A New Paradigm for Patient-Reported Outcomes Assessment in Foot and Ankle Research: Computerized Adaptive Testing

**Foot & Ankle Category:** Other

**Author(s):**
Charles L. Saltzman, MD  
Man Hung, PhD  
Florian Nickisch, MD  
Timothy C. Beals, MD  
Daniel Clegg, MD  
Tom Greene, PhD

**Introduction**
Measuring, reporting and comparing outcomes are essential for improving health care delivery. The problems with current health status scales include patient fatigue, floor/ceiling effects and validity/reliability. The lack of valid, precise and full range instruments prompted the NIH to develop the Patient-Reported Outcomes Measurement Information System (PROMIS) facilitating computerized adaptive testing (CAT). From the PROMIS 124-item physical function bank a 79-item lower extremity (LE) bank was derived with particular relevance to foot and ankle problems. This study compared the CAT of the LE item bank (LE CAT) to two widely accepted scales – the Foot and Function Index (FFI) and the sport subscale of the Foot and Ankle Ability Measure (FAAM). Primary instrument characteristics measured were 1) average time for completion, 2) reliability and 3) precision across the range of lower extremity function.

**Methods**
Foot and ankle surgery patients from one center completed questionnaires administered on IPads. For the 287 subjects reported in this study, mean age was 47.5 years, 60% were male, and > 60% described their daily activities as either heavy labor, or mostly walking and standing. 12% had rheumatoid arthritis and 28% had pain or deformity in the hips, thighs, or knees that limited their activities. 29% also had pain or deformity in the opposite foot or ankle that restricted their activities. Time to complete each scale was automatically calculated by the testing software. A Rasch partial credit model was applied to calibrate all items; those items that showed misfit to the model were deleted. The remaining items within each instrument were then recalibrated and then equating technique was used to set all three instruments to a common scale. Finally, person and item measures and standard errors were computed to examine instrument precision and coverage.

**Results**
Time to Completion It took on average 63 seconds for a patient to complete the LE CAT, 132 seconds to complete the sports FAAM and 241 seconds to complete the FFI. Fit and Reliability Seven FFI items showed an infit and outfit MNSQ of 2.0; thus they were deleted from further analyses of the data. After recalibration of the remaining 16 FFI items, none showed any serious misfit. Item and person reliabilities for the LE CAT, FFI and sports FAAM were all exceptionally high: 0.96, 0.99, 0.99 and 0.98, 0.94 0.91, respectively. Precision The LE CAT had the best overall precision across the entire
range of physical functioning. All three questionnaires had relatively good precision in the mid and part of the upper range of function, but only the LE CAT was able to maintain precision in the lower range of function. The FFI covered the upper and mid levels well, but lacked items in the lower range; but FAAM sports lacked items to cover both the upper and the lower ends.

**Conclusion**
Using a computerized adaptive testing approach to select questions from a lower extremity item bank (LE CAT) reduced patient burden and improved precision compared to two commonly used foot and ankle scales.