# Assessing Physical Function in Patients With Lower Extremity Bone Metastases

Authors: Stein Janssen, MD; Nuno Cools Paulino Pereira, MD; Kevin Raskin, MD; Marco Ferrone, MD; Francis Hornicek, MD, PhD; Santiago Lozano-Calderon, MD, PhD; Joseph Schwab, MD

**Background:** Bone metastatic disease is common in the lower extremity and is a major contributor to deteriorating physical function and quality of life in patients with cancer. Physical function is an important outcome measure to determine success of surgical treatment. Because function can vary substantially among patients with metastatic bone disease, finding a reliable instrument that covers the full range of lower extremity function, while minimizing questionnaire burden to the patient and researcher, is important for both clinical and research practice.

**Questions/Purposes:** This study aimed to assess which questionnaire –PROMIS Physical Function, PROMIS Neuro-QoL Mobility, Toronto Extremity Salvage Score (TESS), Lower Extremity Function Score (LEFS), MSTS score– 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, reliability and validity among questionnaires, and (3) difference in completion time.

**Patients and methods:** Eighty-two English speaking patients above 18 years of age with a metastatic bone lesion, myeloma, or lymphoma of the lower extremity (pelvis, acetabulum, femur, or tibia) were enrolled in this IRB approved prospective study. Five questionnaires –PROMIS Physical Function, PROMIS Neuro-QoL Mobility, TESS, LEFS, and MSTS score– were completed by all patients in random order using Assessment Center (survey tool) on a tablet computer. The PROMIS Physical Function and PROMIS Neuro-QoL were administered through Computer Adaptive Testing (CAT), a dynamic selection of items wherein the response to each item guides the system's choice of the next item.

We used exploratory factor analysis –a method that correlates questionnaires with a common underlying mathematically derived trait– to assess if questionnaires measure the same concept. Subsequently, we assessed pairwise associations between questionnaires using Spearman rank correlation to further explore their relation.

Coverage of the questionnaires was demonstrated by the floor (score at the lowest possible limit) and ceiling (score at the upper limit) effect. To assess validity, we evaluated how well the outcome measures distinguished –concurrent validity– among patients self rated degree of disability using Kruskal-Wallis analysis.

Reliability was assessed by calculating the standard error of measurement (SEM) as a function of ability per questionnaire using Item Response Theory analysis. The ability score per questionnaire is set to a standard scale –theta– with a mean of 0 and a standard deviation of 1 (Fig. 1). The SEM quantifies the degree to which a measurement contains error (below 0.32 is accepted as precise; dashed line in Fig. 1).

Difference in completion time among questionnaires was tested using the Friedman test. Subsequently, pairwise comparisons were made using the Wilcoxon signed-rank test.

**Results:** There were 30 men (37%) and 52 (63%) women, with a median age of 63 years. The femur was most commonly affected (71%). Breast (33%), renal cell (12%), and prostate (9.8%) were the most common metastatic lesions. Nine (11%) patients had a lymphoma, and 8 (9.8%) multiple myeloma.

All five questionnaires measure the same construct as demonstrated by high interquestionnaire correlations (rho > 0.7) and high correlations (rho > 0.7) with the underlying trait. The lowest interquestionnaire correlation was between the MSTS score and PROMIS Neuro-QoL Mobility (rho = 0.72, p < 0.001) and the highest correlation was between the TESS and LEFS (rho = 0.90, p < 0.001).

Floor effect was not seen in any of the questionnaires, while Ceiling effect was seen in all, but was highest for the PROMIS Neuro-QoL Mobility questionnaire (6.1%). All questionnaires differentiate well –indicating concurrent validity– among patients self rated degree of disability (p < 0.001).

The standard error of measurement was below the threshold of excellent measurement precision – indicating reliability– over a wide range of ability levels for the PROMIS-Physical Function, TESS, and LEFS (Fig. 1).

Completion time differed among questionnaires (p < 0.001) and was shortest for the PROMIS Physical Function questionnaire (Median 45 seconds), and longest for the TESS (Median 249 seconds).

**Conclusions:** This study demonstrated that the PROMIS Physical Function is superior to the PROMIS Neuro-QoL Mobility, TESS, LEFS, and MSTS score as a result of its reliability over a wide range of ability levels, validity, brevity, and good coverage through Computer Adaptive Testing.