MP, JW, EW, CvW, DC. Acquisition of data: MP, AF, EH, SS, CS. Analysis and interpretation of data: MP, JW, EW, CvW, DC. Drafting of manuscript and critical revision: MP, JW, EW, CvW, DC.

Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of The Ottawa Hospital Research Ethics Board and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Funding Financial support for this study was provided entirely by a grant from the Hans K. Uthoff Fund. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report.

Conflict of interest Markian Pahuta, Aaron Frombach, Emile Hashem, Stewart Spence, Christina Sun, Eugene K. Wai, Joel Werier, Carl van Walraven, and Doug Coyle have no conflicts of interest.

Data availability statement The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*. USA: Oxford University Press; 2005.
2. Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health*. 2009;12(Suppl 1):S27–30.
3. Cleemput I, Neyt M, Thiry N, De Laet C, Leys M. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *Int J Technol Assess Health Care*. 2011;27(1):71–6.
4. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health*. 2009;12(1):S5–9.
5. Furlong W, Feeny D, Torrance GW, Barr R, Horsman J. *Guide to design and development of health-state utility instrumentation*. Westminster: Hamilton; 1990.
6. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA*. 1996;276(15):1253–8.
7. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, Kuntz KM, Meltzer DO, Owens DK, Prosser LA, Salomon JA, Sculpher MJ, Trikalinos TA, Russell LB, Siegel JE, Ganiats TG. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *J Am Med Assoc*. 2016;316(10):1093–103.
8. Engel L, Bansback N, Bryan S, Doyle-Waters MM, Whitehurst DGT. Exclusion criteria in national health state valuation studies: a systematic review. *Med Decis Mak*. 2016;36(7):798–810.
9. Evans M, Jensen HH, Bogelund M, Gundgaard J, Chubb B, Khunti K. Flexible insulin dosing improves health-related quality-of-life (HRQoL): a time trade-off survey. *J Med Econ*. 2013;16(11):1357–65.
10. Evans M, Khunti K, Mamdani M, Galbo-Jorgensen CB, Gundgaard J, Bogelund M, Harris S. Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. *Health Qual Life Outcomes*. 2013;11:90.
11. Prosser LA, Payne K, Rusinak D, Shi P, Messonnier M. Using a discrete choice experiment to elicit time trade-off and willingness-to-pay amounts for influenza health-related quality of life at different ages. *Pharmacoeconomics*. 2013;31(4):305–15.
12. Harris S, Mamdani M, Galbo-Jorgensen CB, Bogelund M, Gundgaard J, Groleau D. The effect of hypoglycemia on health-related quality of life: Canadian results from a multinational time trade-off survey. *Can J Diabetes*. 2014;38(1):45–52.
13. Lloyd JC, Yen T, Pietrobon R, Wiener JS, Ross SS, Kokorowski PJ, Nelson CP, Routh JC. Estimating utility values for vesicoureteral reflux in the general public using an online tool. *J Pediatr Urol*. 2014;10(6):1026–31.
14. Sears ED, Shin R, Prosser LA, Chung KC. Economic analysis of revision amputation and replantation treatment of finger amputation injuries. *Plast Reconstr Surg*. 2014;133(4):827–40.
15. van Nooten FE, Koolman X, Busschbach JJV, Brouwer WBF. Thirty down, only ten to go?! Awareness and influence of a 10-year time frame in TTO. *Qual Life Res*. 2014;23(2):377–84.
16. Goodwin E, Green C, Spencer A. Estimating a Preference-Based Index for an eight-dimensional health state classification system for multiple sclerosis. *Value Health*. 2015;18(8):1025–36.
17. Hutchins R, Viera AJ, Sheridan SL, Pignone MP. Quantifying the utility of taking pills for cardiovascular prevention. *Circ Cardiovasc Qual Outcomes*. 2015;8(2):155–63.
18. van Nooten FE, van Exel NJA, Koolman X, Brouwer WBF. 'Married with children' the influence of significant others in TTO exercises. *Health Qual Life Outcomes*. 2015;13:94.
19. Augestad LA, Stavem K, Kristiansen IS, Samuelsen CH, Rand-Hendriksen K. Influenced from the start: anchoring bias in time trade-off valuations. *Qual Life Res*. 2016;25(9):2179–91.
20. Jorgensen TR, Emborg C, Dahlen K, Bogelund M, Carlborg A. The effect of the medicine administration route on health-related quality of life: results from a time trade-off survey in patients with bipolar disorder or schizophrenia in 2 Nordic countries. *BMC Psychiatry*. 2016;16:244.
21. Olofsson S, Norrlid H, Persson U. Preferences for improvements in attributes associated with basal insulin: a time trade-off and willingness-to-pay survey of a diabetic and non-diabetic population in Sweden. *J Med Econ*. 2016;19(10):945–58.
22. Weernink MGM, Groothuis-Oudshoorn CGM. Valuing treatments for Parkinson disease incorporating process utility: performance of best-worst scaling, time trade-off, and visual analogue scales. *Value Health*. 2016;19(2):226–32.
23. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
24. Mitera G, Loblaw A, Sahgal A, Danielson B. Canadian-led international development of a European organization for research and treatment of cancer quality of life module for malignant spinal cord compression: results of phase I. *Int J Radiat Oncol*. 2010;78(3):S604.

25. Lenert LA, Sturley A, Watson ME. iMPACT3: internet-based development and administration of utility elicitation protocols. *Med Decis Making*. 2002;22:464–74.
26. Maranzano E, Bellavita R, Rossi R, De Angelis V, Frattegiani A, Bagnoli R, Mignogna M, Beneventi S, Lupattelli M, Ponticelli P, Biti GP, Latini P. Short-course versus split-course radiotherapy in metastatic spinal cord compression: results of a phase III, randomized, multicenter trial. *J Clin Oncol*. 2005;23(15):3358–65.
27. Patchell RA, Tibbs PA, Regine WF, Payne R, Saris S, Kryscio RJ, Mohiuddin M, Young B. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomised trial. *Lancet (London, England)*. 2005;366:643–8.
28. Maranzano E, Trippa F, Casale M, Costantini S, Lupattelli M, Bellavita R, Marafioti L, Pergolizzi S, Santacaterina A, Mignogna M, Silvano G, Fusco V. 8 Gy single-dose radiotherapy is effective in metastatic spinal cord compression: results of a phase III randomized multicentre Italian trial. *Radiother Oncol*. 2009;93(2):174–9.
29. Fehlings MG, Nater A, Tetreault L, Kopjar B, Arnold P, Dekutoski M, Finkelstein J, Fisher C, France J, Gokaslan Z, Massicotte E, Rhines L, Rose P, Sahgal A, Schuster J, Vaccaro A. Survival and clinical outcomes in surgically treated patients with metastatic epidural spinal cord compression: results of the prospective multicenter AOSpine Study. *J Clin Oncol*. 2015;34:268–76.
30. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. Describing health states. Methodologic issues in obtaining values for health states. *Med Care*. 1984;22(6):543–52.
31. Schünemann HJ, Ståhl E, Austin P, Akl E, Armstrong D, Guyatt GH. A comparison of narrative and table formats for presenting hypothetical health states to patients with gastrointestinal or pulmonary disease. *Med Decis Mak*. 2004;24:53–60.
32. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737–45.
33. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes*. 2007;5(1):70.
34. Wee H-L, Li S-C, Xie F, Zhang X-H, Luo N, Feeny D, Cheung Y-B, Machin D, Fong K-Y, Thumboo J. Validity, feasibility and acceptability of time trade-off and standard gamble assessments in health valuation studies: a study in a multiethnic Asian population in Singapore. *Value Health*. 2008;11(Suppl 1):S3–10.
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
36. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol*. 1989;42(5):403–8.
37. Guyatt G, Walter S, Norman G. Measuring change over time—assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40(2):171–8.
38. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness. *J Clin Epidemiol*. 2000;53(5):459–68.
39. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.
40. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care*. 2001;17:479–87.
41. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ*. 1996;15(2):209–31.
42. Lamers LM, Stalmeier PFM, Krabbe PFM, Busschbach JJV. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Mak*. 2006;26(2):173–81.
43. Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Med Care*. 2005;43(3):221–8.
44. Devlin NJ, Hansen P, Kind P, Williams A. Logical inconsistencies in survey respondents' health state valuations—a methodological challenge for estimating social tariffs. *Health Econ*. 2003;12(7):529–44.
45. Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PLoS One*. 2012;7(2):e31115.
46. Augustovski F, Rey-Ares L, Irazola V, Oppe M, Devlin NJ. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ*. 2013;14(Suppl 1):S25–31.
47. Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien BJ. Are health states 'timeless'? The case of the standard gamble method. *J Clin Epidemiol*. 1999;52(11):1047–53.
48. Stiggelbout AM, Kiebert GM, Kievit J, Leer JW, Stoter G, de Haes JC. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Mak*. 1994;14:82–90.
49. Pliskin J, Shepard D, Weinstein M. Utility functions for life years and health status. *Oper Res*. 1980;28(1):206–24.
50. Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Med Decis Mak*. 2005;25(4):460–7.