

Reliability and Validity of a Japanese-language and Culturally Adapted Version of the Musculoskeletal Tumor Society Scoring System for the Lower Extremity

Shintaro Iwata MD, Kosuke Uehara MD, Koichi Ogura MD,
Toru Akiyama MD, Yusuke Shinoda MD, Tsukasa Yonemoto MD,
Akira Kawai MD

Received: 2 November 2015 / Accepted: 2 May 2016 / Published online: 11 May 2016
© The Association of Bone and Joint Surgeons® 2016

Abstract

Background The Musculoskeletal Tumor Society (MSTS) scoring system is a widely used functional evaluation tool for patients treated for musculoskeletal tumors. Although the MSTS scoring system has been validated in English and Brazilian Portuguese, a Japanese version of the MSTS scoring system has not yet been validated.

Questions/purpose We sought to determine whether a Japanese-language translation of the MSTS scoring system

for the lower extremity had (1) sufficient reliability and internal consistency, (2) adequate construct validity, and (3) reasonable criterion validity compared with the Toronto Extremity Salvage Score (TESS) and SF-36 using psychometric analysis.

Methods The Japanese version of the MSTS scoring system was developed using accepted guidelines, which included translation of the English version of the MSTS into Japanese by five native Japanese bilingual musculoskeletal oncology surgeons and integrated into one document. One hundred patients with a diagnosis of intermediate or malignant bone or soft tissue tumors located in the lower extremity and who had undergone tumor resection with or without reconstruction or amputation participated in this study. Reliability was evaluated by test-retest analysis, and internal consistency was established by Cronbach's alpha coefficient. Construct validity was evaluated using the principal factor analysis and Akaike information criterion network. Criterion validity was evaluated by comparing the MSTS scoring system with the TESS and SF-36.

Results Test-retest analysis showed a high intraclass correlation coefficient (0.92; 95% CI, 0.88–0.95), indicating high reliability of the Japanese version of the MSTS

The institution of one or more of the authors (SI, KO, TY, AK) has received, during the study period, funding from the National Cancer Center Research and Development Fund (Grant 26-A-4).

Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that may pose a conflict of interest with respect to the submitted article.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

Each author certifies that his or her institution approved the human protocol for this investigation, and all investigations were conducted in conformity with the ethical principles of research.

This work was performed at the Chiba Cancer Center, The University of Tokyo Hospital, National Cancer Center Hospital, and Saitama Medical Center, Jichi Medical University.

S. Iwata, T. Yonemoto
Division of Orthopedic Surgery, Chiba Cancer Center, Chiba,
Japan

K. Uehara, K. Ogura
Department of Orthopaedic Surgery, The University of Tokyo
Hospital, Tokyo, Japan

K. Ogura (✉), A. Kawai
Department of Musculoskeletal Oncology, National Cancer
Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
e-mail: ogura-ky@umin.ac.jp

T. Akiyama
Department of Orthopaedic Surgery, Saitama Medical Center,
Jichi Medical University, Saitama, Japan

Y. Shinoda
Department of Rehabilitation Medicine, The University of
Tokyo, Tokyo, Japan

scoring system, although a considerable ceiling effect was observed, with 23 patients (23%) given the maximum score. Cronbach's alpha coefficient was 0.87 (95% CI, 0.82–0.90), suggesting a high level of internal consistency. Factor analysis revealed that all items had high loading values and communalities; we identified a central role for the items “walking” and “gait” according to the Akaike information criterion network. The total MSTS score was correlated with that of the TESS ($r = 0.81$; 95% CI, 0.73–0.87; $p < 0.001$) and the physical component summary and physical functioning of the SF-36.

Conclusions The Japanese-language translation of the MSTS scoring system for the lower extremity has sufficient reliability and reasonable validity. Nevertheless, the observation of a ceiling effect suggests poor ability of this system to discriminate from among patients who have a high level of function.

Introduction

Advances in the treatment of musculoskeletal tumors, including accurate radiographic diagnostic tools, more effective and precise chemotherapeutic regimens, and various alternative reconstruction techniques, have led to a change in paradigm among musculoskeletal oncology surgeons; now, the goal of a treatment is not only to save lives but also to improve the quality of life. Extirpation of tumor tissue is expected to cause some degree of functional impairment; therefore, efforts should be made to decrease such impairment. Studies have reported on the functional outcomes of impairments on musculoskeletal tumor treatment [9, 10, 14, 16, 17, 19, 21].

The Musculoskeletal Tumor Society (MSTS) scoring system was developed in 1985 and revised in 1993 as a physician-derived tool to measure a patient's functional outcome and quality of life after musculoskeletal tumor treatment [5]. Since then, this system has been used in numerous studies to evaluate functional outcomes [9, 10, 14, 16, 17, 19, 22], making it one of the most widely used functional evaluation tools. The original scoring system was written in English and subsequently has been translated into Brazilian Portuguese and Japanese. Although the original and Brazilian Portuguese versions have been validated [8, 13, 22], the Japanese-language and cross-culturally adapted version of the MSTS scoring system has not yet been validated, which is a considerable shortcoming in light of the substantial volume of musculoskeletal oncology research now being done in Japan. Moreover, several tools are available for analyzing physical function or health-related quality of life, such as the Toronto Extremity Salvage Score (TESS) [4] or SF-36 [3], which are patient-derived assessments. Few studies have

shown a correlation between the MSTS scores and those of the two other methods [7, 8, 12].

Therefore, we performed a validation analysis of the Japanese-language translation of the MSTS scoring system for the lower extremity, focusing on psychometric characteristics; specifically, we sought to determine whether the MSTS scoring system had (1) sufficient reliability and internal consistency, (2) adequate construct validity, and (3) adequate criterion validity compared with the TESS and SF-36.

Materials and Methods

This study was designed as a cross-sectional study, and study approval was obtained from the institutional review boards of the participating institutes. The patients meeting the following eligibility criteria were included in the study: (1) a diagnosis of intermediate or malignant bone or soft tissue tumors located in the lower extremity or pelvic girdle, according to the 2013 WHO classification [6]; (2) age between 12 and 85 years; (3) a minimum interval of 6 months after the most recent definitive surgery; and (4) confirmed absence of local recurrence or distant metastasis after definitive surgery. Patient recruitment was conducted from August 2014 to December 2014, and 100 patients agreed to participate (Table 1).

The MSTS scoring system is the most widely used physician-completed scoring system [5] and is based on an analysis of factors pertinent to the patient as a whole and of those specific to the affected limb. It contains six items: pain, function, emotional acceptance, use of any external support, walking ability, and gait. These items were included in 1983 based on recommendations of a committee of the International Symposium in Limb Salvage and after modifications by the MSTS in 1993 [5]. Each of these items is assigned a value of 0 to 5 points, and the total scores are divided by the maximum possible number of points (30 points). The score is subsequently obtained by multiplying the calculated point value by 100.

Translation and Cross-cultural Adaptation

The Japanese version of the MSTS scoring system approved by the Japanese Orthopaedic Association Musculoskeletal Tumor Committee was made available to members in 2010. To develop the version, a translation to Japanese was prepared along with a cross-cultural adaptation of the MSTS scoring system.

The English version of the MSTS scoring system was translated separately by five native Japanese bilingual

Table 1. Descriptive characteristics of the study population

Characteristic	Number of patients (%)
Overall	100 (100)
Age, years; mean [SD]	48.6 (range, 14–82) [17.9]
Sex	
Male	47 (47)
Female	53 (53)
Time from surgery, months; mean [SD]	47.3 (range, 7–229) [50.1]
Tumor location	
Soft tissue	63 (63)
Thigh	37 (37)
Knee	6 (6)
Inguinal region	5 (5)
Pelvis	4 (4)
Buttock	4 (4)
Lower leg	4 (4)
Foot	3 (3)
Bone	37 (37)
Femur	19 (19)
Pelvic bone	13 (13)
Tibia	5 (5)
Histologic diagnosis	
Soft tissue	
Myxoid/round liposarcoma	13 (13)
Well-differentiated liposarcoma	10 (10)
Undifferentiated pleomorphic sarcoma	8 (8)
Synovial sarcoma	8 (8)
Myxofibrosarcoma	6 (6)
Leiomyosarcoma	3 (3)
Others	15 (15)
Bone	
Osteosarcoma	16 (16)
Chondrosarcoma	8 (8)
Giant cell tumor of bone	6 (6)
Malignant fibrous histiocytoma of bone	4 (4)
Others	3 (3)
Type of surgery	
Ablative surgery	3 (3)
Amputation	2 (2)
Rotationplasty	1 (1)
Limb salvage surgery	97 (97)
Resection only	67 (67)
Resection + prosthesis	11 (11)
Resection + biological reconstruction	14 (14)
Curettage	5 (5)

musculoskeletal oncology surgeons with proficiency in Japanese and English. Because each sentence was relatively simple, no professional medical interpreter was used.

Subsequently, all independent translations were compared and combined into one document. Back-translation was not performed; meanwhile, the final version was approved by all translators. During the translation, all translators consented to require no modification from the point of view of cross-cultural adaptation since each description of the item in the original version fits well to the Japanese lifestyle and appeared appropriate. After release of the Japanese version, we have been using it in the clinical setting, and so far, we have had no difficulties with interpretation of the questionnaire.

Psychometric Characteristics

To validate the Japanese version of the MSTS scoring system, psychometric analysis of 100 patients was performed. Investigators who were not surgeons but were research nurses or therapists evaluated each patient's score during an interview in our clinic.

Reliability, referring to the consistency or reproducibility of measurement, was evaluated by test-retest analysis, which consists of two administrations of the same test to the same patient on two different occasions. A second survey was performed for the same patient 2 to 5 weeks after the first survey by either the same investigator or a different investigator. The test-retest reliability of the MSTS scoring system was assessed by calculating the intraclass correlation coefficient (ICC). The ICC was calculated on the basis of the responses of the first (test) and second surveys (retest) for each item and the total score. In general, an ICC greater than 0.9 is regarded as an index of high reproducibility. In addition to test-retest analysis, floor and ceiling effects were calculated for each item and total score. These effects were considered present if greater than 15% of the respondents achieved the lowest (floor effect) or highest (ceiling effect) number of points [18].

Validity refers to a measure of how well a test measures what it claims to measure. In this study, we used three types of analysis for validity. First, internal consistency, which reflects the strength of the relationship among the six items in the system, was identified by calculating Cronbach's alpha coefficient. A coefficient of 0.80 was used as the cutoff, indicating sufficient internal consistency [11], namely, the system includes inconsistent items if the coefficient is less than 0.80. Second, construct validity, which refers to the degree to which the system assesses the underlying theoretical construct it is supposed to measure, generally is evaluated by principal component analysis, which is a statistical method examining the latent structure of the six items in the system. The number of factors to take into account was determined by Kaiser's criteria (eigenvalue > 1 rule) and a scree plot. Next, principal

component analysis followed by repeated varimax rotations was done to calculate each item's factor loading, which represents how much the item explains a variable. Factor loadings can range from -1 to 1 and values close to -1 or 1 indicate that the item strongly affects the variable. In addition to this conventional method, the degree of correlation among the items in the system was evaluated using the Akaike information criterion (AIC) network to examine the latent structure of the system's construct validity; this is a graphic modeling method used to assess the relationship among the items [1]. The Categorical Data Analysis Program (Institute of Statistical Mathematics, Tachikawa, Japan) was used to perform crosstable analyses involving all combinations of the two items in the MSTS system. The program simultaneously searched for the best subset and categorization of explanatory items and automatically indicated matching combinations using the AIC [15]. Third, criterion validity, which measures how well one measure predicts an outcome for another measure, was evaluated by comparing the Japanese version of the MSTS scoring system with evaluation systems already adequately validated, including the TESS and SF-36. The TESS is a disease-specific, self-assessment questionnaire developed for patients with musculoskeletal tumors in the extremities [4]; similarly, the SF-36 is also a self-assessment questionnaire for comprehensive evaluation of health-related quality of life [3]. All patients participating in this study completed the TESS and SF-36 questionnaires at the same time as the first analysis of the MSTS score. The correlation of these measures was assessed using Spearman's correlation coefficient.

The investigators were not surgeons but were research nurses or therapists. The investigators evaluated each patient's score and interviewed each patient in our clinic throughout the study. Total scores generated by the Japanese version of the MSTS scoring system for all 100 patients

ranged from 13 to 100, and the mean score was 82 (SD, 20.4). There were no missing data in any of the surveys.

Statistical Analysis

All statistical analyses were performed using SPSS version 18.0 (SPSS Inc, Chicago, IL, USA). The scores were reported as mean values \pm SD. The threshold for significance was set at a probability less than 0.05.

Results

Reliability and Floor and Ceiling Effects

The ICC between the test and retest of the total score obtained by the Japanese version of the MSTS scoring system was 0.92 (95% CI, 0.88–0.95), indicating high reliability of the system. "Support" showed the highest ICC (ICC, 0.93; 95% CI, 0.90–0.95), whereas "emotional acceptance" had the lowest ICC (ICC, 0.69; 95% CI, 0.57–0.78) (Table 2). No patients obtained the lowest possible total score of 0, indicating the absence of floor effects. In contrast, the highest possible total score of 100 was observed in 23 patients (23%). This suggests a considerable ceiling effect for the total score of the Japanese version of the MSTS scoring system. The same trend was observed for each item; therefore, all items revealed ceiling effects with an absence of floor effects.

We also performed a sensitivity analysis to clarify whether the patients with intermediate tumors (underwent a less-invasive surgery) may contribute to the ceiling effect observed (Fig. 1). The results showed that the average MSTS scores of all the patients with intermediate tumors was 95, whereas that of all the patients without intermediate tumors was 80.

Table 2. Summary of test-retest data

Variable	Test				Retest				ICC	95% CI	Cronbach's alpha
	Minimum (%)	Maximum (%)	Mean	SD	Minimum (%)	Maximum (%)	Mean	SD			
Total (%)	13 (1)	100 (23)	82	20.4	10 (1)	100 (18)	83	19.4	0.92	0.88–0.95	0.87
Pain	0 (1)	5 (69)	4.6	0.8	0 (1)	5 (67)	4.6	0.8	0.85	0.78–0.90	
Function	0 (2)	5 (53)	4	1.3	0 (2)	5 (51)	4.1	1.3	0.88	0.83–0.92	
Emotional acceptance	1 (13)	5 (42)	3.7	1.4	0 (1)	5 (41)	3.8	1.3	0.69	0.57–0.78	
Support	0 (8)	5 (75)	4.1	1.7	0 (7)	5 (74)	4.2	1.7	0.93	0.90–0.95	
Walking	1 (4)	5 (55)	4.2	1.1	1 (3)	5 (48)	4.2	1.1	0.76	0.66–0.83	
Gait	1 (4)	5 (52)	4.1	1.2	1 (4)	5 (47)	4.1	1.2	0.87	0.83–0.91	

ICC = intraclass correlation coefficient; * total score =100 points, best score; 0 = worst score.

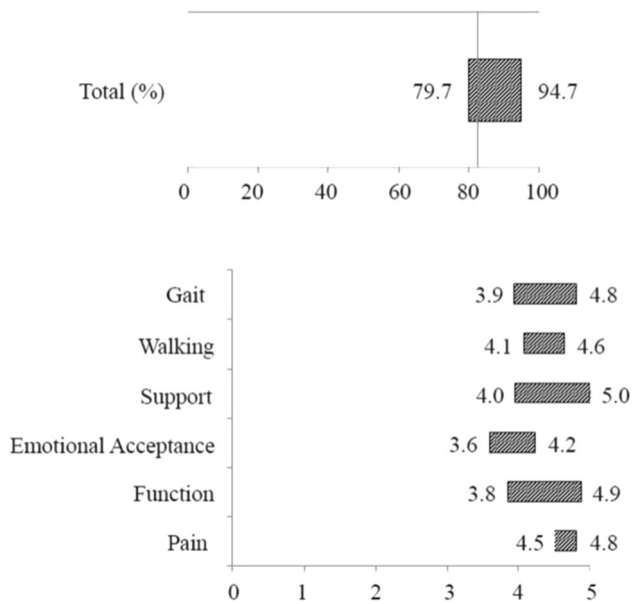


Fig. 1 The sensitivity analysis shows the high performance of the Japanese version of the MSTS scoring system especially in patients with lower function. The average MSTS scores for the patients who underwent less invasive surgery is shown to the right of the bars, and those of the patients who underwent invasive surgery is shown to the left.

Internal Consistency

The overall Cronbach's alpha coefficient was 0.87 (95% CI, 0.82–0.90), suggesting a high level of internal consistency (Table 2).

Construct Validity

On the basis of the results of principal component analysis, the construct validity of the Japanese version of the MSTS scoring system is high, based on the result of principal component analysis. First, we used a scree plot to determine how many factors to extract in the factor analysis (Fig. 2). The scree plot displays the eigenvalues (amount of variation in the total sample accounted for by that factor) in descending order of their magnitude against the number of the factor. A sharp break in the plot suggests the optimal number of the factor. According to the result, the appropriate number of the factor was considered to be one. On the basis of the factor-loading pattern, all items had high loading values and communalities of the first factor (Table 3).

The AIC network, which is another index of construct validity, identified “walking” and “gait” as having a central role among the six factors of the system. Minimal distance assortments, that is, degrees of independence, for

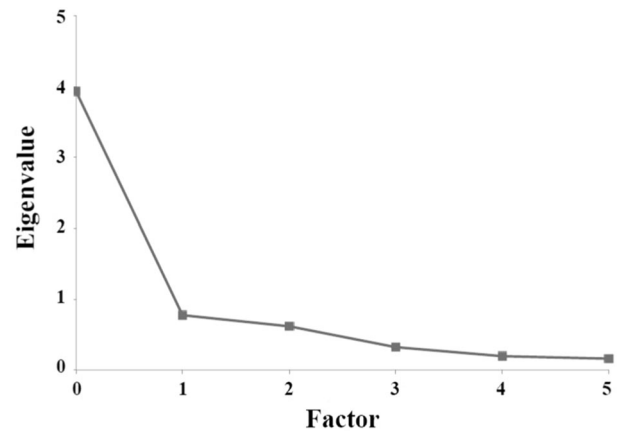


Fig. 2 The scree plot with eigenvalue and factor number of the Japanese version of the MSTS scoring system is shown.

Table 3. Factor loading of MSTS scoring system

Item*	Loading
Pain	0.68
Function	0.91
Emotional acceptance	0.57
Support	0.80
Walking	0.92
Gait	0.92

* Each item refers to the individual five items of the MSTS rating scale; a single item score of 5 points = the best score, and 0 = the worst score.

the two-item groupings was observed, and the AIC network of these six items was observed graphically with the spatial association of the calculation of each item (Fig. 3). This AIC network showed that “walking” and “gait” were related to all the other items; however, the other four items were unrelated to each other. These results were consistent with the highest loading values for “walking” and “gait” in the factor analysis.

Criterion Validity

Criterion validity analysis confirmed that the total score of the Japanese version of the MSTS scoring system was highly correlated with that of the TESS and the physical component summary and physical functioning of the SF-36. The criterion validity was evaluated by correlating the total score of the system with the total score of the TESS and each of the components of the SF-36 (Table 4). The total score of the MSTS system correlated well with that of the TESS ($r = 0.81$; 95% CI, 0.73–0.87; $p < 0.001$). In addition, the physical component summary and physical functioning of the SF-36 correlated with the total MSTS

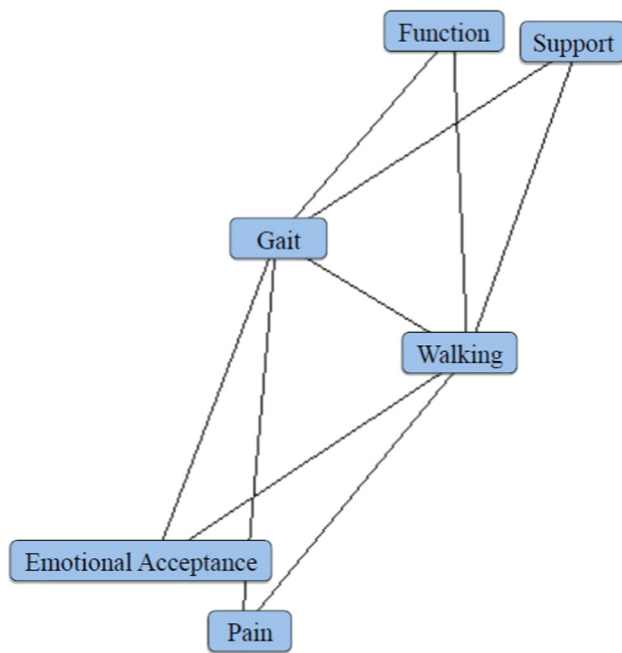


Fig. 3 The Akaike information criterion network of the Japanese version of the MSTS scoring system shows a central role for the items “walking” and “gait”.

score and physical items of the MSTS system (function, support, walking, and gait). However, the mental component summary of the SF-36 did not correlate with the total MSTS score ($r = 0.04$; 95% CI, -0.15 to 0.24) or any of the items of the MSTS scoring system.

Discussion

The MSTS scoring system is a widely used functional evaluation tool for patients undergoing musculoskeletal tumor resection. The Japanese-language version of the MSTS scoring system was made available to members in 2010. The MSTS scoring system has been validated twice [8, 20]; however, the Japanese version has not yet been validated. We aimed to clarify the validity of the Japanese-language version of the MSTS scoring system using psychometric analysis. Our study showed that the Japanese version has sufficient reliability with high ICC by test-retest analysis but also with a high ceiling effect, a favorable internal consistency with a high Cronbach’s alpha coefficient, adequate construct validity indicated by factor analysis and the AIC network, and reasonable criterion validity compared with the TESS and SF-36.

Our study has several limitations. First, the study was performed with Japanese patients using a Japanese version of the MSTS scoring system. The original MSTS scoring system was officially introduced in Japan in 2000 by the Japanese Orthopaedic Association [20], and the Japanese

version was developed by the Japanese Orthopaedic Association Tumor Committee in 2010. Intercultural differences might affect the outcomes; however, the items included in the MSTS scoring system appear to be basic factors. Therefore, there might be little or no difference between the patients from Japan and those from other parts of the world regarding these items. Moreover, several indices of psychometric analyses, such as Cronbach’s alpha coefficient and ICC, were similar to those reported in studies of patients with other ethnicities [4, 8, 13]. Second, we included patients who achieved relatively high MSTS scores compared with patients in other studies [2, 7, 8, 13]. This may have been the case because only a few patients with intermediate tumors (based on the WHO classification of bone and soft tissue tumors [6], in terms of biological potential) were included. Seventeen (17%) patients in our cohort had a diagnosis of an intermediate tumor, such as a well-differentiated liposarcoma (atypical lipomatous tumor) or giant cell tumor of bone, and these patients undergo smaller surgical procedures, such as marginal resection or tumor curettage, respectively, than do patients with malignant tumors who undergo wide resection, leading to superior function. In addition, we performed a sensitivity analysis to analyze the performance of the MSTS scoring system with and without patients who were high-functioning. The result indicated the inferior discrimination capability of the MSTS scoring system for patients with higher function. From these perspectives, this system may not be appropriate for patients undergoing less invasive surgery.

Sufficient reliability of the Japanese version of the MSTS scoring system was confirmed by psychometric analysis; however, we observed a considerable ceiling effect. In this study, 23% of the patients surveyed (23 of 100) achieved the highest possible score, indicating that the Japanese version of the MSTS scoring system has a ceiling effect. This is important because it may reflect low sensitivity in terms of discriminating the patients who have superior function. The ceiling effect was noted even among patients with a malignant tumor; however, the proportion of patients achieving the highest score of this group was lower than the proportion from the intermediate tumor group (20% and 35%, respectively). Nevertheless, the MSTS scoring system contains only six items, which represent basic functions or symptoms, and the Japanese version also achieved a high ICC and a high Cronbach’s alpha coefficient. The effort required by patients for completing the MSTS questionnaire would be much less compared with the TESS (32 items) or SF-36 (36 items). In general, increasing the number of items makes the scale more reliable; the tradeoffs are lower convenience, high time consumed, and more effort required for patients and investigators. The six items in the MSTS system appear

Table 4. Criterion validity

Reference score	Mean	SD	Total	Spearman's correlation coefficients (95% CI)						
				Pain	Function	Emotional acceptance	Support	Walking	Gait	
TESS	86.7	16.5	0.81*** (0.73–0.87)	0.39*** (0.22–0.55)	0.73*** (0.63–0.81)	0.45*** (0.28–0.59)	0.63*** (0.49–0.73)	0.69*** (0.57–0.78)	0.72*** (0.62–0.80)	
SF-36 (physical component summary)	44.7	16.5	0.68*** (0.56–0.77)	0.33*** (0.14–0.49)	0.65*** (0.52–0.75)	0.34*** (0.16–0.50)	0.61*** (0.47–0.72)	0.62*** (0.48–0.72)	0.62*** (0.48–0.72)	
SF-36 (mental component summary)	52.8	9.1	0.043 (–0.15 to 0.24)	0.021 (–0.17 to 0.22)	–0.07 (–0.26 to 0.13)	0.17 (–0.02 to 0.36)	–0.12 (–0.31 to 0.08)	–0.01 (–0.20 to 0.19)	–0.065 (–0.26 to 0.13)	
SF-36 (physical functioning)	48.6	19.2	0.66*** (0.54–0.76)	0.30** (0.12–0.47)	0.62*** (0.48–0.73)	0.35*** (0.16–0.50)	0.57*** (0.42–0.69)	0.59*** (0.44–0.70)	0.62*** (0.50–0.74)	
SF-36 (role physical)	44.4	13.4	0.62*** (0.48–0.73)	0.33** (0.15–0.50)	0.53*** (0.37–0.66)	0.31** (0.12–0.48)	0.52*** (0.36–0.65)	0.54*** (0.39–0.67)	0.52*** (0.36–0.65)	
SF-36 (bodily pain)	50.7	9.1	0.52*** (0.36–0.65)	0.41*** (0.23–0.56)	0.46*** (0.29–0.60)	0.27** (0.08–0.44)	0.42*** (0.24–0.57)	0.49*** (0.32–0.62)	0.41*** (0.24–0.56)	
SF-36 (general health)	51	9.4	0.31** (0.12–0.47)	0.20 (0.01–0.38)	0.20 (0.002–0.38)	0.26** (0.07–0.44)	0.19 (–0.01 to 0.37)	0.32** (0.13–0.48)	0.24* (0.05–0.41)	
SF-36 (vitality)	51.1	8.9	0.33** (0.15–0.49)	0.12 (–0.08 to 0.30)	0.29** (0.10–0.45)	0.23* (0.04–0.41)	0.21* (0.02–0.39)	0.24* (0.05–0.42)	0.29** (0.10–0.46)	
SF-36 (social functioning)	47.6	12.2	0.58*** (0.44–0.70)	0.23** (0.11–0.46)	0.45*** (0.28–0.59)	0.43*** (0.28–0.58)	0.43*** (0.26–0.58)	0.49*** (0.32–0.62)	0.47*** (0.30–0.61)	
SF-36 (role emotional)	46.8	12.4	0.47*** (0.30–0.61)	0.22* (0.02–0.39)	0.41*** (0.23–0.56)	0.23* (0.37–0.41)	0.47*** (0.29–0.61)	0.43*** (0.25–0.57)	0.39*** (0.21–0.54)	
SF-36 (mental health)	51.7	9	0.25* (0.05–0.42)	0.11 (–0.09 to 0.30)	0.11 (–0.08 to 0.30)	0.27** (0.26–0.58)	0.06 (–0.14 to 0.24)	0.18 (–0.02 to 0.36)	0.13 (–0.06 to 0.32)	

TESS = Toronto Extremity Salvage Score; *0.05 > p ≥ 0.01; ** 0.01 > p ≥ 0.001; *** p < 0.001.

adequately reliable despite being concise. These advantages of the MSTS scoring system have been described, adding support to the validation of the system [8, 13].

Psychometric analysis also provides adequate validity of the items included in the Japanese version of the MSTS scoring system. In the current study, construct validity analyses were performed to reveal the latent structure of the MSTS scoring system using principal component analysis and the AIC network. In other words, we performed those analyses to assess whether the Japanese version of the MSTS scoring system measures the functional outcome of the patients with musculoskeletal tumors. According to the results of the factor analysis, the system exhibits high unidimensionality, that is, the Japanese version of the MSTS scoring system can be used as a reliable scale. There was no item showing low factor loading, specifically so regarding items associated with physical activities, such as function, support, walking, and gait. In addition, the central roles of walking and gait identified in the AIC network supported the highest factor loading values for these parameters. Criterion validity showed that a high correlation with total scores of the MSTS scoring system was observed with the TESS and SF-36 physical component summary. Furthermore, these two components significantly correlated with the items associated with the MSTS components regarding the physical activities mentioned above.

Interpretation of the two nonphysical items, pain and emotional acceptance, is a bit less clear in the context of the Japanese version of the MSTS score that we studied. These two items showed adequate, although relatively lower, factor loading according to the factor analysis (0.68 and 0.57, respectively). In addition, criterion validity analysis resulted in no correlation being observed between each of these two items and any component of the SF-36 or the total score of the TESS (Table 4). This indicates that these two can be relatively inappropriate compared with the physical items. The MSTS scoring system is a physician-reported questionnaire, and this type of assessment would not be suitable for evaluating the adequacy of treatment without the bias of the physician, especially regarding the psychological factors. Another reason for less validity of emotional acceptance is a defect in the questionnaire. This item comprises the expressions “enthused”, “satisfied”, “accepted”, and “dislike”, and these expressions appear not to be definitive. The TESS system, one of the most frequently used patient-oriented evaluation systems, was translated into Japanese in 2015 [12]. Data accumulation regarding the outcome evaluation from various aspects would be required in the future.

With better outcomes for patients with cancer achieved through improved cancer therapy, there will be greater demand for a comprehensive outcome assessment of

treatment, including psychometric assessment and health-related quality of life

Through this study, we confirmed that the Japanese version of the MSTS scoring system is a reliable and adequate evaluation system for assessing the functional outcome in patients with musculoskeletal tumors. In addition, the MSTS scoring system is widely used, and continued accumulation of results is valuable. Therefore, it should be used as a standard tool in the future. However, we also found that the system is not sufficient for evaluation of nonphysical factors such as emotional support. Therefore, a new or adjunct evaluation system should be developed and introduced that more completely and comprehensively assesses patient-derived outcomes and that is applicable to patients with higher function.

Acknowledgments We thank Kazuo Saita MD (Department of Orthopaedic Surgery, Saitama Medical Center, Jichi Medical University, Saitama, Japan), Hirokazu Chuman MD (Department of Musculoskeletal Oncology, National Cancer Center Hospital, Tokyo, Japan), and Hirotaka Kawano MD (Department of Orthopaedic Surgery, The University of Tokyo Hospital, Tokyo, Japan) for clinical assistance, and Misuzu Mori, Yoko Kato (Department of Musculoskeletal Oncology, National Cancer Center Hospital, Tokyo, Japan), and Keiko Hishiki (Department of Orthopaedic Surgery, Saitama Medical Center, Jichi Medical University, Saitama, Japan) for administrative support with data collection. Furthermore, we thank Enago (Crimson Interactive KK, Tokyo, Japan) for reviewing the English language.

References

1. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In Petrov BN, Caski F, eds. *Proceedings of the Second International Symposium on Information Theory*. Budapest, Hungary: Akademiai Kiado; 1973:267–281.
2. Aksnes LH, Bauer HC, Jebesen NL, Folleras G, Allert C, Haugen GS, Hall KS. Limb-sparing surgery preserves more function than amputation: a Scandinavian sarcoma group study of 118 patients. *J Bone Joint Surg Br*. 2008;90:786–794.
3. Brazier JE, Harper R, Jones NM, O’Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305:160–164.
4. Davis AM, Wright JG, Williams JI, Bombardier C, Griffin A, Bell RS. Development of a measure of physical function for patients with bone and soft tissue sarcoma. *Qual Life Res*. 1996;5:508–516.
5. Enneking WF, Dunham W, Gebhardt MC, Malawar M, Pritchard DJ. A system for the functional evaluation of reconstructive procedures after surgical treatment of tumors of the musculoskeletal system. *Clin Orthop Relat Res*. 1993;286:241–246.
6. Fletcher CD, Bridge JA, Hogendoorn P, Mertens F. *WHO Classification of Tumours of Soft Tissue and Bone*. Geneva, Switzerland: WHO Press; 2013.
7. Ginsberg JP, Rai SN, Carlson CA, Meadows AT, Hinds PS, Spearing EM, Zhang L, Callaway L, Neel MD, Rao BN, Marchese VG. A comparative analysis of functional outcomes in adolescents and young adults with lower-extremity bone sarcoma. *Pediatr Blood Cancer*. 2007;49:964–969.

8. Lee SH, Kim DJ, Oh JH, Han HS, Yoo KH, Kim HS. Validation of a functional evaluation system in patients with musculoskeletal tumors. *Clin Orthop Relat Res.* 2003;411:217–226.
9. Malek F, Somerson JS, Mitchel S, Williams RP. Does limb-salvage surgery offer patients better quality of life and functional capacity than amputation? *Clin Orthop Relat Res.* 2012;470:2000–2006.
10. Mei J, Zhu XZ, Wang ZY, Cai XS. Functional outcomes and quality of life in patients with osteosarcoma treated with amputation versus limb-salvage surgery: a systematic review and meta-analysis. *Arch Orthop Trauma Surg.* 2014;134:1507–1516.
11. Nunnally JC, Bernstein IH. *Psychometric Theory.* New York, NY: McGraw-Hill; 1994.
12. Ogura K, Uehara K, Akiyama T, Iwata S, Shinoda Y, Kobayashi E, Saita K, Yonemoto T, Kawano H, Chuman H, Davis AM, Kawai A. Cross-cultural adaptation and validation of the Japanese version of the Toronto Extremity Salvage Score (TESS) for patients with malignant musculoskeletal tumors in the lower extremities. *J Orthop Sci.* 2015;20:1098–1105.
13. Rebolledo DC, Vissoci JR, Pietrobon R, de Camargo OP, Baptista AM. Validation of the Brazilian version of the musculoskeletal tumor society rating scale for lower extremity bone sarcoma. *Clin Orthop Relat Res.* 2013;471:4020–4026.
14. Satcher RL Jr, O'Donnell RJ, Johnston JO. Reconstruction of the pelvis after resection of tumors about the acetabulum. *Clin Orthop Relat Res.* 2003;409:209–217.
15. Seichi A, Hoshino Y, Doig T, Akai M, Tobimatsu Y, Iwata T. Development of a screening tool for risk of locomotive syndrome in the elderly: the 25-question Geriatric Locomotive Function Scale. *J Orthop Sci.* 2012;17:163–172.
16. Sherman CE, O'Connor MI, Sim FH. Survival, local recurrence, and function after pelvic limb salvage at 23 to 38 years of followup. *Clin Orthop Relat Res.* 2012;470:712–727.
17. Tang X, Guo W, Yang R, Tang S, Ji T. Synthetic mesh improves shoulder function after intraarticular resection and prosthetic replacement of proximal humerus. *Clin Orthop Relat Res.* 2015;473:1464–1471.
18. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.
19. Teunis T, Nota SP, Hornicek FJ, Schwab JH, Lozano-Calderon SA. Outcome after reconstruction of the proximal humerus for tumor resection: a systematic review. *Clin Orthop Relat Res.* 2014;472:2245–2253.
20. The Japanese Orthopaedic Association Committee of Tumors. *General Rules of Clinical and Pathological Studies on Malignant Bone Tumours (in Japanese).* Tokyo, Japan: Kanehara Shuppan Co; 2000.
21. van de Sande MA, Dijkstra PD, Taminiau AH. Proximal humerus reconstruction after tumour resection: biological versus endoprosthesis reconstruction. *Int Orthop.* 2011;35:1375–1380.
22. Wada T, Kawai A, Ihara K, Sasaki M, Sonoda T, Imaeda T, Yamashita T. Construct validity of the Enneking score for measuring function in patients with malignant or aggressive benign tumours of the upper limb. *J Bone Joint Surg Br.* 2007;89:659–663.