

## HEALTH SERVICES RESEARCH

OPEN

## Calculating Utilities From the Spine Oncology Study Group Outcomes Questionnaire

*A Necessity for Economic and Decision Analysis*

Markian A. Pahuta, MD, PhD,<sup>a</sup> Felicity Fisk, MD,<sup>b</sup> Anne L. Versteeg, MD, MHSc, PhD,<sup>c</sup> Charles G. Fisher, MD, MHSc,<sup>d</sup> Arjun Sahgal, MD,<sup>e</sup> Ziya L. Gokaslan, MD,<sup>f</sup> Jeremy J. Reynolds, MD,<sup>g</sup> Ilya Laufer, MD,<sup>h</sup> Aron Lazary, MD,<sup>i</sup> Laurence D. Rhines, MD,<sup>j</sup> Stefano Boriani, MD,<sup>k</sup> Chetan Bettegowda, MD, PhD,<sup>l</sup> and Nicolas Dea, MD, MSc<sup>d</sup>, on behalf of the AO Spine Knowledge Forum Tumor<sup>f</sup>

**Study Design.** General population utility valuation study.

**Objective.** The aim of this study was to develop a technique for calculating utilities from the Spine Oncology Study Group Outcomes Questionnaire v2.0 (SOSGOQ2.0).

From the <sup>a</sup>Division of Orthopedic Surgery, McMaster University, Hamilton, ON, Canada; <sup>b</sup>Department of Orthopedics, Henry Ford Health System, Detroit, MI; <sup>c</sup>Department of Radiation Oncology, University Medical Center Utrecht, Utrecht, The Netherlands and Department of Orthopaedic Surgery, University of Toronto, Toronto, Ontario, Canada; <sup>d</sup>Division of Spine, Surgery Vancouver Spine Surgery Institute, Vancouver General Hospital University of British Columbia, Vancouver, Canada; <sup>e</sup>Department of Radiation Oncology, Sunnybrook Odette Cancer Center and University of Toronto, Toronto, Ontario, Canada; <sup>f</sup>Department of Neurosurgery, The Warren Alpert Medical School of Brown University and Rhode Island Hospital and The Miriam Hospital, Providence, RI; <sup>g</sup>Oxford University Hospitals NHS Foundation Trust, Oxford, UK; <sup>h</sup>Department of Neurosurgery, NYU Grossman School of Medicine, New York, NY; <sup>i</sup>National Center for Spinal Disorders and Buda Health Center, Budapest, Hungary; <sup>j</sup>Department of Neurosurgery, Division of Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX; <sup>k</sup>Spine Surgery Division, IRCCS Istituto Ortopedico Galeazzi, Milan, Italy; and <sup>l</sup>Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD.

Acknowledgment date: June 23, 2020. First revision date: October 19, 2020. Second revision date: November 19, 2020. Third revision date: December 22, 2020. Acceptance date: December 24, 2020.

The manuscript submitted does not contain information about medical device(s)/drug(s).

Relevant financial activities outside the submitted work: board membership, consultancy, grants, travel/accommodations/meeting expenses.

This study was organized and funded by AO Spine through the AO Spine Knowledge Forum Tumor, a focused group of international spine oncology experts. AO Spine is a clinical division of the AO Foundation which is an independent medically-guided not-for-profit organization. Study support was provided directly through the AO Spine Research Department.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Address correspondence and reprint requests to Markian A. Pahuta, MD, PhD, Division of Orthopedic Surgery, McMaster University, 237 Barton Street East, Floor 5N, Hamilton, ON L8L 2X2, Canada; E-mail: pahutam@mcmaster.ca

DOI: 10.1097/BRS.0000000000003981

Spine

**Summary of Background Data.** The ability to calculate quality-adjusted life-years (QALYs) for metastatic spine disease would enhance treatment decision-making and facilitate economic analysis. QALYs are calculated using *utilities*.

**Methods.** Using a hybrid concept-retention and factorial analysis shortening approach, we first shortened the SOSGOQ2.0 to eight items (SOSGOQ-8D). This was done to lessen the cognitive burden of the utility valuation exercise. A general population sample of 2730 adults was then asked to evaluate 12 choice sets based on SOSGOQ-8D health states in a Discrete Choice Experiment. A utility scoring rubric was then developed using a mixed multinomial-logit regression model.

**Results.** We were able to reduce the SOSGOQ2.0 to an SOSGOQ-8D with a mean error of 0.003 and mean absolute error of 3.078 compared to the full questionnaire. The regression model demonstrated good predictive performance and was used to develop a utility scoring rubric. Regression results revealed that participants did not regard all SOSGOQ-8D items as equally important.

**Conclusion.** We provide a simple technique for converting the SOSGOQ2.0 to utilities. The ability to evaluate QALYs in metastatic spine disease will facilitate economic analysis and patient counseling. We also quantify the importance of individual SOSGOQ-8D items. Clinicians should heed these findings and offer treatments that maximize function in the most important items.

**Key words:** decision-making, health economics, health services research, health related quality-of-life, neoplasm metastasis, quality of life, quality-adjusted life years, resource allocation, spinal neoplasms, spine, survey, surveys and questionnaires, utilities.

**Level of Evidence:** 3  
**Spine 2021;46:1165–1171**

Treatment decisions for spine metastases, the most common site of skeletal metastases, are challenging because multiple patient and treatment-related factors including performance status, survival and risk of adverse events, must be considered.<sup>1</sup> Furthermore, with

limited life expectancy and scarce resources, the value of care in these patients is hotly debated.<sup>2,3</sup> Due to this controversy the economic value of metastatic spine care is desperately needed.<sup>4,5</sup> The ability to calculate quality-adjusted life-years (QALYs) for metastatic spine disease would facilitate economic analysis and thus enhance treatment decision-making and resource utilization.

QALYs are required in economic analysis because economic decisions are based on the incremental cost-effectiveness ratio which is simply the cost per QALY gained.<sup>6</sup> QALY analysis could also help patients and clinicians jointly assess the trade-offs between survival, health-related quality-of-life (HRQoL) benefits, recovery, and potential complications to reach an optimal treatment decision.<sup>6,7</sup> QALYs are calculated using *utilities* which are a number, typically between 0 and 1, that quantifies the preference for (*i.e.*, desirability of) a health state.<sup>6</sup> The utility of perfect health is set at 1 and the utility of dead is set at 0. Studies on metastatic spine disease have calculated utilities by converting to generic health measure responses (*e.g.*, the EQ-5D, SF-6D, and HUI-3) to utilities using multiattribute utility functions (MAUFs).

Literature examining the economic value of treatments for spinal metastases has been inconclusive. Uncertainty regarding the economic value of treating metastatic spine disease may be a consequence of psychometric limitations of generic health measures, which demonstrate suboptimal validity and responsiveness in conditions relevant to spinal metastases: cancer, spinal cord injury, and musculoskeletal disorders.<sup>6,8-13</sup> In an effort to improve HRQoL assessments in patients with metastatic spine disease, the AO Spine Knowledge Forum Tumor (former Spine Oncology Study Group, SOSG) developed, revised, and validated a spine oncology-specific outcome questionnaire (SOSGOQ2.0).<sup>14</sup> However, since a MAUF has not been developed for the SOSGOQ2.0, this outcome measure cannot be used to calculate QALYs.

We report the development of a MAUF for the SOSGOQ2.0. We first performed shortening of the SOSGOQ2.0 to eight items (SOSGOQ-8D) which was needed to minimize the cognitive burden of the utility valuation exercise.<sup>6</sup> We then conduct a utility valuation exercise to develop and validate a MAUF for the SOSGOQ-8D using a Discrete Choice Experiment (DCE) with a general population sample.

## MATERIALS AND METHODS

### Descriptive System

A hybrid concept-retention and factorial analysis approach was used for shortening the SOSGOQ2.0 to the SOSGOQ-8D.<sup>15,16</sup> To maintain content validity and clinical relevance, we planned to retain all four neurologic function single questions (Neurologic:Legs, Neurologic:Arms, Neurologic:Bladder, and Neurologic:Bowel). Therefore, it was necessary to select a single item from each of the remaining four domains (Physical Function, Pain, Mental Health, Social Function) in the SOSGOQ2.0 to achieve an eight item

questionnaire. The full analysis is reported in Supplemental Digital Content 1, Text, <http://links.lww.com/BRS/B730>.

The optimal SOSGOQ-8D questions for the non-neurologic domains were: items 3 (self-care), 11 (average pain), 17 (anxiety), 19 (personal relationships). These items were included in the final SOSGOQ-8D which is shown in Table 1.

### Subjects

Participants were recruited from an online market research panel (Toluna Influencers).<sup>17</sup> Quota sampling was used to ensure that the study sample was representative of the general United States population in terms of region, gender, and age based on 2017 United States Census Bureau Population Estimates Program data.<sup>18</sup> The utility valuation protocol was submitted for IRB review and deemed exempt

### Health States

The SOSGOQ-8D was converted to a set of health states consisting of eight attributes corresponding to each of the SOSGOQ-8D items and (the duration of survival in the given health state).<sup>19</sup> SOSGOQ-8D rating level phrasing was used in the health state descriptions. Duration of survival was set at: 1 year, 2 years, 5 years, and 10 years.<sup>20</sup> Health states were phrased in the second person and structured as declarative sentences.<sup>21</sup>

### Discrete Choice Experiment Valuation Task

Utility valuation was conducted using a DCE questionnaire. DCE methodology is simpler than traditional utility valuation and is therefore better suited for online studies.<sup>22,23</sup> In the DCEs for this study, participants were presented with pairs of health states (choice sets) and asked to select the more desirable health state. Choice sets were presented in a table with differing attributes highlighted (Figure 1).<sup>20</sup>

### Choice Set Selection

To select a manageable subset of choice sets, we first removed clinically unrealistic health states to form a pool of clinically relevant health states. Next, to lessen the cognitive burden of the DCEs, choice sets were developed in which no more than five attributes differed between health states.<sup>20</sup> A D-efficient collection of 100 nondominated choice sets was organized into blocks of 10 blocks using the modified Federov algorithm with Ngene software (Supplemental Digital Content 6, Table, <http://links.lww.com/BRS/B735>).<sup>22,24</sup> The design was developed using parameter values from a general population utility valuation study for a cancer-specific quality of life questionnaire using DCE methodology.<sup>20</sup> To assess whether participants understood the DCE task, one dominated choice set (one health state is clearly preferable) was added to each block to test for logic. To assess whether participants engaged in the DCE task, one choice set was repeated in each block, but with health state order reversed to test for internal consistency. Therefore, there were a total of 12 choice sets in each block. There were three levels of randomization in the survey. First, participants were randomized to one of the 10 blocks.

**TABLE 1. The Eight-item Spine Oncology Study Group Outcomes Questionnaire**

Domain (Abbreviation)	Level	Stem	Descriptor
Physical Function (PF)	1	Your spine limits your ability to care for yourself...	Not at all
	2		A little bit
	3		Somewhat
	4		Quite a bit
	5		Very much
Neurologic: Legs (L)	1	You have weakness in your legs...	None
	2		Mild occasionally
	3		Mild constantly
	4		Moderate constantly
	5		Severe constantly
Neurologic: Arms (A)	1	You have weakness in your arms...	None
	2		Mild occasionally
	3		Mild constantly
	4		Moderate constantly
	5		Severe constantly
Neurologic: Bowel (Bow)	1	You have difficulty controlling your bowel function beyond episodes of diarrhea/constipation...	Never
	2		Rarely
	3		Sometimes
	4		Often
	5		Very often
Neurologic: Bladder (Blad)	1	You have difficulty controlling your bladder function...	Never
	2		Rarely
	3		Sometimes
	4		Often
	5		Need a catheter
Pain (P)	1	On average, your back/neck pain is...	None
	2		Very mild
	3		Mild
	4		Moderate
	5		Severe
Mental Health (MH)	1	You feel anxious about your health related to your spine...	Never
	2		Rarely
	3		Sometimes
	4		Often
	5		Very often
Social Function (SF)	1	Your spine condition affects your personal relationships...	Never
	2		Rarely
	3		Sometimes
	4		Often
	5		Very often

Second, the order choice set in each block was randomized. Finally health state order was randomized in each choice set.

**Survey Procedures**

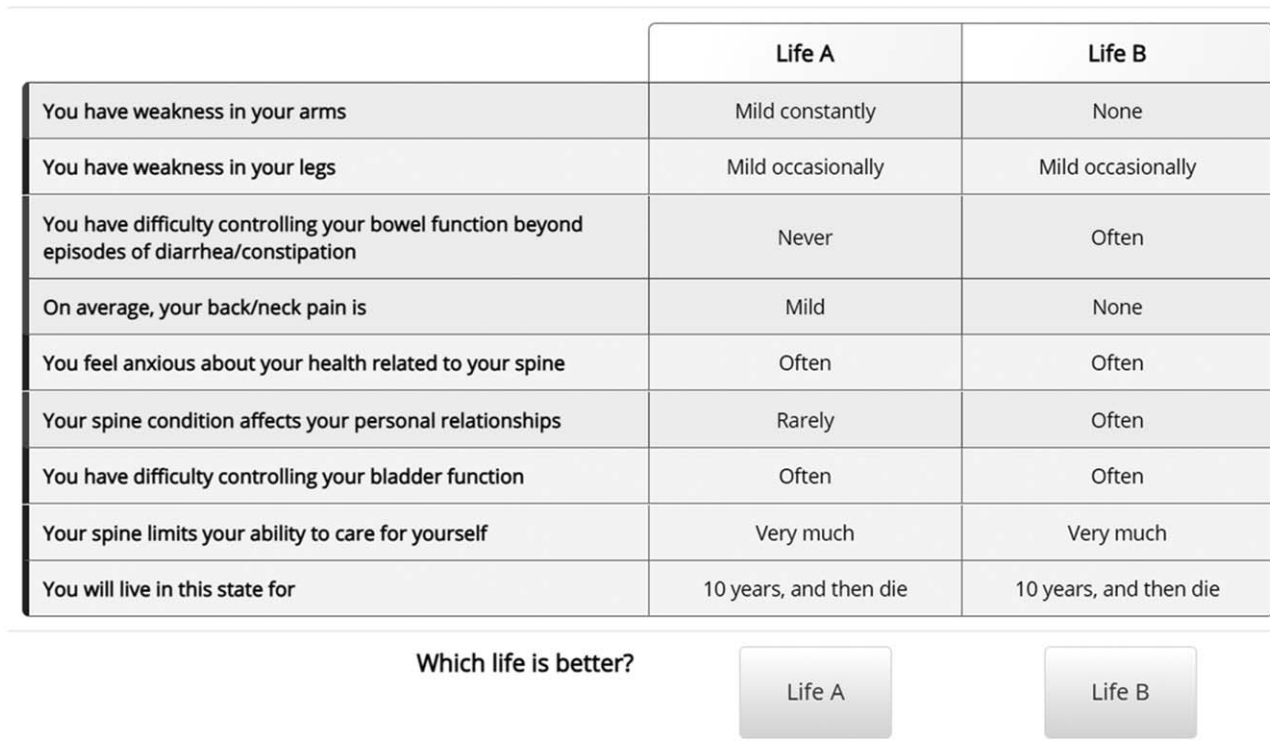
The market research company sent panel members an e-mail invitation to participate in our study. Interested panel members were redirected to a secure website hosting the utility valuation exercise.<sup>17,21</sup> Participants first read brief background information on metastatic spine disease. Next, participants were provided with an explanation of DCEs and shown a worked example. Participants then completed a practice DCE and provided feedback before completing the study DCEs. The

website closed by asking patients to provide five-point Likert rating for the statement “[t]his survey was difficult.”

**Statistical Analysis**

Participants who spent an average of at least eight seconds per choice set; selected the clearly dominant alternative in the dominated choice; and provided consistent responses for the repeat choice set were deemed to have engaged in and understood the DCE tasks. Only these participants were included in analyses.<sup>17,19,25</sup>

A multiattribute utility function was estimated from DCE responses using a mixed multinomial-logit regression model



**Figure 1.** Choice set presentation in online Discrete Choice Experiment. Differing attributes highlighted in green.

(MIXL) using the “mixl” library in the statistical programming language R.<sup>19,26–29</sup> The regression model incorporated the main survival duration effect, and two-way interactions between survival duration and each SOSGOQ-8D item.<sup>19</sup> Each parameter was treated as a random effect to account for participant heterogeneity in the repeated DCE tasks. The random effects were modeled with 1000 draws from a normal distribution. In the base regression model, all SOSGOQ-8D items were coded as nominal categorical (dummy) predictors to avoid assumptions of linear or extra-linear effects. The base regression model was simplified by removing nonsignificant predictors, and combining adjacent predictors to maintain a monotonic decreasing relationship. Model performance during the simplification procedure was monitored using McFadden’s  $\rho^2$ .<sup>30</sup> Values between 0.2 and 0.4 indicate very good model fit and are analogous to an  $R^2$  value between 0.7 and 0.9 for linear regression.

In an effort to strengthen the generalizability of the regression analysis, we implemented validation by allocating participants to a training set and validation set in a 3:1 ratio.<sup>31</sup> Regression models were fit using only the training set. The performance of the simplified regression model was assessed by prediction accuracy for choice set selections by participants in the validation set using 1000 draws from the MIXL model. Prediction accuracy was quantified using the area under the curve (AUC) interpreted using the following thresholds: excellent, 0.9 to 1; good, 0.8 to 0.9; fair, 0.7 to 0.8; poor, 0.6 to 0.7; and failed, 0.5 to 0.6.<sup>32</sup>

Regression coefficients quantify the impact of dysfunction in a particular SOSGOQ-8D item on utility. Since level

1 for all SOSGOQ-8D items is nondysfunctional, this level imparts no change in utility. Under this scheme, utilities can be calculated by substituting the sum of the product of predictors and coefficients for each SOSGOQ-8D item in the formula

$$Utility = 1 - PF - L - A - Bow - Blad - P - MH - SF. \tag{1}$$

A worked example is provided in the Results section.

Since, the MIXL model treats each coefficient as a normal (“bell-curve”) random variable, regression results consisted of a mean and standard deviation for each coefficient. In this way MIXL techniques model heterogeneity (differences between individuals) of the utility impact of dysfunction in the SOSGOQ-8D items. Thus to predict how a single individual values the utility of each SOSGOQ-8D item, a random draw is made from the normal distributions estimated by the MIXL model. The mean coefficient values are the expected values for a single individual. In accordance with best practices in health economics, a SOSGOQ-8D utility scoring rubric was developed using mean values.<sup>6</sup> The importance of individual SOSGOQ-8D items was quantified by calculating the difference in utilities between the best and worst levels of the attribute.<sup>33</sup>

### Sample Size Calculation

Sample size was guided by S-efficiency which is a measure of the minimum sample size to estimate statistically significant regression parameters at the 95% level.<sup>34</sup> The minimum

sample for the DCE design shown in Supplemental Digital Content 6, Table is 2039 participants, <http://links.lww.com/BRS/B735>. Since we partitioned data in a 3:1 ratio, the sample size for the training set was 2039, and for the validation set was 691.

## RESULTS

Of the participants who began the DCE exercise, 95.3% completed the task. All geographic, sex, and age quotas based on the 2017 United States Census Bureau Population Estimates Program were met.<sup>18</sup> There were no statistically or qualitatively significant differences between the training set and validation set in terms of sex, age, or census region (Table 2). Most of the participants in the training and validation sets (69% and 70%, respectively) did not agree with the statement that “this survey was difficult.”

Five pairs of adjacent coefficients were collapsed to simplify the base regression model: L2 and L3, A1 and A2, MH1 and MH2, MH2 and MH3, and SF3 and SF4. Model simplification did not have an adverse effect on performance with the training set as McFadden’s  $\rho^2$  remained unchanged at 0.26 which is indicative of an excellent fit. The simplified regression model had excellent external validity as it predicted DCE choices in the validation set well with an AUC of 0.82 (95% confidence interval [CI]: 0.81–0.83).

The final MIXL regression results are shown in Supplemental Digital Content 7, Table, <http://links.lww.com/BRS/B736>. Statistically significant standard deviation for the majority of coefficients indicated the presence of heterogeneity between participants; therefore, use of a MIXL model was appropriate. Regression results revealed that participants did not regard all SOSGOQ-8D items equally important. The rank order of importance scores for the mean coefficient values (decreasing) was: Neurologic: Bowel/ Neurologic: Bladder, Pain, Physical Function/Neurologic: Legs, Social Function, Neurologic: Arms, and Mental Health.

To calculate utilities with Eq. (1), utility values for each item are obtained from the scoring rubric shown in Table 3. SOSGOQ-8D responses must first be converted to numerical levels using Table 1. To illustrate the use of Eq. (1) and the scoring rubric, we will calculate the utility for Patient A who provided the SOSGOQ-8D responses: PF5, L3, A2, Bow1, Blad1, P4, MH4, SF5. For the PF domain, Patient A provided a rating of 5, the corresponding number in Table 2 is 0.35. For the L domain, the patient provided a rating of 3, the corresponding number in Table 2 is 0.05. This process is repeated for each domain and the values substituted into the Equation 1

$$Utility = 1 - 0.35 - 0.05 - 0 - 0 - 0.15 - 0.04 - 0.26 = 0.15.$$

## DISCUSSION

In this article, we expanded the use of the SOSGOQ2.0 to QALY calculation. In Part 1 we formally shortened the SOSGOQ2.0 to an eight-item SOSGOQ-8D questionnaire.<sup>16</sup> This step was necessary to lessen the cognitive burden of the utility valuation exercise which was reported in Part 2.

We quantified a MAUF for the SOSGOQ-8D for the US general population using DCE methods. The MAUF demonstrated good out-of-sample prediction accuracy with an AUC of 0.82 (95% CI: 0.81–0.83). We provided a worked example for a hypothetical patient to illustrate how to calculate utilities using Eq. (1) and Table 3. This utility value quantifies the desirability of Patient A’s health state relative to perfect health (PF1, L1, A1, Bow1, Blad1, P1, MH1, SF1) from the general population perspective. A utility of 0.15 means that the general population would regard 365 days of life in Patient A’s health as equivalent to 365 days  $\times$  0.15 = 55 days of life in perfect health. In other words, if given the option between living 1 year in Patient A’s health state, or only living 55 days in perfect health, on average, members of the general population would choose

**TABLE 2. Respondent Demographic Characteristics**

	Training Set, N = 2039	Validation Set, N = 691
Sex, no. (%)		
Female	1034 (51)	366 (53)
Male	1005 (49)	325 (47)
Age, y, no. (%)		
18–29	434 (21)	153 (22)
30–39	358 (18)	112 (17)
40–49	325 (16)	114 (16)
50–59	356 (17)	115 (17)
60–69	299 (15)	98 (14)
≥70	267 (13)	99 (14)
Census region, no. (%)		
Northeast	364 (18)	120 (17)
Midwest	412 (20)	155 (22)
South	771 (38)	263 (38)
West	492 (24)	153 (23)



**TABLE 3. SOSGOQ-8D Utility Scoring Rubric**

Level	SOSGOQ-8D Item							
	PF	L	A	Bow	Blad	P	MH	SF
1	0	0	0	0	0	0	0	0
2	0.09	0.05	0	0.06	0.05	0.05	0	0.06
3	0.15	0.05	0.09	0.16	0.12	0.13	0.04	0.06
4	0.28	0.13	0.09	0.32	0.24	0.15	0.04	0.18
5	0.35	0.31	0.24	0.39	0.38	0.36	0.10	0.26

*To use this table, SOSGOQ-8D responses must be converted to numerical levels using Table 1. The appropriate values from this table are then substituted in Eq. (1) to calculate utilities. A indicates Arms; Blad, bladder; Bow, bowel; L, Legs; MH, Mental Health; P, Pain; PF, Physical Function; SF, Social Function.*

to live a shorter duration with better health. The SOSGOQ-8D MAUF also provides useful insights for clinicians.

The MAUF can also be used to quantify the importance of each SOSGOQ-8D item. Importance scores are listed in Table 3 and quantify the how much individuals discount life in the worst level of each SOSGOQ-8D item relative to the best. For example, an importance score 0.33 for both sphincter function domains means that individuals would be willing to trade  $365 \text{ days} \times 0.33 = 120$  days of life to reverse sphincter dysfunction from its worst state. In contrast, individuals are only willing to trade  $365 \text{ days} \times 0.15 = 55$  days of life to reverse arm dysfunction from its worst state. Notably, lower extremity neurologic function, which is a common outcome for metastatic spine disease studies, was less important than sphincter function and pain. It was equally important to physical function which was represented by the ability to care for self. Clinicians should heed these findings and offer treatments that maximize function in the most important attributes.

In Part 1 we formally shortened the SOSGOQ2.0 to an eight-item SOSGOQ-8D questionnaire using the “data-splitting” technique for internal validation.<sup>35,36</sup> Based on *a priori* selection criteria, SOSGOQ2.0 items 3, 11, 17, and 19 were selected because these items performed best on the validation set. Other decision rules for internal validation yield different sets of items. For example, items 4, 13, 17 and 18 were the best performing on the training set and were the second-best on the validation set. An independent test set, in addition to the training and validation sets used, would be required to evaluate different decision rules. Since the SOSGOQ-8D identified in this study have clinical face validity, we feel it is appropriate to use them in future work without further validation with an additional test set.

It is important for readers to appreciate that utilities provided by the general population (*ex ante*) are not equivalent to utilities obtained from patients who have experienced the health states of interest (*ex post*).<sup>37</sup> Although *ex ante* utilities are theoretically restricted to system policy decisions, *ex ante* utilities have become the de facto standard for individual patient decision making. This is because the widely used utilities obtained from generic health surveys such as the EuroQol-5D, Short Form-6D, and Health Utilities Index 3 are actually *ex ante* valuations.<sup>6</sup> Therefore generating *ex ante* utilities for the SOSGOQ-8D conforms with conventions in the literature.

We expanded the use of the SOSGOQ2.0 to QALY calculation. Eq. (1) and Table 3 can be used together to convert SOSGOQ-8D responses to utilities. The regression modeling exercise revealed the relative importance of SOSGOQ-8D items to the general population. These data can be used to help inform population level health care decision-making, such as the allocation of limited resources for specific treatments. The results of this study can also help clinicians counsel patients.

Supplemental Digital Content 2, <http://links.lww.com/BRS/B731>.

Supplemental Digital Content 3, <http://links.lww.com/BRS/B732>.

Supplemental Digital Content 4, <http://links.lww.com/BRS/B733>.

Supplemental Digital Content 5, <http://links.lww.com/BRS/B734>.

### ➤ Key Points

- Consideration of QALYs can enhance resource allocation and patient counseling.
- We developed a simple technique for converting the SOSGOQ to QALYs weights (utilities)
- Respondents did regard all SOSGOQ domains equally important; therefore, clinicians should heed these findings and offer treatments that maximize function in the most important domains.

### Acknowledgments

The authors thank Dr. Nick Bansback for his helpful insights and suggestions on methodology for the MAUF derivation. The authors and AO Spine also thank the following principal investigators and study sites who contributed to the AO Spine EPOS0 and MTRON studies.

- Daniel Sciubba, Johns Hopkins University School of Medicine
- Jorrit Jan Verlaan, Universitair Medisch Centrum Utrecht
- Michael Johnson, Winnipeg Health Science Centre
- Addisu Mesfin, University of Rochester

- John O'Toole, Rush University Medical Center
- John Shin, Massachusetts General Hospital
- Michael Weber, McGill University Health Centre
- Michael Fehlings, Toronto Western Hospital
- James Schuster, Hospital of the University of Pennsylvania
- Dean Chou, University of California San Francisco
- Tony Goldschlager, Monash Health
- Naresh Kumar, National University Hospital of Singapore
- Michelle Clarke, Mayo Clinic
- Paul Arnold, University of Kansas Medical Center Research Institute

Supplemental digital content is available for this article. Direct URL citations appearing in the printed text are provided in the HTML and PDF version of this article on the journal's Web site ([www.spinejournal.com](http://www.spinejournal.com)).

## References

1. Paulino Pereira NR, Janssen SJ, Raskin KA, et al. Most efficient questionnaires to measure quality of life, physical function, and pain in patients with metastatic spine disease: a cross-sectional prospective survey study. *Spine J* 2017;17:953–61.
2. Ryu S, Rock J, Jain R, et al. Radiosurgical decompression of metastatic epidural compression. *Cancer* 2010;116:2250–7.
3. Lee I, Omodon M, Rock J, et al. Stereotactic radiosurgery for high-grade metastatic epidural cord compression. *J radiosurgery SBRT* 2014;3:51–8.
4. Thomas KC, Nosyk B, Fisher CG, et al. Cost-effectiveness of surgery plus radiotherapy versus radiotherapy alone for metastatic epidural spinal cord compression. *Int J Radiat Oncol Biol Phys* 2006;66:1212–8.
5. Furlan JC, Chan KK-WKW, Sandoval GA, et al. The combined use of surgery and radiotherapy to treat patients with epidural cord compression due to metastatic disease: a cost-utility analysis. *Neuro Oncol* 2012;14:631–40.
6. Drummond MF, Sculpher MJ, Torrance GW, et al. *Methods for the Economic Evaluation of Health Care Programmes*. USA: Oxford University Press; 2005.
7. Kind P, Lafata JE, Matuszewski K, et al. The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health* 2009;12 suppl 1:S27–30.
8. Teckle P, Peacock S, McTaggart-Cowan H, et al. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health Qual Life Outcomes* 2011;9:106.
9. Lorgelly PK, Doble B, Rowen D, et al. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. *Qual Life Res* 2017;26:1163–76.
10. Hill MR, Noonan VK, Sakakibara BM, et al., SCIRE Research Team. Quality of life instruments and definitions in individuals with spinal cord injury: a systematic review. *Spinal Cord* 2010;48:438–50.
11. Whitehurst DGT, Noonan VK, Dvorak MFS, et al. A review of preference-based health-related quality of life questionnaires in spinal cord injury research. *Spinal Cord* 2012;50:646–54.
12. Bansback N, Ara R, Karnon J, et al. Economic evaluations in rheumatoid arthritis: a critical review of measures used to define health States. *Pharmacoeconomics* 2008;26:395–408.
13. DeVine J, Norvell DC, Ecker E, et al. Evaluating the correlation and responsiveness of patient-reported pain with function and quality-of-life outcomes after spine surgery. *Spine (Phila Pa 1976)* 2011;36 (21 suppl):S69–74.
14. Street J, Lenehan B, Berven S, et al. Introducing a new health-related quality of life outcome tool for metastatic disease of the spine: content validation using the International Classification of Functioning, Disability, and Health; on behalf of the Spine Oncology Study Group. *Spine (Phila Pa 1976)* 2010;35:1377–86.
15. Beaton DE, Wright JG, Katz JN, et al. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Jt Surg Am* 2005;87:1038–46.
16. Lemetayer F, Montel S, Recchia S, et al. Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *J Clin Epidemiol* 2013;66:710–8.
17. Pahuta MA, Wai EK, Werier J, et al. A general population utility valuation study for metastatic epidural spinal cord compression health states. *Spine (Phila Pa 1976)* 2019;44:943–50.
18. United States Census Bureau. State Population by Characteristics: 2010-2018. Population Estimates Program. Available at: <https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html>. Published 2018.
19. Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. *J Health Econ* 2012;31:306–18.
20. Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. *Qual Life Res* 2016;25:637–49.
21. Pahuta M, Formbach A, Mitera G, et al. Validation of the self-administered online assessment of preferences (SOAP) utility elicitation tool. *Can J Surg* 2016;59 (3 suppl 2):S40.
22. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Heal* 2013;16:3–13.
23. Janssen EM, Marshall DA, Hauber AB, et al. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability?. *Expert Rev Pharmacoecon Outcomes Res* 2017;17:531–42.
24. ChoiceMetrics. Ngene 1.1.1 User Manual & Reference Guide. 2018.
25. Harrison M, Marra C, Shojanian K, et al. Societal preferences for rheumatoid arthritis treatments: evidence from a discrete choice experiment. *Rheumatol (United Kingdom)* 2015;54:1816–25.
26. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics* 2019;37:201–26.
27. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Heal* 2016;19:300–15.
28. R Core Team. R: A Language and Environment for Statistical Computing. 2018.
29. Molloy J, Schmid B, Becker F. Mixl: an open-source R Package for estimating complex choice models on large datasets. *Zurich* 2019; doi:10.3929/ethz-b-000334289.
30. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications*, 1st ed. Cambridge, UK: Cambridge University Press; 2000.
31. Steyerberg EW, Harrell FE, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
32. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law Hum Behav* 2005;29:615–20.
33. Kromer C, Schaarschmidt M-L, Schmieder A, et al. Patient preferences for treatment of psoriasis with biologicals: a discrete choice experiment. *PLoS One* 2015;10:e0129120.
34. Rose JM, Bliemer MCJ. Sample size requirements for stated choice experiments. *Transportation (Amst)* 2013;40:1021–41.
35. Steyerberg EW, Harrell FE, Borsboom GJJM, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
36. Picard RR, Berk KN. Data splitting. *Am Stat* 1990;44:140.
37. Nord E, Daniels N, Kamlet M. QALYs: some challenges. *Value Heal* 2009;12 (suppl 1):S10–5.