Surgical decision-making for long bone metastases

Janssen, S.J.

Publication date
2018
Document Version
Other version
License
Other

CHAPTER 4

A Comparison Of Questionnaires For Assessing Physical Function In Patients With Lower Extremity Bone Metastases

S.J. Janssen
N.R. Paulino Pereira
K.A. Raskin
M.L. Ferrone
F.J. Hornicek
C.N. van Dijk
S.A. Lozano-Calderon
J.H. Schwab

Journal of Surgical Oncology 2016 Nov;114(6):691-696

Presented at:
Musculoskeletal Tumor Society Annual Meeting 2015, Orlando, Florida, USA.

Poster at:
Massachusetts General Hospital Clinical Research Day 2015, Boston, Massachusetts, USA.
American Academy of Orthopaedic Surgeons Annual Meeting 2016, Orlando, Florida, USA.

ABSTRACT

Objectives

To assess: (1) the degree to which five different questionnaires measure physical function in patients with lower extremity bone metastases, (2) differences in coverage and reliability, and (3) difference in completion time.

Design

Prospective survey study.

Setting

Two tertiary care referral centers for orthopaedic oncology.

Participants

100 of 115 (87%) consecutively invited patients with lower extremity metastases participated in this study between June 2014 and September 2015.

Outcome Measures

The PROMIS Physical Function Cancer, PROMIS Neuro-QoL Mobility, Toronto Extremity Salvage Score (TESS), Lower Extremity Function Score (LEFS), and Musculoskeletal Tumor Society score (MSTS) questionnaires.

Results

All questionnaires measured the same concept; demonstrated by high correlations (>0.7). Floor effect was absent, while ceiling effect was present in all, but highest for the PROMIS Neuro-QoL Mobility (7%). The standard error of measurement was below the threshold –indicating reliability– over a wide range of ability levels for the PROMIS-Physical Function, TESS, and LEFS. Completion time differed between questionnaires ($p<0.001$) and was shortest for the PROMIS questionnaires.

Conclusions

The PROMIS Physical Function Cancer questionnaire is the most useful for measurement of physical function in patients with lower extremity bone metastases. This is due to its reliability over a wide range of ability levels, validity, brevity, and good coverage.

INTRODUCTION

Metastatic disease to bone is common in the lower extremity and is a major contributor to deteriorating function and quality of life.[1-4] Physical function is an important outcome measure to determine success of treatment.[3-5] The Musculoskeletal Tumor Society (MSTS) score and Toronto Extremity Salvage Score (TESS) are most commonly used to evaluate function in patients with bone metastases.[4,6,7] The MSTS score is developed to be completed by the clinician who evaluates function through six items.[7,8] The TESS is the first patient reported outcome for patients with musculoskeletal tumors and assesses function through 30 items.[6,9] Other –non-tumor-specific– patient reported outcome measures exist to assess physical function; however, questionnaires often lack responsiveness (i.e. cannot detect subtle differences in function) for the condition under study, suffer from floor and ceiling effect (i.e. situations in which a subject scores at the lower or upper limit of the questionnaire), or are burdensome due to the number of questions.[10] Therefore, the National Institute of Health funded the Patient-Reported Outcomes Measurement Information System (PROMIS) to develop standardized item banks to measure physical, mental, and social health.[10,11] PROMIS questionnaires allow for Computer Adaptive Testing (CAT) –a dynamic selection of items wherein the response to each item guides the system's choice of the next item– resulting in a tailored series of questions, reducing floor and ceiling effect and questionnaire burden, while maintaining precision.[11,12]

Because function can vary substantially among patients with bone metastases, finding a reliable and valid instrument covering the full range of lower extremity function, while minimizing burden, is beneficial for both clinical and research practice. We therefore sought to determine the most effective, reliable, and efficient questionnaire by comparing the MSTS score, TESS, two PROMIS CAT questionnaires –PROMIS Physical function Cancer and Neuro-QoL Mobility–, and the generic Lower Extremity Function Scale (LEFS).

We assessed which questionnaire is most useful for assessment of physical function in patients with lower extremity bone metastases. Specifically, our study questions evaluated: (1) the degree to which questionnaires measure the same underlying trait (physical function), (2) differences in coverage and reliability, and (3) difference in completion time.

METHODS

Study Design

Our institutional review board approved this prospective study and verbal informed consent was obtained. One hundred fifteen consecutive new or followup patients who presented to two tertiary care referral centers for orthopaedic oncology between June 2014 and September 2015 were invited. We approached English-speaking patients aged 18 years

or older with a metastatic bone lesion, myeloma, or lymphoma of the lower extremity. Patients were enrolled either before or immediately after their visit with the surgeon. Twelve (10%) patients declined participation and 3 (3%) did not finish all questionnaires, leaving 100 patients for analysis. Patients were included once to avoid an effect of learning curve. We collected data using Assessment Center (Northwestern University Feinberg School of Medicine, Chicago, IL, USA, 2007) on a tablet computer. Assessment Center is a free online data collection tool allowing researchers to create surveys and includes the PROMIS CAT item banks.

We calculated that a minimum sample size of 100 participants would provide 90% statistical power (beta 0.10; alpha 0.05) to detect an R-squared of 0.10 (moderate effect size: 0.32) on a bivariate correlation model between questionnaires.

Outcome Measures

Date of birth, race, sex, and tumor type were collected from medical records. Location of the lesion, education, marital status, presence of other disabling conditions, prior surgery or radiation therapy for the lesion, and presence of other bone or visceral metastases were collected from the patient. This was followed by the PROMIS Pain Intensity questionnaire. Subsequently, the PROMIS Physical Function Cancer (CAT), PROMIS Neuro-QoL Mobility (CAT), TESS lower extremity, LEFS, and MSTS lower extremity questionnaires were completed by the patient in random order (Assessment Center has a build in option to randomize the order of questionnaires). The survey consisted of 77 to 93 questions –depending on the number of questions in the CAT questionnaires– and the median completion time was 12 minutes (interquartile range: 9 to 15 minutes).

The PROMIS Pain Intensity is generic and assesses how much pain a patient had over the past seven days and the pain intensity “right now”.

The PROMIS CAT questionnaires consist of a dynamic selection of items. Computer Adaptive Testing administers a minimum of 4 questions and the response to each item guides the system’s choice of the next most appropriate item until either the standard error of measurement (a measure of preciseness) drops below a specified level, or the participant answered the maximum number of questions (set at 12).[10-12] Every PROMIS item comprises 5 response options on a Likert scale. The final score for PROMIS measures is represented by the t-score, a standardized score with a mean of 50 (United States general population) and a standard deviation of 10. A higher t-score represents more of the concept (i.e. better physical function) being measured.

The PROMIS-Physical Function Cancer CAT is an item bank that evaluates general disability in cancer patients.[10-12]

The PROMIS Neuro-QoL Mobility CAT is an item bank that measures lower extremity disability in patients with chronic neurological disease.[13]

The Toronto Extremity Salvage Score (TESS) assesses disability in patients with musculoskeletal tumors.[6,14] The questionnaire consists of 30 items rated on a 5-point Likert scale, but has an additional option "this task is not applicable for me". The raw score is converted to a score ranging from 0 to 100, with a higher score indicating better function.

The Lower Extremity Functional Scale (LEFS) measures lower extremity function in patients with lower extremity conditions (e.g. osteoarthritis).[15] The questionnaire consists of 20 items, and every item is rated on a 5-point Likert scale. The raw score is converted to a score from 0 to 100, with a higher score indicating better function.

The Musculoskeletal Tumor Society (MSTS) score is developed to be completed by the clinician, and consists of six items rated on a 6-point Likert scale.[7,8] The raw score is converted to a score from 0 to 100, with a higher score indicating better function. Although the score is developed to be completed by the clinician, we administered the lower extremity version to patients.

Statistical Analysis

Categorical variables were presented as frequencies and continuous variables were presented as median with interquartile range as histograms suggested non-normal distributions.

We used exploratory factor analysis to assess if included questionnaires measure the same concept (physical function). Exploratory factor analysis is a method that looks for a common underlying –mathematically derived– trait measured by all questionnaires and subsequently correlates the included questionnaires with this trait.[16,17] A factor score of 1 indicates perfect correlation of the questionnaire with the underlying trait and 0 indicates no correlation. Subsequently, we assessed pairwise correlations between questionnaires using the Spearman rank correlation coefficient to further explore if questionnaires measure the same. A coefficient of 1 indicates perfect correlation, while a correlation of 0 indicates no correlation.[18] We considered a correlation coefficient above 0.7 as strong, indicating that questionnaires measure the same.[18,19]

Coverage of the questionnaires was demonstrated by the floor and ceiling effect; floor effect is a term used to describe a situation in which a subject has scores at the lowest possible limit (floor), ceiling effect reflects scores at the upper limit (ceiling).[19,20]

Reliability –precision of measurement– was assessed by calculating the standard error of measurement (SEM) as a function of ability per questionnaire using Item Response Theory.[21] The ability score per questionnaire is set to a standard scale with a mean of 0 and a standard deviation of 1 (Figure 1). The SEM quantifies the degree to which a measurement contains error; i.e. a patient will, by definition, have a "true" ability score and the SEM describes the likely range of actual scores such a patient might have with a specific questionnaire as a result of the unreliability of the questionnaire.[21,22] A SEM below 0.32 is accepted as precise, and corresponds to a Cronbach alpha internal consistency

of 0.90.[21,22] The SEM for the range of ability scores per questionnaire is calculated through Graded Response Models, which fit ordered polytomous items.

Difference in completion time between questionnaires was tested using the Friedman test. Subsequently, pairwise comparisons were made using the Wilcoxon signed-rank test; the alpha level was adjusted using Bonferroni correction for multiple comparisons (Bonferroni alpha level: 0.005, [0.05/10]).

All statistical analyses were performed using Stata® 14.0 (StataCorp LP, College Station, TX, USA). A two-tailed p value less than 0.05 was considered statistically significant unless stated otherwise.

RESULTS

Among the 100 included patients, 59 were women and the median age was 63 years (Table 1). Sixty-six patients had previous surgery and 56 had previous radiation therapy.

All five questionnaires measure the same construct as demonstrated by substantial correlation (factor scores > 0.7) of every questionnaire with the underlying trait (Table 2). This

Table 1: Baseline characteristics of patients with lower extremity bone metastases (n = 100)

	Median (Interquartile Range)
Age (in years)	63 (54 - 70)
Sex	**n**
Men	41
Women	59
Tumor location	
Femur	71
Acetabulum	14
Pelvis	12
Tibia	2
Fibula	1
Race	
White	93
Black or African American	7
Highest grade completed	
High school or below	32
Some college or Bachelor's degree	41
Professional degree	27

Table 1: Baseline characteristics of patients with lower extremity bone metastases (n = 100) (*continued*)

	Median (Interquartile Range)
Marital status	
Married	66
Widowed	12
Single	11
Separated/divorced	7
Living with partner	4
Presence of other disabling conditions*	
Yes	26
No	73
Previous surgery for lesion	
Yes	66
No	34
Previous radiotherapy for lesion	
Yes	56
No	41
Unknown	3
Tumor distribution	
Bone metastases	
Breast	29
Renal cell	11
Prostate	9
Lung	8
Melanoma	4
Bladder	3
Leiomyosarcoma	3
Colorectal	2
Thyroid	2
Stomach	1
Hepatocellular	1
Oesophageal	1
Neuroendocrine	1
Sarcoma	1
Primary bone tumor	
Lymphoma	12
Myeloma	12

*One patient did not complete this question.

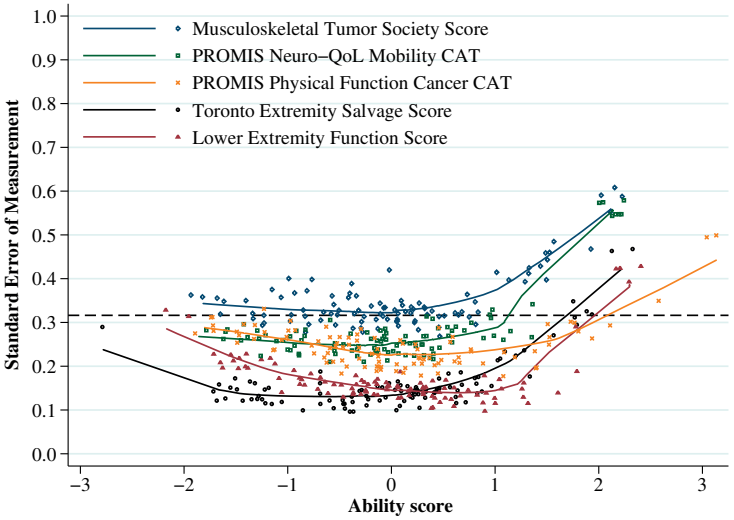

Figure 1: The standard error of measurement –a measure of precision– is demonstrated as a function of the ability scores per questionnaire. Each dot represents a patients' ability and the corresponding standard error of measurement for the specific questionnaire. A higher standard error of measurement means less measurement precision. The ability score is standardized with a mean of 0 and a standard deviation of 1 for all questionnaires to allow for comparison between questionnaires. CAT = Computer Adaptive Testing. QoL = Quality of Life.

finding is supported by high (> 0.7) interquestionnaire correlations, indicating substantial shared variance (Table 3).

Floor effect was not seen in any of the questionnaires, meaning that all questionnaires differentiate well among disabled patients. Ceiling effect was seen in all questionnaires, but was highest (7%) for the PROMIS Neuro-QoL Mobility questionnaire (Table 4), meaning that this questionnaire performed worst when differentiating higher functioning patients.

The standard error of measurement was below the threshold (0.32) of excellent measurement precision –indicating reliability– for all questionnaires in the midrange ability

Table 2: Factor loadings (correlations) from exploratory factor analysis measuring the underlying trait

	Factor1 (underlying trait)
PROMIS Physical Function CAT*	0.878
PROMIS Neuro-QOL Mobility CAT*	0.883
Toronto Extremity Salvage Score	0.922
Lower Extremity Function Scale	0.960
Musculoskeletal Tumor Society Score	0.901

CAT = Computer Adaptive Testing, QOL = Quality of life, PROMIS = Patient Reported Outcomes Measurement Information System

Table 3: Spearman rank correlation coefficient between questionnaires

	Promis Physical Function CAT	Promis Neuro-QOL Mobility CAT	Toronto Extremity Salvage Score	Lower Extremity Function Scale	Musculoskeletal Tumor Society Score
PROMIS Physical Function CAT	1.000				
PROMIS Neuro-QOL Mobility CAT	0.780	1.000			
Toronto Extremity Salvage Score	0.848	0.847	1.000		
Lower Extremity Function Scale	0.867	0.843	0.917	1.000	
Musculoskeletal Tumor Society Score	0.819	0.766	0.868	0.873	1.000

CAT = Computer Adaptive Testing, QOL = Quality of life, PROMIS = Patient Reported Outcomes Measurement Information System; *p* values for all interquestionnaire correlations was below 0.001

scores (theta between -1 and 1), except for the MSTS score (Figure 1). The MSTS score was least reliable as indicated by a high standard error of measurement for the complete range of ability scores. The PROMIS Neuro-QoL Mobility becomes less reliable at ability levels above 1, this means that it loses precision when assessing function in relatively well functioning patients (in line with the ceiling effect). The PROMIS-Physical Function, TESS, and LEFS questionnaires have excellent precision (low standard error of measurement) over a broad range of ability levels, meaning that these questionnaires measure physical function precisely regardless of how poor or good the physical function of the patient is.

Completion time differed among questionnaires ($p < 0.001$) and was shortest for the PROMIS Physical Function and PROMIS Neuro-QoL Mobility, and longest for the TESS (Table 5). There was no difference in completion time between the PROMIS Physical Function and PROMIS Neuro-QoL Mobility questionnaires ($p = 0.75$), but these two questionnaires were quicker to complete than all other questionnaires ($p < 0.001$) (Table 5).

DISCUSSION

We aimed to determine if questionnaires measure the same concept of physical function in patients with bone metastases of the lower extremity and which is most useful –effective, reliable, and efficient– for its assessment. We found evidence that all five included questionnaires measure the same concept of physical function. The PROMIS Physical Function, LEFS, and TESS questionnaires are most reliable and have adequate coverage, but the PROMIS Physical Function is the quickest to complete, through its Computer Adaptive Testing administration.

Our study has several limitations. First, the TESS items contain the option “this task is not applicable for me”; we considered items missing if this option was indicated by the patient

Table 4: Item completion rate, floor and ceiling effect, and score distribution per functional outcome questionnaire

Questionnaire:	Number of items per questionnaire	Item completion rate (%)	Median score (IQR)	Range	Possible range	Floor effect n (%)	Ceiling effect n (%)
PROMIS Physical Function CAT*	4 - 12	456/456 (100)	36 (31 - 43)	19.3 - 68.6	14.6 - 68.6	0 (0)	2 (2.0)
PROMIS Neuro-QOL Mobility CAT*	4 - 12	491/491 (100)	40 (34 - 45)	22.4 - 62.3	15.7 - 62.3	0 (0)	7 (7.0)
Toronto Extremity Salvage Score	30	2789/3000 (93.0)	75 (51 - 89)	2.9 - 100	0 - 100	0 (0)	2 (2.0)
Lower Extremity Function Scale	20	1990/2000 (99.5)	51 (31 - 70)	6.7 - 100	0 - 100	0 (0)	4 (4.0)
Musculoskeletal Tumor Society Score	6	598/600 (99.7)	57 (37 - 70)	4.2 - 100	0 - 100	0 (0)	4 (4.0)

* PROMIS Physical Function v1.1 cancer item bank using Computer Adaptive Testing (CAT), PROMIS Neuro-Quality Of Life (QOL) v1.0 lower extremity (mobility) item bank using Computer Adaptive Testing (CAT), PROMIS = Patient Reported Outcomes Measurement Information System

Table 5: Duration of questionnaire in seconds

				Multiple comparison:			
	Median (IQR)	Range	*p* value	PROMIS Physical Function CAT	PROMIS Neuro-QOL Mobility CAT	Toronto Extremity Salvage Score	Lower Extremity Function Scale	Musculoskeletal Tumor Society Score
PROMIS Physical Function CAT	45 (33 - 62)	13 - 527	<0.001	-				
PROMIS Neuro-QOL Mobility CAT	44 (35 - 63)	14 - 884		0.75	-			
Toronto Extremity Salvage Score	244 (178 - 308)	114 - 800		<0.001	<0.001	-		
Lower Extremity Function Scale	158 (121 - 212)	62 - 743		<0.001	<0.001	<0.001	-	
Musculoskeletal Tumor Society Score	73 (55 - 97)	35 - 556		<0.001	<0.001	<0.001	<0.001	-

CAT = Computer Adaptive Testing, QOL = Quality of life. IQR = interquartile range, PROMIS = Patient Reported Outcomes Measurement Information System

(7.0%) and did not include these in the overall score calculation as per the questionnaires' guidelines.[6] We feel that patients were more likely to indicate this option when they were more disabled and therefore see this as an important limitation. However, we believe that this did not compromise our reliability and completion time analyses. Second, patients completed questionnaires at the outpatient clinic; we did not include patients admitted to the emergency room or ward. We might have missed a subset of more disabled patients, which could have influenced the floor effect. However, all questionnaires differentiated well among the most limited patients. Third, we did not assess test-retest reliability. We see this as a minor limitation as the concept being measured –physical function– might change rapidly in patients with metastatic cancer, rendering test-retest reliability analyses less valid.[23] Fourth, the survey was lengthy and questions within and between surveys seem very similar, which might have bothered patients, resulting in less representative answers. We believe this is a minor limitation as we randomized the order of surveys. Fifth, the study included a heterogeneous patient sample –i.e. patients were included regardless of disease and treatment stage–; however, we see this as a strength rather than a limitation as we aimed to compare questionnaires over a wide range of functional levels. Unfortunately, this did not allow us to assess the impact of specific treatments or disease on physical function.

Exploratory factor analysis can uncover the underlying concept being measured by a number of questionnaires. This statistical method describes variability among observed, correlated variables in terms of a potentially lower number of unobserved variables –traits– to identify relationships between measured variables. In terms of our study; are the five questionnaires comparable in what they are measuring? We demonstrated high factor loadings and substantial interquestionnaire correlations. This suggests that questionnaires measure the same concept in patients with lower extremity bone metastases despite the different conditions for which the questionnaires were originally developed.[6-8,10-13,15] Previous studies were limited comparing only two (MSTS score and TESS) of our five selected questionnaires; an interquestionnaire correlation of 0.48 was found in a sample of 57 patients with soft tissue sarcoma of the thigh; however, the MSTS score was completed by the clinician and not by the patient.[6] Tracking physical function for research studies and clinical use should be done using patient reported outcome measures, as clinicians' outcome assessment tends to underestimate a patients' symptoms and their impact on daily live.[24-28]

In our study, coverage was good –no floor and minimal ceiling effect– for all questionnaires, except for the PROMIS Neuro-QoL Mobility demonstrating a ceiling effect of 7%. This means that the PROMIS Neuro-QoL Mobility questionnaire cannot distinguish higher functioning patients.[13] We therefore discourage the use of this questionnaire in patients with extremity bone metastatic disease. The PROMIS Physical Function, TESS, LEFS, and MSTS score differentiate well among both disabled (no floor effect) and higher functioning

(minimal ceiling effect) patients with lower extremity bone metastases. A study by Davis et al.[29] compared the TESS and MSTS score in 97 patients with lower extremity sarcomas, and found 2% floor effect for the TESS and MSTS score (0% in our study), but a ceiling effect of 33% for the MSTS score (4% in our study). However, the MSTS score was completed by the clinician, potentially overestimating the patients' function. In addition, patients were younger in comparison to our study. A study by Binkley et al.[15] demonstrated no floor or ceiling effect (4% ceiling effect in our study) of the LEFS questionnaire in a sample of patients with lower extremity conditions. The PROMIS Physical Function questionnaire had no floor and only 2% ceiling effect; this superior coverage with only 4 to 12 questions is a result of the dynamic item selection via Computer Adaptive Testing. PROMIS Physical Function questionnaires are also available as short-forms (fixed number of questions). However, these short-forms suffer from ceiling effects as demonstrated by a study in 4,840 cancer patients; 34% ceiling effect for the 4-item short-form, 25% for the 6-item short-form, 13% for the 10-item short-form, and 12% for the 16-item short form.[12] We therefore prefer to use Computer Adaptive Testing for outcome assessment.

We found substantial reliability (i.e. measurement precision) of the PROMIS Physical Function, TESS, and LEFS over a large range of ability levels using Item Response Theory. Traditionally, reliability has been assessed using a single index of internal consistency: Cronbach alpha. However, the issue of a single index is that it assumes the reliability to be the same for all participants over the range of ability scores.[21] The advantage of Item Response Theory is that it demonstrates measurement precision over a range of ability scores.[21] Ability scores at the limits of a questionnaire tend to have more error than those in the middle of the ability range, which is also seen in our study.[21] This means that the precision of measurement decreases when measuring relatively disabled or higher functioning patients. In general, longer questionnaires will provide more information about a patients' physical function and would therefore have less measurement error; however, the dynamic selection of items through Computed Adaptive Testing can optimize reliability while minimizing the number of items. As the MSTS score is to be completed by a clinician, it might be less well understood by patients; this could have compromised its reliability in our study. We therefore discourage the use of MSTS score to capture patient reported outcomes. The relatively high standard error of measurement of the PROMIS Neuro-QoL Mobility questionnaire in the higher functioning patients is in line with its ceiling effect: higher functioning patients are not distinguished nor measured accurately by this questionnaire.

Obtaining research data in the least amount of time is key in reducing questionnaire burden and attrition of patients. We found that the PROMIS Physical Function and PROMIS Neuro-QoL Mobility are quickest to complete; 73% of the participants completed these questionnaires within a minute, compared to an average completion time of more than 4 minutes for the TESS, and almost 3 minutes for the LEFS.

In conclusion, this study demonstrated that the PROMIS Physical Function Cancer questionnaire is the most useful to determine physical function for patients with lower extremity bone metastases. This is due to its reliability over a wide range of ability levels, validity, brevity, and good coverage through Computer Adaptive Testing.

REFERENCES

1. Bickels J, Dadia S, Lidar Z. Surgical management of metastatic bone disease. J Bone Joint Surg Am. Jun 2009;91(6):1503-1516.
2. Quinn RH, Randall RL, Benevenia J, Berven SH, Raskin KA. Contemporary management of metastatic bone disease: tips and tools of the trade for general practitioners. J Bone Joint Surg Am. Oct 16 2013;95(20):1887-1895.
3. Steensma M, Healey JH. Trends in the surgical treatment of pathologic proximal femur fractures among Musculoskeletal Tumor Society members. Clin Orthop Relat Res. Jun 2013; 471(6):2000-2006.
4. Talbot M, Turcotte RE, Isler M, Normandin D, Iannuzzi D, Downer P. Function and health status in surgically treated bone metastases. Clin Orthop Relat Res. 2005;438:215-220.
5. Steensma M, Boland PJ, Morris CD, Athanasian E, Healey JH. Endoprosthetic treatment is more durable for pathologic proximal femur fractures. Clin Orthop Relat Res. 2012;470: 920-926.
6. 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. Oct 1996; 5(5):508-516.
7. 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. Jan 1993(286):241-246.
8. 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. Dec 2013;471(12):4020-4026.
9. Cheng EY. Prospective quality of life research in bony metastatic disease. Clin Orthop Relat Res. Oct 2003(415 Suppl):S289-297.
10. Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. J Clin Oncol. Nov 10 2007;25(32):5106-5112.
11. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol. Jan 2008;61(1):17-33.
12. Jensen RE, Potosky AL, Reeve BB, et al. Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res. May 3 2015.
13. Gershon RC, Lai JS, Bode R, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. Qual Life Res. Apr 2012;21(3):475-486.
14. Clayer M, Davis A. Can the Toronto Extremity Salvage Score produce reliable results when used online? Clin Orthop Relat Res. Jun 2011;469(6):1750-1756.

15. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther. Apr 1999;79(4):371-383.
16. Revicki DA, Chen WH, Harnam N, et al. Development and psychometric analysis of the PROMIS pain behavior item bank. Pain. Nov 2009;146(1-2):158-169.
17. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. Pain. Jul 2010;150(1):173-182.
18. Dancey C, Reidy J. Statistics without Maths for Psychology: using SPSS for Windows: London: Prentice Hall; 2004.
19. Tyser AR, Beckmann J, Franklin JD, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. J Hand Surg Am. Oct 2014;39(10):2047-2051 e2044.
20. Hung M, Nickisch F, Beals TC, Greene T, Clegg DO, Saltzman CL. New paradigm for patient-reported outcomes assessment in foot & ankle research: computerized adaptive testing. Foot Ankle Int. Aug 2012;33(8):621-626.
21. Frank B. Baker. The Basics of Item Response Theory: ERIC Clearinghouse on Assessment and Evaluation; 2001.
22. Tighe J, McManus IC, Dewhurst NG, Chis L, Mucklow J. The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. BMC Med Educ. 2010;10:40.
23. Davidshofer KR, Murphy CO. Psychological testing: principles and applications. 6 ed: Upper Saddle River, N.J.: Pearson/Prentice Hall; 2005.
24. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ. 2015;350:g7818.
25. Nelson E, Conger B, Douglass R, et al. Functional health status levels of primary care patients. JAMA. Jun 24 1983;249(24):3331-3338.
26. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med. Mar 11 2010; 362(10):865-869.
27. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. Dec 2 2009;101(23):1624-1632.
28. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. Aug 2008;14(8): 530-539.
29. Davis AM, Bell RS, Badley EM, Yoshida K, Williams JI. Evaluating functional outcome in patients with lower extremity sarcoma. Clin Orthop Relat Res. Jan 1999(358):90-100.