J Shoulder Elbow Surg (2017) ■■, ■■–■■

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

ORIGINAL ARTICLE

Psychometrics of the Patient-Reported Outcomes Measurement Information System Physical Function instrument administered by computerized adaptive testing and the Disabilities of Arm, Shoulder and Hand in the orthopedic elbow patient population

Man Hung, PhD*, Maren W. Voss, MS, Jerry Bounsanga, BS, Yushan Gu, BS, Erin K. Granger, MPH, Robert Z. Tashjian, MD

Department of Orthopaedic Surgery Operations, University of Utah, Salt Lake City, UT, USA

Background: This study evaluated the psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF) instrument administered through computerized adaptive testing (CAT) compared with the traditional full-length Disabilities of the Arm, Shoulder and Hand (DASH).

Methods: The PROMIS PF CAT and the DASH were administered to 1759 patients seeking care for elbow conditions. This study used Rasch partial credit modeling to analyze the instruments with item fit, internal reliability, response category thresholds, dimensionality, local independence, gender differential item functioning, and floor and ceiling effects.

Results: The PROMIS PF CAT and DASH had satisfactory item fit for all but 1 item on both measures. Internal reliabilities were high for both measures. Two items on the DASH and 4 items on the PF CAT showed nonordered category thresholds. Unidimensionality was adequate, and local independence was supported for both instruments. Gender bias was found for 4 items on the PF CAT and 12 items on the DASH. Both measures had adequate instrument targeting and satisfactory floor and ceiling effects.

Conclusion: The PROMIS PF CAT and the DASH both showed sufficient unidimensionality, good item fit, and good local independence with the exception of high levels of gender item bias, particularly for the DASH. Further scale evaluation should address item bias and item response categories for these instruments. Overall, the PROMIS PF CAT is an effective outcome instrument to measure function in patients with elbow disorders that requires significantly fewer questions than the DASH.

Level of evidence: Basic Science Study; Validation of Outcome Instruments

Keywords: PROMIS; DASH; physical function; psychometric; orthopedics; elbow

Value-based propositions in health care have been the focus of health care reform[24] and are evidenced by increases in bundling payment models[11] and patient safety-based reimbursement penalties.[27] This increased emphasis on value over volume requires a metric or measuring system to accurately

The University of Utah Institutional Review Board approved this study (IRB #00053404).

*Reprint requests: Man Hung, PhD, Department of Orthopaedics, University of Utah School of Medicine, 590 Wakara Way, Salt Lake City, UT 84108, USA.

E-mail address: Man.Hung@hsc.utah.edu (M. Hung).

1058-2746/$ - see front matter © 2017 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.
https://doi.org/10.1016/j.jse.2017.10.015

assess and compare treatment effects. Precise patient-reported outcome (PRO) measurement is critical to the process of improving care while controlling costs.[30] Orthopedics as a profession needs to take steps to understand the psychometric properties of outcome measures to collect useful data that can show measurable improvements to justify treatment costs.[2]

The widely used Disabilities of the Arm, Shoulder and Hand (DASH) has been shown to be a reliable, responsive, and valid measure of upper extremity disability.[12,17,26] However, depending on the population being assessed, it is prone to floor and ceiling effects.[16] The DASH is a 30-item questionnaire that can take as much as 5 minutes for patients to complete, adding to testing burden, particularly if it is administered with other instruments.[28] Newly developed measures using computerized adaptive testing (CAT) can minimize test burden and improve the patient's experience.[31]

The National Institutes of Health funded the Patient-Reported Outcomes Measurement Information System (PROMIS) to develop and improve existing PRO instruments.[7] Included is the PROMIS Physical Function (PF) CAT version 1.2, a 124-item instrument that assesses upper extremity, lower extremity, central function, and functioning in activities of daily living.[14] The CAT administration uses computer algorithms to select targeted questions that assess functional ability without redundancy. This is accomplished as individual items are selected from the 124-item bank that incorporate or are altered based on a patient's prior answers. The PROMIS PF CAT has shown high correlation with its full-length test version, while significantly reducing testing time and test burden.[6,19,20,22,34] Reduced questionnaire length has been associated with higher completion rates,[25] an important factor in obtaining high-quality data for value-driven health care.

As new instruments are developed, it is important to thoroughly evaluate their psychometric properties in defined patient populations and to assess the relative performance of new measures compared with commonly used legacy measures. This study evaluated the psychometric characteristics of the PROMIS PF CAT and the DASH in an upper-extremity orthopedic population with elbow conditions.

Materials and methods

Patients visiting a university orthopedic clinic for elbow conditions between February 2014 and March 2017 took the PF CAT and DASH on hand-held tablet computers (iPad; Apple, Inc., Cupertino, CA, USA) as part of their routine clinic care and before seeing the medical professionals. Data were collected using the mEVAL via the institution's electronic medical record system through a secure wireless network and stored in the institution's enterprise data warehouse. Standard CAT algorithms from the PROMIS Assessment Center were applied. All patients aged 18 years and older were included and consecutively enrolled based on the inclusion criteria of upper extremity elbow conditions. Non-English speakers were included if patients felt comfortable with translators going through the questionnaire side by side.

The DASH is a 30-item questionnaire. Its scores range from 0 to 100, with higher scores representing higher levels of functioning. The PROMIS PF CAT draws from a 124-item test bank, with an average test length of 6 to 7 items in an orthopedic population.[19] It has calibrated a T-score scale with mean of 50 and standard deviation (SD) of 10 in the general population.[32] T scores can help in standardizing and better representing the general population in a consistent way.[1] Higher scores on the PROMIS PF reflect higher levels of function. The instruments were administered at the first clinic visit or within the 7 days before the visit of a new patient with an upper extremity condition. Patients were emailed before their visit to complete the health assessment at home, and if not completed prior, they completed the assessment on the tablet at check-in.

Analyses

Psychometric evaluation of the PROMIS PF CAT and the DASH was completed using the Rasch partial credit model (PCM). This model is based on logistic latent trait item-response theory (IRT).[29] IRT poses many advantages for test development because it generates item-rich information while linking to the underlying latent trait.[8] Rasch modeling can determine the fit of individual items to the latent trait, and the PCM can model fit with polytomous items (items with more than 2 possible scores), because it does not require the response thresholds to be uniform across items. Rasch PCM has been used frequently in modern instrument development and evaluation.[15,21] The analyses in this study included assessments of (1) item fit, (2) internal reliability, (3) response category threshold, (4) dimensionality, (5) local independence, and (6) differential item functioning.

Item fit

The Rasch PCM is based on a form of Guttman scaling, where items and persons can be placed on the same linear continuum. In this analysis we first compared the response patterns in the observed data to the expected patterns determined by the model to see whether Rasch PCM is suitable for analysis of the data. The items were considered to fit the model if the observed responses were not significantly different from the expected responses. The outfit mean square (MNSQ) statistic was used here as an indicator of model fit. An outfit MNSQ of 1 represents a perfect fit, less than 1.4 is a good fit, and less than 2 is an acceptable fit.[4]

Internal reliability

Internal reliability is evaluated by the person separation index (PSI), which assesses how well an item distinguishes between groups of patients. Interpretation of the PSI is similar to the Cronbach α. The advantage of PSI over the Cronbach α is that the PSI has no upper limit to its value. A PSI of 0.8 or above is a satisfactory indicator of internal reliability.

Response category threshold

Item response categories should be adequate to capture the differences between responses. Otherwise, a different response format should be considered, such as collapsing the response categories. Checking whether the response thresholds (ie, the transition points