The Validity, Reliability, and Responsiveness of Commonly Used Orthopedic Outcome Measures, Cancer Specific Measures, and Patient Reported Functional and Quality of Life Measures

Justin E. Bird MD*, Joseph E. Niland, Valerae Lewis MD, Theresa Nalty, PhD, PT, NCS

Justin Bird, MD
E-mail: jebird@mdanderson.org
Joseph Niland, MS
E-mail: jeniland@mdanderson.org
Valerae Lewis, MD
E-mail: volewis@mdanderson.org
Theresa Nalty, PhD, PT, NCS
E-mail: tnalty@mdanderson.org

*Department of Orthopaedic Oncology, University of Texas MD Anderson Cancer Center, Houston TX

Background: Outcome measures are used in orthopedics to capture patient's self-reported level of disability, quality of life, functional ability, and to determine the relative effectiveness of interventions. These tools are used by healthcare providers to determine a plan of care, document change, and demonstrate the effectiveness and appropriateness of care. Older instruments appear to have been developed at a time when researchers were not utilizing item response theory to evaluate the items within the outcome measure. Even though some of the orthopedic measures have been used for years they may not have been shown to have a high level of test-retest reliability, responsiveness to change over time, or validity. In addition, the majority of the orthopedic scores, designed for arthritic patients, may not be adequate in the orthopedic oncology practice to capture change over time. Cancer patients are often in chronic pain which may only be partially alleviated by orthopedic surgery. A cancer patient in remission who receives a shoulder replacement due to arthritis will likely have a different functional outcome on a shoulder score than a cancer patient who has metastasis to the bone requiring a joint replacement with adjunct chemotherapy or radiation. Clinicians and investigators should select outcome measures that have been developed with appropriate patient input for item generation and reduction and have demonstrated higher levels of validity and reliability.

Purpose: 1.) Evaluate current orthopedic outcome measures to assess their levels of validity, reliability and responsiveness 2.) Educate providers on how to select an outcome tool that is pertinent to the patient population based on an understanding of validity, reliability, and responsiveness.

Methods: Commonly used orthopedic, oncology, functional, and quality of life outcome measures were searched in Scopus to determine the convergent validity, reliability (test-retest and internal consistency), responsiveness, Minimal Detectable Change (MDC), and Minimal Clinically Important Difference (MCID). Convergent validity establishes the association of a tool to a previously validated measure of a similar construct. Test-retest reliability assesses the degree to which test scores are consistent from one test administration to the next. Internal consistency measures the consistency of results between items within the test. Responsiveness is the ability of a measure to detect change over time in the construct of interest. The MDC is defined as the minimal change that falls outside the measurement error in the score of an instrument used to measure a symptom. The MCID is of clinical relevance because it is the amount of change between scores that can be considered as a clinically important difference in the patient's disability given the specific diagnosis. A floor effect indicates the tool is unable to assess deterioration in the condition and a ceiling effect indicates the tool is unable to assess improvement in the condition.

Interpretation of values: In this study we ranked convergent validity on an order of magnitude of high (0.70 to 1.00, moderate (0.40 to 0.69), or unacceptable (<.40). Internal consistency was scored as excellent (α=0.91 to 0.95), good (α=0.81 to 0.90), acceptable (α=0.71 to 0.80), or, unacceptable redundancy (α >0.95). Test-retest

reliability was ranked as excellent (0.80 to 1.0), good (0.64 to 0.79), moderate (0.51 to 0.63), or unacceptable (< 0.51). Responsiveness was operationalized in our study by either the area under the curve (**AUC**), the effect size (**ES**) ***or*** the standardized response mean (**SRM**). AUC was ranked excellent (>0.84) or acceptable (0.74 to 0.83). SRM was ranked acceptable (0.9 to 1.9). ES was large (0.8 to 0.89), moderate (0.5 to 0.79), or poor (<0.5).

Results: See Table 1 for the validity and reliability of general orthopedic scores and Table 2 for orthopedic oncology scores.

Table 1

Frequency of publication in ( )	Test-Retest Reliability	Internal Consistency	Convergent Validity	Responsiveness	MDC / MCID
Clinician Reported Measures					
ASES (8010)	**E**	**G**	**M** with MEPS **H** with pain	**E** for AUC	MDC 9.7 MCID (shoulder) 6.4, except RTC 12-17
Mayo Elbow Score (546)		**U**	**M** with DASH **M** with ASES	**L** (ES=1.12 to 2.71)	MCID 15
Mayo Wrist Score (330)					
Harris Hip Score (2844)					MCID 7-9
Knee Society Score (2247)				**U** (SRM 0.8)	
American Foot & Ankle Score(1868)					
Patient Reported Measures					
VAS Pain (3747)					VAS MDC 3cm
Borg RPE (55)					
Borg Dyspnea (67)					
Neck Disability Index (1004)	**M, G, E** depending on Dx	**A, G, E** depending on Dx	**H** with DASH (whiplash)	**P** for radiculopathy	MDC 8.4 to 19.6, MCID 13.5 to 19
Oswestry Disability Index (3749)	**U** for pain standing travel, sex, sleep, accept lifting , sitting	**U** for walking, pain, lifting, sitting, standing, sleeping, etc (all except travel)	**M** with SF36 social, Mod with VAS	**U** (AUC .71). Responsive only to those who get worse	MDC 11.74-13.67, MCID 9.5 to o12.8 for acute pain and 15.35 for chronic pain
Oxford Elbow Score (35)		**G** for pain, function, psychosoc	**H** with DASH and SF36 for function **M** with MEPS, DASH, SF36 for pain only	**L** (ES >1.49) overall score, 1.15 for pain and function, 1.13 psychosoc	
DASH (1415)	**E**		**M** with SF36	**U** (SRM 0.74-0.80) for baseline to after treatment and **A** (SRM 0.92-1.40) for patients who said they were better	MDC 10 MCID 10
Quick DASH (1761)	**E**	**E**	**H** with DASH and VAS	**U** (SRM .79)	MDC 11 MCID 19

WOMAC (4513)	**E**	**E**	**M** with SF36	**A** for pain, function, stiffness	MDC 9-10 function, MDC 5.51 pain MCID 9-12 on 100 scale
Foot & Ankle Disability Index (94)					
Michigan Hand Outcomes Questionnaire (405)	**E**	**E**	**M** with SF 12	**E** for function, work, and pain	MDC 11-23 pain, MDC 13 function, MDC 8 work, MCID 11-23 based on dx. Ceiling effect with trauma

Table 2

Frequency of publication in ( )	Test-Retest Reliability	Internal Consistency	Convergent Validity	Responsiveness	MDC / MCID
Clinician Reported Measures					
ASIA (1038)			**M** with 10 MWT (ASIA-D)		MDC 12.95 light touch, 1 motor, pin prick 7.8
Karnofsky (8993)			**M** with Neuro QOL UE		
MSTS (778)	**E**	**G**	**M** with TESS		
Moran Biagini Neurologic Classification (1)					
ECOG (15770)			**U** with ESAS		
Spine Instability Neoplastic Score (38)					
Timed Up and Go test (2956)	**G to E**		**H** with Berg B **M** with walking speed and Tinetti		MDC 2.9 to 4.85 sec
Timed 6 meter (of 10 meter) walk test (163)	**E**		**H** with Berg Balance		MDC 0.7 to 0.82 m/s MCID 0.13 to 0.16 m/s
Berg Balance Scale (1826)	**E**	**G**	**H** with 10 MWT and TUG	**M to A** (ES=.66 to .97)	MDC 6.5, 8 for older adults

9 hole peg test (268)	**E**		**H** with motoricity Test	**M** (ES = .52 to .66)	MDC 2.6 dom hand, 1.3 non-dom
Patient Reported Measures					
MDASI (410), MDASI spine (4)	**G**	**G**			
Brief Fatigue Inventory (451)					
SF 36 (28328) SF36 v2 (106)	**E**	**A**	**M** with WOMAC **M** with Knee Society	**A** all domains except soc fx and role emo	Floor and ceiling effects role emot. Ceiling effects Phys func, role phys, and pain
PROMIS sexual function (10)	**G**	**A to G**	**M** with FSFI **H** with IIEF		
Barthel Index (7225)		**G**	**H** with FIM	**A** (SRM 1.2)	MDC 4.02 MCID 1.85
Edmonton Symptom (324)	**G**	**G**	**U** with ECOG		
Spine Oncology Surgery Group Outcomes Questionnaire (2)					
Toronto Extremity Salvage Score (139)		**G**	**M** with MSTS and EORTC QOQ C30		
PROMIS Global Health short form (12)					
PROMIS Physical Function 10 q (56)			**M** with pain		MDC 4-6 T scores
MSKCC Bowel (21)	**G to E**	**G**			
Neurogenic Bladder Symptom Score (5)	**E**	**G**			
Rivermead Mobility (299)	**G to E**	**A to R**	**H** with FIM and timed Walk Test	**A** (SRM 1.14 to 1.94)	MDC 2.2
PROMIS NeuroQOL UE Function (14)		**A to E**	**M** with EQ 5D		
EuroQOL EQ-5D (6959)			**M** with PROMIS Neuro QOL	**P** (ES=.10)	
EORTC QLQ-C30 (4003)	**E**	**A to E**			
Pelvic Girdle Questionnaire (3)	**E**		**E** with UIQ, POPIQ, CRAIT		
Pelvic Floor Impact Questionnaire (247)	**G**	**A to R**		**M to L** (ES .67 to 1.25)	MDC 64.8 MCID 36

Key:

Test-Retest E= Excellent (0.8 to 1.0); G=Good (0.64 to 0.79); M=Moderate (0.51 to 0.63); U= Unacceptable <.51
Inter Consistency: E= Excellent (0.91 to 0.95); G=Good (0.81 to 0.90); A= Acceptable (0.71 to 0.80); R=Redundant >.95; U <0.71
Convergent Validity: H=High (0.7 to 1.0); M=Moderate (0.4 to 6.9); U=Unacceptable <0.40
Responsivenss: ***AUC***: E=Excellent >0.84; A=Acceptable (0.74 to 0.83); U=Unacceptable <0.73 ***SRM:*** A=Acceptable (0.9 to 1.9);
ES: L=Large (0.8 to 0.89); M=Moderate (0.5 to 0.79); P=Poor <0.5
MDC: Minimal Detectable change
MCID: Minimal Clinically Important Difference

Conclusions: Of the 45 measures (Table 1 and Table 2) only 10 (22%) have data demonstrating responsiveness (per our rankings of excellent, acceptable, or large). 78% of these highly regarded outcome measures either lack data demonstrating the ability to detect changes over time (responsiveness) or are poor measures to assess responsiveness. It is critical that orthopaedic oncologists know which outcome measures have high levels of validity, reliability and responsiveness for the particular populations we treat, given that these outcome measures are tied to value (outcomes related to cost) and ultimately reimbursement. It is imperative that orthopaedic oncologists identify the outcome tools that we should use in our practice given the specific population of patients we treat. We strongly recommend that the outcome measures currently in high use be rigorously tested to determine their levels of responsiveness for orthopedic oncology patients. Whenever possible, the clinician-reported and patient-reported outcome measures with the highest levels of validity, reliability and responsiveness should be utilized.