


Psychometric Evaluation and Adaptation of the Spine Oncology Study Group Outcomes Questionnaire to Evaluate Health-Related Quality of Life in Patients With Spinal Metastases

Anne L. Versteeg, MD ¹; Arjun Sahgal, MD²; Laurence D. Rhines, MD³; Daniel M. Sciubba, MD⁴; James M. Schuster, MD, PhD⁵; Michael H. Weber, MD, PhD⁶; Peter Pal Varga, MD⁷; Stefano Boriani, MD⁸; Chetan Bettegowda, MD, PhD⁴; Michael G. Fehlings, MD, PhD⁹; Michelle J. Clarke, MD¹⁰; Paul M. Arnold, MD¹¹; Ziya L. Gokaslan, MD¹²; Charles G. Fisher, MD, MHS¹³; and AOSpine Knowledge Forum Tumor¹⁴

BACKGROUND: The Spine Oncology Study Group Outcomes Questionnaire (SOSGOQ) was developed as the first spine oncology-specific health-related quality of life (HRQOL) measure. This study evaluated the psychometric properties and clinical validity of the SOSGOQ in a diverse cohort of patients with spinal metastases. **METHODS:** An international, multicenter, prospective observational cohort study including patients with spinal metastases who underwent surgery and/or radiotherapy was conducted by the AOSpine Knowledge Forum Tumor. Demographic, tumor, and treatment data were collected. HRQOL was evaluated using the SOSGOQ and Medical Outcomes Study Questionnaire Short Form 36 Health Survey (SF-36) at baseline and fixed follow-up times. Construct validity was assessed using multitrait scaling analyses, confirmatory factor analyses, and correlation with the SF-36 and NRS pain score. Test-retest reliability was assessed in a subgroup of patients between 12 weeks after treatment and the retest 4 to 9 days later. **RESULTS:** A total of 238 patients were enrolled at 9 centers across North America; 153 of these patients had HRQOL data available at 12 weeks after treatment. Multitrait scaling analyses and confirmatory factor analyses resulted in a refined version of the SOSGOQ with 4 domains and 4 single items. The revised SOSGOQ (SOSGOQ2.0) demonstrated strong correlations with SF-36 and the ability to discriminate between clinically distinct patient groups. Reliability of the SOSGOQ2.0 was demonstrated to be good, with an intraclass correlation coefficient ranging from 0.58 to 0.92 for the different domains. **CONCLUSIONS:** The SOSGOQ2.0 is a reliable and valid measure with which to evaluate HRQOL in patients with spinal metastases. It is recommended to use the SOSGOQ2.0 together with a generic HRQOL outcome measure to comprehensively assess HRQOL and increase sensitivity and specificity. *Cancer* 2018;124:1828-38. © 2018 The Authors. *Cancer* published by Wiley Periodicals, Inc. on behalf of *American Cancer Society*. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

KEYWORDS: metastases, patient-reported outcomes, quality of life, reliability, spine, validity.

INTRODUCTION

A diagnosis of bone metastases often represents an incurable yet treatable disease. The population of patients with advanced stages of cancer, including metastatic spinal disease, is growing due to improved medical treatment options, which allow for better and longer disease control.¹ The main treatment goal for these patients is to improve or maintain

Corresponding author: Charles Fisher, MD, FRCSC, Blusson Spinal Cord Centre, Division of Spine, Department of Orthopaedics, Vancouver General Hospital, University of British Columbia, 6th Fl, 818 W10th Ave, Vancouver, BC Canada, V5Z 1M9; charles.fisher@vch.ca

¹Department of Orthopedics, University Medical Center Utrecht, Utrecht, The Netherlands; ²Department of Radiation Oncology, Sunnybrook Odette Cancer Centre, University of Toronto, Toronto, Ontario, Canada; ³Division of Surgery, Department of Neurosurgery, The University of Texas MD Anderson Cancer Center, Houston, Texas; ⁴Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland; ⁵Department of Neurosurgery, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania; ⁶Division of Surgery, Montreal General Hospital, McGill University, Montreal, Quebec, Canada; ⁷National Center for Spinal Disorders and Buda Health Center, Budapest, Hungary; ⁸Spine4 Spine Surgery Division, IRCCS Galeazzi Orthopedic Institute, Milan, Italy; ⁹Division of Orthopedic Surgery, Toronto Western Hospital, University of Toronto, Toronto, Ontario, Canada; ¹⁰Department of Neurosurgery, Mayo Clinic, Rochester, Minnesota; ¹¹Department of Neurosurgery, The University of Kansas Hospital, Kansas City, Kansas; ¹²Department of Neurosurgery, The Warren Alpert Medical School of Brown University and Rhode Island Hospital, The Miriam Hospital, Providence, Rhode Island; ¹³Division of Spine, Department of Orthopaedics, Vancouver General Hospital, University of British Columbia, Vancouver, British Columbia, Canada; ¹⁴Research Department, AOSpine International, Davos, Switzerland

We are grateful to the local clinical research personnel and support staff of the collaborating centers for their active participation. We acknowledge the study support that was provided by AOSpine's Research department and Clinical Investigation and Documentation unit. We especially thank Christian Knoll (AOCID statistician) and Louise Masse (professor at the University of British Columbia and psychometric measurement specialist) for their statistical and psychometric measurement support.

Additional supporting information may be found in the online version of this article.

DOI: 10.1002/cncr.31240, **Received:** August 30, 2017; **Revised:** November 6, 2017; **Accepted:** December 10, 2017, **Published online** February 6, 2018 in Wiley Online Library (wileyonlinelibrary.com)

health-related quality of life (HRQOL) for their remaining lifetime. Therefore, determining patient-reported HRQOL is important to optimize and evaluate life-extending and supportive care treatments.

Many generic and cancer-specific outcome measures, including those for bone metastases, exist.² Nevertheless, these outcome measures are nonspecific for spine-related functions, and thus are less sensitive for assessing changes over time due to treatment or the progression of spinal disease. A disease-specific instrument in addition to a generic (cancer) HRQOL measure will increase the sensitivity and specificity of HRQOL assessments. In response to the absence of spine oncology-specific outcome measures, the Spine Oncology Study Group Outcomes Questionnaire (SOSGOQ) was developed and assessed for face and content validity.³ Content validity was evaluated by correlating the SOSGOQ items to the World Health Organization International Classification of Functioning, Disability and Health and face validity by content expert evaluation, both of which demonstrated excellent results.³

With content validity confirmed, the next step is to evaluate the hypothesized structure of the SOSGOQ in a clinical setting. Therefore, the objective of the current study was to examine the psychometric properties and clinical validity of the SOSGOQ in a cohort of patients who had undergone treatment for spinal metastases.

MATERIALS AND METHODS

Design

The AOSpine Knowledge Forum Tumor initiated an international, multicenter, prospective observational cohort study in August 2013 at 10 spine centers across North America and Europe (ClinicalTrials.gov identifier NCT01825161). The prespecified study objectives included evaluation of the validity and reliability of the English version of the SOSGOQ. Patients were eligible for inclusion if they had a diagnosis of metastatic spinal disease, were aged 18 to 75 years, and had undergone surgery and/or radiotherapy for the treatment of spinal metastases from any primary tumor. Patients with a central nervous system tumor or a primary spinal bone tumor were excluded. The ethics board of each participating spine center approved the protocol. All patients provided written informed consent for study participation.

Demographic, medical history, diagnostic, treatment, adverse events, and HRQOL data were collected prospectively. HRQOL and pain scores were evaluated at baseline and at 6, 12, 24, 52, and 104 weeks after treatment or until

death with the English version of the SOSGOQ (version 1.0³), the English version of the Medical Outcomes Study Questionnaire Short Form 36 Health Survey (SF-36, version 2.0; Medical Outcomes Trust, Boston, Massachusetts⁴), and the numeric rating scale (NRS) pain score. All data were stored in a secure Web-based application (Research Electronic Data Capture [REDCap]; Vanderbilt University, Nashville, Tennessee).

The SF-36⁵ and the NRS pain score are widely used generic outcome instruments and have been described in detail previously. The hypothesized structure of the SOSGOQ consists of 5 HRQOL domains (Physical Function, Neurological Function, Pain, Mental Health, and Social Function) and an additional set of questions during follow-up for evaluating treatment satisfaction. Scores for the SF-36 and the SOSGOQ were calculated if at least 50% of the items within the domain were answered. All domain scores were transformed to a scale of 0 to 100, in which a higher score represents a better QOL.

Statistical Analysis

Demographic data were summarized using descriptive statistics. A subject-to-item ratio of 1:7⁶ was considered to be sufficient power with which to perform a factor analysis and assess validity; the minimum required sample size was 140 subjects with 20 items in the SOSGOQ to be evaluated (excluding the posttherapy domain).

The results first were analyzed to define the structure (construct validity) of the SOSGOQ. The concurrent validity, clinical validity, and reliability were determined using the modified structure of the SOSGOQ (SOSGOQ2.0). All statistical analyses were performed using SAS statistical software (version 9.4; SAS Institute Inc, Cary, North Carolina). Statistical significance was defined as $P < .05$.

Construct Validity

The internal structure of the SOSGOQ was evaluated using multitrait scaling analysis and confirmatory factor analyses (CFA). Convergent validity was evaluated by correlating the item score with the total score of its own domain, and divergent validity was evaluated by correlating the item score with the total score of the other SOSGOQ domains (Spearman rank). A scaling success was defined as a significantly higher item correlation with its own domain compared with the other SOSGOQ domains, with a minimum correlation of 0.40.⁷

CFA first was performed based on the prior hypothesized structure of the SOSGOQ³ followed by testing of a modified structure on a clinical conceptual basis supported by the results of the multitrait scaling analysis and

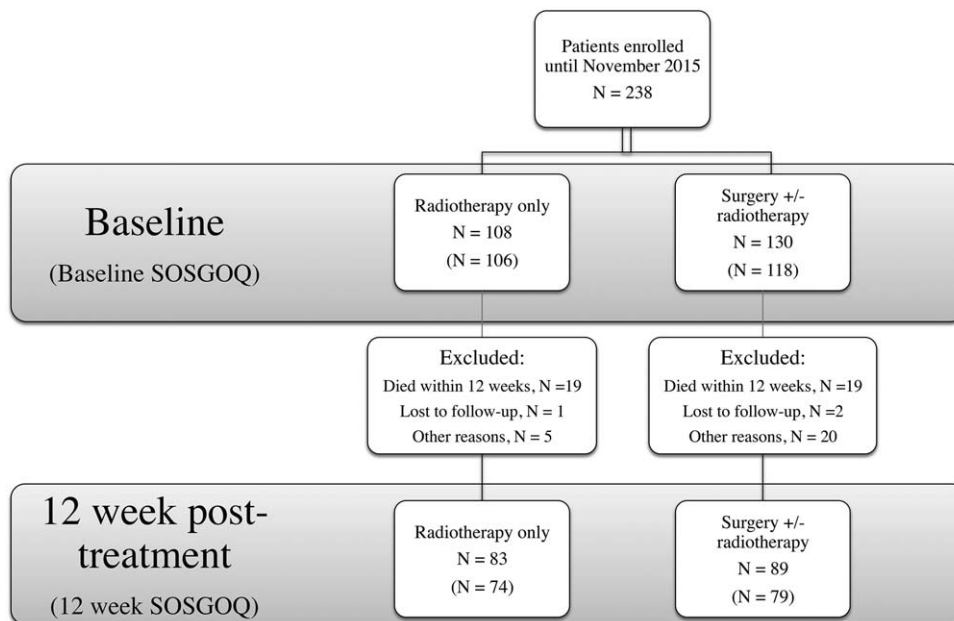


Figure 1. Patient selection flow diagram. SOSGOQ indicates Spine Oncology Study Group Outcomes Questionnaire.

modification indices. Model fit was evaluated with the root mean square error of approximation (RMSEA [≤ 0.08]),⁸ the 90% confidence interval (90% CI) of the RMSEA (upper bound limit of 0.1), the comparative fit index (CFI; ≥ 0.90),⁹ and the standardized root mean residual (SRMR; ≤ 0.08).¹⁰

Concurrent validity was evaluated by Spearman rank correlation of the SOSGOQ2.0 domains to the domains of the SF-36 and the NRS pain score. The domains that are conceptually related were expected to demonstrate a correlation of at least 0.40. Analyses were performed with the 12-week posttreatment data to meet the normality assumption; baseline data were used to perform sensitivity analyses of the 12-week data.

Clinical Validity

Clinical validity was examined using the ability of the SOSGOQ2.0 to discriminate between patient groups. At baseline, patients with an Eastern Cooperative Oncology Group (ECOG) performance score of 0 or 1 were compared with patients with an ECOG score of ≥ 2 . Furthermore, to assess the responsiveness to change, changes in ECOG status from baseline to 12 weeks after treatment (stable/improved vs deteriorated) were associated with changes in the SOSGOQ2.0 scores.

Reliability and Reproducibility

The test-retest reliability of the SOSGOQ was assessed at 2 centers between the 12-week posttreatment assessment

and the retest 4 to 9 days later using the intraclass correlation coefficient (ICC).¹¹

Internal consistency of the domains of the SOSGOQ2.0 was evaluated using Cronbach alpha¹² at baseline and follow-up; the minimum acceptable alpha value was defined at .70.¹³

RESULTS

A total of 238 patients from 9 centers across North America (4 in Canada and 5 in the United States) were enrolled in the prospective observational cohort study until November 2015. Of the 238 patients, 130 underwent surgery with or without additional radiotherapy and 108 patients received only radiotherapy (Fig. 1). The breast (26%) was the most common primary tumor site, followed by the lung (18%) and renal cell (16%). A summary of the baseline characteristics of the patients is shown in Table 1. At 12 weeks, a total of 172 patients had data available, 38 (16%) patients had died within the first 12 weeks of follow-up, 3 patients were lost to follow-up, 11 patients did not complete the 12-week follow-up visit, and 14 patients (6%) dropped out of the study for other reasons (withdrawal of consent, withdrawn by investigator). The SOSGOQ was completed at baseline by 224 patients (94%) and by 153 patients (63%) at 12 weeks after treatment.

At baseline, 44 items from the SOSGOQ (1%) and 47 items from the SF-36 (0.6%) were missing. The item addressing bowel and bladder function (item 8) in the

TABLE 1. Summary of Baseline Characteristics

Characteristic	No. (%)
Age at treatment, (n = 238), y ^a	59 (10.3)
Sex (n = 238)	
Female	129 (54.2)
Male	109 (45.8)
Primary tumor (n = 238)	
Breast	62 (26.1)
Lung	44 (18.5)
Kidney	39 (16.4)
Prostate	21 (8.8)
Other	72 (30.3)
Time since primary diagnosis (n = 238), mo ^b	26.5 (5.0-76.0)
ECOG PS (n = 234)	
0-1	156 (66.7)
2-4	78 (33.3)
Location (n = 238) ^c	
Cervical	38 (16)
Thoracic	152 (64)
Lumbar/sacral	103 (43)
Presence of other metastases (n = 238)	
Visceral	80 (33.6)
Brain	23 (9.7)
ASIA	
ASIA E	181 (77.4)
ASIA D	41 (17.5)
ASIA A-C	12 (5.1)

Abbreviations: ASIA, American Spinal Injury Association; ASIA A, complete; ASIA B, incomplete (sensory but not motor function is preserved below the neurological level and includes the sacral segments S4-S5); ASIA C, incomplete (motor function is preserved below the neurological level, and more than one-half of key muscles below the neurological level have a muscle grade <3); ASIA D, incomplete (motor function is preserved below the neurological level, and at least one-half of key muscles below the neurological level have a muscle grade of ≥3); ASIA E, normal (motor and sensory function are normal); ECOG PS, Eastern Cooperative Oncology Group performance status.

^a Mean ± standard deviation.

^b Median (interquartile range).

^c Multiple levels possible.

SOSGOQ was the most commonly reported missing item across different time points. The items addressing strength in the arms (item 6) and bowel and bladder function (item 8) demonstrated positive skewness, with only 7% and 3%, respectively, of the patients reporting moderate to severe symptoms.

Internal Structure of the SOSGOQ

A total of 133 patients from 9 centers had complete data available with which to evaluate the structure of the SOSGOQ. The multitrait scaling analysis at 12 weeks after treatment according to the prior hypothesized structure of the SOSGOQ³ demonstrated that item 7 (need for a walking aid) and item 20 (leaving the house) had a strong correlation (correlation coefficient [R] = -0.71 and R = 0.65, respectively) with the Physical Function domain. Furthermore, item 16 (overwhelming pain) also demonstrated a strong correlation with the Pain domain (R = -0.60). Item 6 (arm strength), item 8 (bowel and

bladder function), and item 15 (level of energy) demonstrated only a moderate correlation with their own scale (R = -0.53 to 0.59). The correlation of the other items exceeded the minimum correlation of 0.40 and these items were found to have a stronger correlation with their own domain compared with the other domains (see Supporting Information).

The CFA according to the hypothesized scale structure of the SOSGOQ confirmed the results of the multitrait scaling analysis, demonstrating unacceptable model fit (RMSEA, 0.12 [90% CI, 0.10-0.13]; CFI, 0.79; and SRMR, 0.11). Based on the conceptual relation of items with other domains and the multitrait scaling analysis, the scale structure was modified. Items 7 (need for a walking aid) and 20 (leaving the house) were moved to the Physical Function domain and item 16 was moved to the Pain domain (cross-loading on mental health). Items 5 (leg strength), 6 (arm strength), 8 (bowel and bladder function) retained as single items, yet item 8 was split into 2 questions to distinctly address bowel and bladder function. Item 15 (level of energy) was not found to contribute any new information and was removed. CFA according to the revised SOSGOQ structure confirmed the structure, demonstrating adequate model fit (RMSEA, 0.074 [90% CI, 0.055-0.092]; CFI, 0.928; and SRMR, 0.06). The revised questionnaire consists of 4 domains and 4 single items and, at follow-up, a set of questions evaluating treatment satisfaction (Fig. 2). Scoring guidelines are outlined in the Supporting Information. Results of the multitrait scaling analysis according to the final structure are shown in Table 2. Sensitivity analysis with the baseline data confirmed the revised structure of the SOSGOQ, demonstrating an adequate model fit (RMSEA, 0.08 [90% CI, 0.068-0.095]; CFI, 0.94; and SRMR, 0.054).

Concurrent Validity With the SF-36 and NRS Pain Score

Correlation between the domains of the SOSGOQ2.0 and the SF-36 and NRS pain score are shown in Table 3. The domains of the SOSGOQ2.0 demonstrated a strong to very strong correlation with the corresponding domains of the SF-36. The SOSGOQ2.0 Pain domain demonstrated a strong correlation with the NRS pain score. The single neurological function items demonstrated a weak correlation with the domains of the SF-36, indicating that these items assess a different aspect that is not assessed by the SF-36.

Clinical Validity

At baseline, patients with an ECOG performance score of 0 or 1 demonstrated a significantly higher score ($P < .001$)

Please think about your level of functioning and symptoms over the past 4 weeks while filling out this questionnaire. © 2018 AOSpine International

I. Physical function

1. What is your level of activity?
 - Full activities without restriction
 - Moderate activities out of house
 - Mobility limited to within house
 - Bed to chair activity
 - Bedridden
2. What is your ability to work (including at home)/study?
 - Unlimited
 - 4-8 hours per day
 - 2-4 hours per day
 - Less than 2 hours per day
 - Not at all
3. Does your spine limit your ability to care for yourself?
 - Not at all
 - A little bit
 - Somewhat
 - Quite a bit
 - Very much
4. Do you require assistance from others to travel outside the home?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Very often
5. What assistance do you need with your walking?
 - None
 - A cane
 - A walker/2 canes
 - Assistance from others
 - Cannot walk at all
6. Do you leave the house for social functions?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Very often

II. A: Neurological function legs

7. Do you have weakness in your legs?
 - None
 - Mild occasionally
 - Mild constantly
 - Moderate constantly
 - Severe constantly

II. B: Neurological function arms

8. Do you have weakness in your arms?
 - None
 - Mild occasionally
 - Mild constantly
 - Moderate constantly
 - Severe constantly

II. C: Neurological function bowel

9. Do you have difficulty controlling your bowel function beyond episodes of diarrhea/constipation?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Very often

II. D: Neurological function bladder

10. Do you have difficulty controlling your bladder function?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Requires catheterization

III. Pain

11. Overall, on average, how much back/neck pain do you have?
 - None
 - Very mild
 - Mild
 - Moderate
 - Severe
12. When you are in your most comfortable position, do you still experience back/neck pain (limiting your sleep)?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Very often
13. How much has your pain limited your mobility (sitting, standing, walking)?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Constantly

Figure 2. Revised structure of the Spine Oncology Study Group Outcomes Questionnaire (SOSGOQ). (From AOSpine International, Used under CC BY-NC-ND 4.0).

across the different SOSGOQ2.0 domains compared with patients with an ECOG performance score of ≥ 2 . Patients with a stable or improved ECOG performance

score at 12 weeks after treatment demonstrated an increase in domain scores compared with a deterioration in the SOSGOQ2.0 domain scores noted for patients

14. How confident do you feel in your ability to manage your pain on your own?

- Not confident at all
- Minimally confident
- Moderately confident
- Mostly confident
- Completely confident

15. When I feel pain, it is awful and I feel that it overwhelms me.

- Never
- Rarely
- Sometimes
- Often
- Very often

IV. Mental health

16. Have you felt depressed?

- Never
- Rarely
- Sometimes
- Often
- Very often

17. Do you feel anxiety about your health related to your spine?

- Never
- Rarely
- Sometimes
- Often
- Very often

V. Social function

18. Does your spine influence your ability to concentrate on conversations, reading, and television?

- Never
- Rarely
- Sometimes
- Often
- Very often

19. Do you feel that your spine condition affects your personal relationships?

- Never
- Rarely
- Sometimes
- Often
- Very often

20. Are you comfortable meeting new people?

- Never
- Rarely
- Sometimes
- Often
- Very often

Post-therapy questions

21. Are you satisfied with the results of your spine tumor management?

- Very satisfied
- Somewhat satisfied
- Neither satisfied nor dissatisfied
- Somewhat dissatisfied
- Very dissatisfied

22. Would you choose the same management of your spine tumor again?

- Definitely yes
- Probably yes
- Not sure
- Probably not
- Definitely not

23. How has treatment of your spine changed your physical function and ability to pursue activities of daily living?

- Much better
- Somewhat better
- No change
- Somewhat worse
- Much worse

24. How has treatment of your spine affected your spinal cord and/or nerve function?

- Much better
- Somewhat better
- No change
- Somewhat worse
- Much worse

25. How has your treatment affected your overall pain from your spine?

- Much better
- Somewhat better
- No change
- Somewhat worse
- Much worse

26. How has treatment of your spine changed your depression and anxiety?

- Much better
- Somewhat better
- No change
- Somewhat worse
- Much worse

27. How has treatment of your spine changed your ability to function socially?

- Much better
- Somewhat better
- No change
- Somewhat worse
- Much worse

Figure 2. (Continued).

with a decline in the ECOG performance score. This difference was significant for all domains except the Mental Health domain (Table 4).

Reliability

A total of 36 patients from 2 centers also completed the retest of the SOSGOQ within 4 to 9 days after the

TABLE 2. Convergent and Divergent Validity at Baseline and at 12 Weeks^a

Scales	Surgery ± RT		RT Only		All Patients	
	Item Own Domain Correlation	Item Other Domain Correlation	Item Own Domain Correlation	Item Other Domain Correlation	Item Own Domain Correlation	Item Other Domain Correlation
Physical Function	0.68-0.84 <i>0.55-0.88</i>	0.14-0.50 <i>0.09-0.55</i>	0.70-0.81 <i>0.71-0.88</i>	0.01-0.68 <i>0.01-0.66</i>	0.72-0.84 <i>0.70-0.91</i>	0.02-0.55 <i>0.05-0.60</i>
Pain	0.60-0.82 <i>0.67-0.78</i>	0.13-0.59 <i>0.17-0.60</i>	0.67-0.91 <i>0.62-0.88</i>	0.02-0.65 <i>0.05-0.68</i>	0.64-0.84 <i>0.69-0.84</i>	0.18-0.61 <i>0.17-0.72</i>
Mental Health	0.89-0.92 <i>0.84-0.89</i>	0.20-0.48 <i>0.19-0.50</i>	0.89-0.92 <i>0.89-0.91</i>	0.01-0.41 <i>0.09-0.28</i>	0.90-0.91 <i>0.86-0.90</i>	0.10-0.37 <i>0.12-0.32</i>
Social Function	0.71-0.85 <i>0.67-0.87</i>	0.25-0.65 <i>0.09-0.68</i>	0.67-0.84 <i>0.67-0.87</i>	0.19-0.63 <i>0.21-0.73</i>	0.68-0.83 <i>0.66-0.88</i>	0.24-0.64 <i>0.16-0.73</i>
Neuro Function: legs	1.00	0.15-0.41 <i>0.26-0.52</i>	1.00	0.01-0.47 <i>0.14- 0.58</i>	1.00	0.05-0.43 <i>0.17-0.58</i>
Neuro function arms	1.00	0.14-0.26 <i>0.23-0.35</i>	1.00	0.07-0.38 <i>0.10-0.44</i>	1.00	0.12-0.25 <i>0.17-0.38</i>
Neuro Function: bowel and bladder	1.00	0.13-0.46 <i>0.12-0.30</i>	1.00	0.06-0.29 <i>0.08-0.30</i>	1.00	0.11-0.36 <i>0.02-0.32</i>

Abbreviations: Neuro, neurological; RT, radiotherapy.

Range of Spearman correlation. Number of patients at baseline was 238; number at 12 weeks after treatment was 153.

^a Italics represent correlations at baseline.

assessment at 12 weeks after treatment. Of these 36 patients, 10 underwent surgical intervention with or without additional postoperative radiotherapy and the remaining 26 patients received radiotherapy only. There were no significant differences noted between patients who completed the retest of the SOSGOQ compared with those included in the validity analyses with regard to age ($P = .043$), sex ($P = .873$), location of the primary tumor ($P = .503$), and ECOG performance score ($P = .191$). The ICC of the SOSGOQ2.0 domains ranged from 0.58 to 0.92, with the Mental Health and posttherapy domains demonstrating the lowest reproducibility with an ICC of 0.62 and 0.58, respectively (Table 5). The Cronbach alphas at baseline and follow-up were acceptable for all domains (Table 5).

DISCUSSION

Patient-reported outcomes are essential in the evaluation of long-term palliative cancer care. The value of a disease-specific questionnaire with the use of generic questionnaires has been widely acknowledged, including the development of many cancer-specific modules by the European Organization for Research and Treatment of Cancer.¹⁴ To the best of our knowledge, the SOSGOQ is the first spine oncology-specific HRQOL questionnaire. The objective of the current study was to assess the psychometric properties and clinical validity of the SOSGOQ in a diverse cohort of patients with spinal metastases. Psychometric evaluation of the SOSGOQ resulted in a slightly revised structure. Evaluation of the revised SOSGOQ

demonstrated that it is a clinically valid and reliable questionnaire with which to evaluate QOL in patients with spinal metastases. It is recommended to use the SOSGOQ together with a generic outcome measure to comprehensively evaluate HRQOL in patients with spinal metastases.

Recently, Janssen et al investigated the construct validity of the SOSGOQ by its correlation to the 5-level EuroQol 5 dimensions (EQ-5D-5L) and exploratory factor analyses in a convenience sample of 82 patients with spinal metastases.¹⁵ Exploratory factor analyses demonstrated the association between items 7 (need for a walking aid) and 20 (leaving the house) and the Physical Function domain and item 16 (intensity of pain) and the Pain domain. Furthermore, the neurological function items and item 15 (level of energy) demonstrated low factor loadings.¹⁵ These results are in agreement with the results of our CFA analyses and the revised SOSGOQ2.0. In the current study, concurrent validity was evaluated by correlating the SOSGOQ2.0 with the SF-36 and the NRS pain score. The SF-36 is a generic yet comprehensive HRQOL measure because it consists of several domains, including multiple items per domain, compared with the less comprehensive EQ-5D-5L, which uses only single items representing the different domains. Correlation of the SOSGOQ2.0 with the SF-36 and NRS pain score demonstrated strong correlations between the domains that were conceptually related, confirming the construct validity of the SOSGOQ2.0.

In contrast to the study by Janssen et al,¹⁵ the current study also assessed the test-retest reliability. The

TABLE 3. Concurrent Validity (Revised SOSGOQ Version 2.0)

Reference Score	SOSGOQ2.0 Physical Function		SOSGOQ2.0 Neurological Item 5		SOSGOQ2.0 Neurological Item 6		SOSGOQ2.0 Neurological Item 8		SOSGOQ2.0 Pain		SOSGOQ2.0 Mental Health		SOSGOQ2.0 Social Function	
	No.	r_s (P)	No.	r_s (P)	No.	r_s (P)	No.	r_s (P)	No.	r_s (P)	No.	r_s (P)	No.	r_s (P)
SF-36v2	153	0.85 (<.001)	153	0.46 (.463)	153	0.26 (.260)	153	0.30 (.299)	146	0.58 (<.001)	146	0.14 (.099)	152	0.51 (<.001)
Physical Functioning	147	0.70 (<.001)	147	0.32 (.321)	147	0.25 (.247)	147	0.21 (.210)	146	0.47 (<.001)	146	0.12 (.135)	146	0.35 (<.001)
Role-Physical	152	0.56 (<.001)	152	0.30 (.300)	152	0.25 (.253)	152	0.33 (.332)	145	0.72 (<.001)	145	0.30 (<.001)	151	0.49 (<.001)
Bodily Pain	153	0.37 (<.001)	153	0.18 (.184)	153	0.12 (.122)	153	0.20 (.200)	146	0.28 (<.001)	146	0.47 (<.001)	152	0.31 (<.001)
General Health	147	0.34 (<.001)	147	0.18 (.176)	147	0.25 (.246)	147	0.20 (.197)	146	0.33 (<.001)	146	0.44 (<.001)	146	0.37 (<.001)
Vitality	153	0.61 (<.001)	153	0.33 (.332)	153	0.17 (.167)	153	0.38 (.381)	146	0.42 (<.001)	146	0.32 (<.001)	152	0.54 (<.001)
Social Functioning	152	0.45 (<.001)	152	0.25 (.254)	152	0.15 (.153)	152	0.26 (.260)	145	0.35 (<.001)	145	0.30 (<.001)	151	0.45 (<.001)
Role Emotional	148	0.23 (.004)	148	0.16 (.157)	148	0.15 (.147)	148	0.23 (.225)	146	0.37 (<.001)	146	0.72 (<.001)	147	0.49 (<.001)
Mental Health	153	-0.41 (<.001)	153	-0.29 (<.001)	153	-0.26 (<.001)	153	-0.37 (<.001)	146	-0.74 (<.001)	146	-0.11 (.197)	152	-0.39 (<.001)
Pain NRS														
Average Daily Pain														

Abbreviations: NRS, numeric rating scale; SF-36 v2, Medical Outcomes Study Questionnaire Short Form 36 Health Survey version 2.0; SOSGOQ, Spine Oncology Study Group Outcomes Questionnaire.

reliability of the SOSGOQ2.0 domains were shown to be excellent, with the exception of the posttherapy questions, Mental Health domain, and Social Function domain, which demonstrated moderate to good reliability. The lower reproducibility of the Mental Health domain may be explained by the small size of the domain (2 questions) and the multifactorial and dynamic nature of mental health in patients with cancer. A recent study by Jim et al¹⁶ demonstrated the daily variation in depression, fatigue, activity, and sleep in women undergoing chemotherapy. The severity of the symptoms, other aspects of the disease, or other concomitant treatments (eg, systemic therapy) may have influenced the mental health state, resulting in decreased reliability. Because QOL may deteriorate quickly in patients with advanced stages of cancer, the retest was administered 4 to 9 days after the initial assessment. As such, a stable health state between 12 weeks posttreatment and the retest was assumed rather than objectively assessed, which may have resulted in an underestimation of the reliability of the SOSGOQ2.0. To the best of our knowledge, the optimal interval between the initial assessment and the retest in patients with advanced cancer is complex and remains uncertain.¹⁷ A short interval is proposed because QOL may change rapidly in this population; however, an interval that is too short may enhance the probability of patients recalling their previous answers, thereby compromising the validity of the reliability measurement.¹⁷

The 2 items in the hypothesized Neurological Function domain assessing arm strength and bowel and bladder control demonstrated the largest floor effect and only moderate correlation with the total domain score, which is in agreement with the study of Janssen et al.¹⁵ This can be explained because the majority of patients included in the current study were treated for metastatic disease in the thoracolumbar spine, which may affect the neurological function of the lower extremity (item 5 leg strength) due to radiculopathy or metastatic spinal cord compression, and will not affect the neurological function of the upper extremity (item 6 arm strength). Moreover, only 8% of the patients had diminished or no bowel or bladder control (item 8) according to physical examination. Therefore, the domain score is predominantly defined by the item addressing leg strength (item 5). Analysis of the missing values demonstrated that item 8 (bowel and bladder function) was the main missing item and also showed low reliability, which may be explained by the double-barreled nature of the question. Because item 8 addresses 2 distinct but related physical functions, answering this question may be difficult and may lead to missing values if only 1

TABLE 4. Sensitivity to Change in the SOSGOQ2.0 Based on Change in ECOG Score

Change in Domain	ECOG Decline N = 48	ECOG Stable/ Improved N = 114	Total N = 162	P
Physical function (no.)	41	102	143	.001 ^a
Mean (SD)	-5.7 (24.8)	8.3 (21.8)	4.3 (23.5)	
Neurological function (no.)	41	102	143	<.001 ^a
Mean (SD)	-8.0 (15.7)	3.6 (15.6)	0.3 (16.4)	
Pain (no.)	41	95	136	<.001 ^a
Mean (SD)	3.1 (26.5)	21.0 (29.5)	15.6 (29.7)	
Mental health (no.)	41	95	136	.196 ^a
Mean (SD)	6.4 (24.8)	11.6 (20.2)	10.0 (21.7)	
Social function (no.)	41	101	142	.012 ^a
Mean (SD)	-1.4 (22.5)	9.8 (24.1)	6.5 (24.1)	

Abbreviations: ECOG, Eastern Cooperative Oncology Group; SD, standard deviation; SOSGOQ2.0, revised Spine Oncology Study Group Outcomes Questionnaire.

^aDerived using the Student *t* test.

TABLE 5. Reliability and Internal Consistency for Each of the 6 Subscales of the SOSGOQ2.0 Assessment

SOSGOQ2.0 Domain	12 Weeks (Test)						12 Weeks (Retest)						ICC ^b	95% CI
	No.	Min	Max	Mean	SD	Alpha ^a	No.	Min	Max	Mean	SD	Alpha ^a		
Physical Function score (6 items)	36	17	100	74.0	20.7	0.77	36	13	100	76.5	20.9	0.65	0.92	0.85-0.96
Neurological Function: legs ^c	36	1	4	2.1	1.1	-	36	1	5	1.9	1.2	-	0.73	0.53-0.85
Neurological Function: arms ^c	36	1	5	1.5	0.9	-	36	1	5	1.4	1.0	-	0.72	0.52-0.84
Pain score (5 items)	36	30	100	70.2	19.8	0.74	36	30	100	72.9	21.6	0.68	0.76	0.58-0.87
Mental Health score (2 items)	36	25	100	74.1	20.1	0.85	36	13	100	80.8	23.4	0.79	0.62	0.38-0.78
Social Function score (3 items)	36	25	100	75.1	20.2	0.74	36	42	100	80.8	18.9	0.63	0.63	0.39-0.79
Posttherapy questions (7 items)	36	25	100	71.5	16.6	0.80	36	36	100	69.4	13.8	0.76	0.58	0.32-0.76

Abbreviations: 95% CI, 95% confidence interval; ICC, intraclass correlation coefficient; Max, maximum; Min, minimum; SD, standard deviation; SOSGOQ2.0, revised Spine Oncology Study Group Outcomes Questionnaire.

^aCronbach alpha.

^bFor item responses at 12 weeks after treatment (retest) calculated using an one-way random effect model.

^cReliability is only displayed for 2 neurological items; the item addressing bowel and bladder function is split into 2 questions in the SOSGOQ2.0 and therefore was not displayed.

of the 2 functions is affected. Therefore, changes to the SOSGOQ have been made to ease future data analyses and interpretation. Item 8 (item 9 & 10 SOSGOQ2.0) has been split into 2 distinct questions addressing bowel and bladder function and, rather than calculating a domain score, it is recommended to analyze these items as individual items because they address distinct functions. A domain score may be calculated to obtain an overall impression of the presence of impairment. In addition, a total score for the SOSGOQ2.0 may be calculated by summing the domain scores of the Physical Function, Social Function, Mental Health, and Pain domains (see Supporting Information). The posttherapy items (items 21-27) should be used during follow-up together with the core items (items 1-20).

The current study has several methodological strengths that enhance the generalizability of the results. First, 9 centers across North America participated in the

validation study and 2 centers participated in the test-retest reliability, resulting in a relatively large sample size for this difficult-to-study patient population. Second, both patients who underwent surgery and/or radiotherapy were enrolled in the study. The different treatment modalities in combination with the wide inclusion and exclusion criteria, with no restrictions concerning ECOG performance status, resulted in a heterogeneous, yet representative, population of patients who undergo treatment for spinal metastatic disease. Therefore, this instrument appears to be well-suited for evaluating any treatment related to metastatic spinal disease.

A limitation of the current study was the relatively high dropout rate (24%) within the first 12 weeks. A significant percentage of the patients (16%) died within the first 12 weeks, indicating the severity of disease. Despite maximized efforts for questionnaire compliance and follow-up, a relatively high rate of loss to follow-up was

encountered, which is inherent to the study population. Nevertheless, sensitivity analyses with the baseline questionnaire responses confirmed the results of the primary analyses and the validity of the SOSGOQ2.0 across a wide range of health states. Second, physical function, including motor function and pain, is assessed in greater depth compared with social function and mental health. This is reflected by the uneven number of items per domain. The instrument's bias toward physical function reflects the impact spinal metastases have on these constructs. Third, despite the multicenter approach, the demonstrated validity of the questionnaire is limited to the English version of the SOSGOQ. Translations of the SOSGOQ2.0 into other languages currently are being performed and validation of the translated versions is planned. Last, the minimal clinically important changes in the SOSGOQ2.0 scores over time for a patient are important for interpretation. A study to assess the minimal clinically important differences in the SOSGOQ currently is planned.

The current study investigated the psychometric qualities and clinical validity of the SOSGOQ using a multicenter, prospective observational study. The refined SOSGOQ2.0 was demonstrated to be a valid and reliable spine oncology-specific outcome measure. In future studies evaluating the management of patients with spinal metastases, it is recommended to use the SOSGOQ2.0 in addition to generic HRQOL outcome measures to achieve a more comprehensive representation of HRQOL and to increase sensitivity to change.

FUNDING SUPPORT

This study was organized and funded by AOSpine International through the AOSpine Knowledge Forum Tumor, a pathology-focused working group of up to 10 international spine experts acting on behalf of AOSpine in the domain of scientific expertise. A research grant for this study was received from the Orthopaedic Research and Education Foundation.

CONFLICT OF INTEREST DISCLOSURES

Anne L. Versteeg has received personal fees from AOSpine International for work performed outside of the current study. Arjun Sahgal has received personal fees for past educational seminars from Elekta AB, Accuray Inc, and Varian Medical Systems; has received a research grant from Elekta AB; and has received fees for travel accommodations/expenses from Elekta and Varian Medical Systems for work performed outside of the current study. Dr. Sahgal also belongs to the Elekta MR-Linac Research Consortium. Laurence D. Rhines has received educational grants from Stryker and Medtronic for work performed outside of the current study. Daniel M. Sciubba has received personal fees from Medtronic, DePuy-Synthes, Stryker, NuVasive, and K2M for work performed outside

of the current study. Paul M. Arnold has received fees for sponsored or reimbursed travel from AOSpine North America; has received personal fees from Spine Wave; has intellectual property rights and interests in, equity, or a position of responsibility with Evoke Medical; has equity with Z-Plasty; has received personal fees from Stryker Orthopaedics and Medtronic Sofamor Danek USA; has received personal fees and travel expenses from Stryker Spine and Medtronic; and has received personal fees from In Vivo, In Vivo Therapeutics, SpineGuard, and Ulrich for work performed outside of the current study. Ziya L. Gokaslan has received research support from AOSpine North America and is a stockholder in Spinal Kinetics for work performed outside of the current study. Charles G. Fisher has received personal fees from Medtronic and NuVasive, has received a grant to his institution from the Orthopaedic Research and Education Foundation (which was transferred from Dr. Fisher to AOSpine Knowledge Forum Tumor), and has received fellowship support to his institution from AOSpine and Medtronic for work performed outside of the current study.

AUTHOR CONTRIBUTIONS

Study design: **Arjun Sahgal, Ziya L. Gokaslan, and Charles G. Fisher.** Conduct: All authors. Statistical analyses: **Anne L. Versteeg, Arjun Sahgal, Ziya L. Gokaslan, and Charles G. Fisher.** Article writing: All authors. Guarantor: **Charles G. Fisher.**

REFERENCES

1. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: a population-based study. *Lancet.* 2015;385:1206-1218.
2. Chow E, Nguyen J, Zhang L, et al; European Organization for Research Treatment of Cancer Quality of Life Group. International field testing of the reliability and validity of the EORTC QLQ-BM22 module to assess health-related quality of life in patients with bone metastases. *Cancer.* 2012;118:1457-1465.
3. Street J, Lenehan B, Berven S, Fisher C. 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;35:1377-1386.
4. Ware JE, Kosinski M. The SF-36 Health Survey (Version 2.0) Technical Note. Boston, MA: Health Assessment Laboratory; 1996.
5. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions and reliability across diverse patient groups. *Med Care.* 1994;32:40-66.
6. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34-42.
7. Hays RD, Hayashi T, Carson S, et al. User's Guide for the Multi-trait Analysis of Program. Santa Monica, CA: Rand Corporation; 1988.
8. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, eds. Testing Structural Equation Models. Newbury Park, CA: Sage; 1993:136-162.
9. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107:238-246.
10. Hancock GR, Mueller RO. Structural Equation Modeling: A Second Course. Charlotte, NC: Information Age Publishing; 2006.
11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420-428.
12. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:296.
13. Nunnally JC. Psychometric Theory. 2nd ed. New York: McGraw-Hill; 1978.

14. Sprangers MA, Cull A; on behalf of the EORTC Quality of Life Study Group. A modular approach to QOL assessment-lines for questionnaire development. *Qual Life Res.* 1993;2:72.
15. Janssen SJ, Teunis T, van Dijk E, et al. Validation of the Spine Oncology Study Group-Outcomes Questionnaire to assess quality of life in patients with metastatic spine disease. *Spine J.* 2017;17:768-776.
16. Jim HS, Small B, Faul LA, Franzen J, Apte S, Jacobsen PB. Fatigue, depression, sleep, and activity during chemotherapy: daily and intraday variation and relationships among symptom changes. *Ann Behav Med.* 2011;42:321-333.
17. Paiva CE, Barroso EM, Carneseca EC, et al. A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review. *BMC Med Res Methodol.* 2014;14:8.