

Correspondence

A proposed set of metrics for standardized outcome reporting in the management of low back pain

Sir—I am writing in response to Clement's et al. article "A proposed set of metrics for standardized outcome reporting in the management of low back pain" published in *Acta Orthopaedica* (Clement et al. 2015). The purpose of the paper was to recommend a set of standard metrics for international use for every day clinical practice to measure outcomes for patients with diverse degenerative lumbar impairments.

The authors work is contemporary and reflects the need to respond to the rapid changes occurring in healthcare reimbursement with growing interest in value-based medicine (Porter and Teisberg 2004, Hart and Connolly 2006, Porter 2009). Value in healthcare was defined as achieving effective patient-reported outcomes efficiently (Porter and Teisberg 2004, Hart and Connolly 2006, Porter 2009). I congratulate the authors on their robust selection of risk-adjustment factors known to influence outcomes and their emphasis on the importance of risk adjustment throughout the patient's cycle of care in order to facilitate meaningful interpretation of outcomes and performance comparisons between providers managing patient with lumbar impairments. Risk adjustment will be essential for the successful implementation of alternative value based purchasing processes in healthcare to prevent denial of services to the most ill and disabled patients (Hart and Connolly 2006, Resnik et al. 2008a and b, Gozalo et al. 2015).

My concern is in regards to the authors' unanimous selection of the Oswestry Disability Index (ODI) as the go to metric for assessing the disability domain or the patient's physical function. I agree that the ODI is psychometrically sound and has been studied extensively in the literature but the authors appear to dismiss other published measures which meet their selection metric criteria. Outcomes in healthcare have evolved with patients and clinicians benefitting from administrative efficiency and measurement precision of computer adaptive tests (CATs) (Hart et al. 2010a). The authors recognized that CAT methods may offer a promising alternative metric choice for measuring patient's function yet cite only one CAT option i.e., the PROMIS tool. The authors reported that at this time CAT metrics were not ready for international use and lacked translation and validation beyond English.

I encourage the authors to expand their efforts going forward for implementing and revising their recommended outcome metric profile by taking a closer look at the abundance of CAT measures now available and commonly cited in the

literature (Hart and Connolly 2006, Deutscher et al. 2008, 2009, 2010, Hart et al. 2009, 2010a and b, 2012, Gozalo et al. 2015). For example, a recent paper published by Hart et al. (2012) conducted a head-to-head comparison between the ODI and a Lumbar CAT (LCAT) measure to assess patient-reported functional status (FS) outcomes. Their results indicated that both measures had similar psychometric characteristics. However, given the need to be efficient as well as precise in estimating measures of FS, results favor the LCAT because the time required to complete FOTO's LCAT is < 2 minutes compared to the increased time (5+ minutes) required for patients to complete and clinicians to score the ODI.

The primary aim, in light of today's rapid changes in the healthcare environment, for creating standard metrics is to promote widespread implementation of psychometrically sound measures to ultimately demonstrate value or quality of care provided to our patients (Porter and Teisberg 2004, Hart and Connolly 2006, Porter 2009). In order to boost implementation during routine clinical practice, data collection process must not only consist of psychometrically strong measures but also be simplified by reducing burden for both the patient and the practitioner. Reduced patient burden, i.e., saving time, during fast-paced and often hectic outpatient environments, will be an essential step in order to encourage provider compliance for collecting data at the patient bedside (Hart and Connolly 2006, Hart et al. 2010a)

Of interest, CAT functional status measures are now commonplace during routine clinical practice throughout the US, Maccabi Healthcare Services, the 2nd largest health provider in Israel, and Canada. Recent published results supported the validity of translation of knee CAT FS items into Hebrew and Russian (Hart et al. 2009, Deutscher et al. 2010, 2011) and validation in Spanish and French are planned.

The science to date and common clinical use supporting CAT measures should be considered and CAT measures should be included as part of any standardized metric registry recommended for assessing patient self-report functional status outcomes during routine practice.

Mark Werneke

CentraState Medical Center, Spine Rehabilitation
901 West Main Street, Freehold, NJ 07728 USA
Email: mwerneke@centrastate.com

Sir—Mr. Werneke has emphasized two important points in his letter: 1) computerized adaptive testing (CAT) tools can facilitate outcome measurements by expediting data collection, and 2) there have been meaningful strides in recent years in the development of these tools for the evaluation of patients with spine pathology. As described in our original article, the intent of the ICHOM Working Group was to harmonize instruments across national and cultural borders in the measurement of outcomes for low back pain (LBP) in order to facilitate international comparisons, both for research and daily clinical practice. Recommending a single instrument per domain necessarily involved trade-offs, as no one instrument was found to be superior to all others on all accounts.

To measure lumbar-related disability (a.k.a. physical function), our Working Group recommended the Oswestry Disability Index (ODI) as the currently preferred tool for several reasons. First, the ODI was considered the most well-known and widely used instrument in the field of LBP. Second, it has been extensively tested and has proven psychometric properties, including validation in 35 languages and cultures (www.proqolid.org/instruments/oswestry_disability_index_odi), more than any other measurement tool for this domain (Chapman et al. 2011). Third, the ODI has been used in numerous studies in recent decades, making relevant historical comparisons relatively easy. Fourth, the time required to complete the ODI, while greater than that associated with CAT tools, was found to be acceptable.

The literature cited by Mr. Werneke describes a lumbar CAT tool (LCAT) to measure physical function that has been designed in conjunction with Focus on Therapeutic Outcomes (FOTO), Inc. (Knoxville, TN, USA), a web-based patient assessment and data management company with a focus on rehabilitation. This tool has been rigorously studied among physical therapy patients by Mr. Werneke's research group over the last ten years; the associated psychometric properties are sound, and it performs on par with a modified version of the ODI. (Hart et al. 2010a, 2010b, 2012). However, it should be noted that this modified version included unvalidated linguistic changes, a copyright violation was eventually conceded, and the term "Oswestry" was retroactively removed from the name of this modified tool (Fritz and Irrgang 2001, <http://ptjournal.apta.org/content/81/2/776/suppl/DC1>). This technicality is a reminder of the careful and painstaking work required in the development and validation of appropriate outcome measurement tools.

Impressively, Mr. Werneke's group has repeatedly described successful implementation of CAT tools with reproducible efficiency gains (Deutscher et al. 2008, 2009, Hart et al. 2010a, Gozalo et al. 2015). They have also described valid results in two languages: English and Hebrew (Hart et al. 2010a). Furthermore, while not directly applicable here, their group should be applauded for their efforts validating and implementing similar CAT's for other anatomic regions, including the knee, shoulder, and cervical

spine (Deutscher et al. 2008, 2009, 2010, Hart et al. 2009, Gozalo et al. 2015).

We recognize that newer tools, including CAT's like the one described by Mr. Werneke, may better match the needs of clinicians in daily practice, and we have assembled a steering committee to monitor and refine our proposed LBP outcome set as evidence mounts concerning their psychometric properties, implementation feasibility, and cross-language availability. To maintain harmonization over time while supporting the development of new instruments, we encourage (and intend to work with) instrument developers to prepare cross-walk algorithms that allow new measurement tools to maintain backwards compatibility with existing, more established instruments, as has been demonstrated successfully in instruments for mental health assessments (Wahl et al. 2014).

In review of existing literature, including the articles cited in Mr. Werneke's letter, we must conclude that none of the currently available CAT tools are yet ready for inclusion in this outcome set. As FOTO's LCAT continues to develop, it may well be translated into numerous languages with proven psychometric properties. Similarly, our original article referred to the Patient Reported Outcome Measurement Information System (PROMIS) instrument for measuring physical function, which is available in CAT form and demonstrates potential for future inclusion in the proposed outcome set. For both FOTO's LCAT and the PROMIS physical function tool, we would seek additional language coverage with appropriate cross-cultural validation as well as mechanisms for backwards compatibility with the ODI to allow historical comparisons before recommending them as suitable alternatives for international implementation.

Carter Clement

Department of Orthopaedic Surgery, University of North Carolina Hospitals, Chapel Hill, NC, USA
Email: carter.clement@gmail.com

Chapman J R, Norvell D C, Hermsmeyer J T, Bransford R J, DeVine J, McGirt M J, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine* 2011; 36(21 Suppl): S54–68.

Clement R C, Welander A, Stowell C, Cha T D, Chen J L, Davies M, Fairbank J C, Foley K T, Gehrchen M, Hagg O, Jacobs W C, Kahler R, Khan SN, Lieberman I H, Morisson B, Ohnmeiss D D, Peul W C, Shonnard N H, Smuck M W, Solberg T K, Stromqvist B H, Hooff M L, Wasan A D, Willem P C, Yeo W, FRITZ P. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop* 2015; 86(5): 523–33

Deutscher D, Hart D L, Horn S D, Dickstein R, Gutvirtz M. Implementing an integrated electronic outcomes and electronic health record process to create a foundation for clinical practice improvement. *Phys Ther* 2008; 88(2): 270–85.

Deutscher D, Horn S D, Dickstein R, Hart DL, Smout R J, Gutvirtz M, Ariel I. Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. *Arch Phys Med Rehabil* 2009; 90(8): 1349–63.

- Deutscher D, Hart D L, Crane P K, Dickstein R. Cross-cultural differences in knee functional status outcomes in a polyglot society represented true disparities not biased by differential item functioning. *Phys Ther* 2010; 90(12): 1730-42.
- Fritz J M, Irrgang J J. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther* 2001; 81(2): 776–88.
- Gozalo P, Resnik R, Silver B. Benchmarking outpatient rehabilitation clinics using functional status outcomes. *Health Serv Res* 2015 Aug 6. doi: 10.1111/1475-6773.12344. [Epub ahead of print]
- Hart D L, Connolly J B. Pay-for-Performance for Physical Therapy and Occupational Therapy: Medicare Part B Services. Final Report. Grant #18-P-93066/9-01: Health & Human Services/Centers for Medicare & Medicaid Services; 2006. <http://www.cms.hhs.gov/TherapyServices/downloads/P4PFinalReport06-01-06.pdf>.
- Hart D L, Deutscher D, Crane P K, Wang Y-C. Differential item functioning was negligible in an adaptive test of functional status for patients with knee impairments who spoke English or Hebrew. *Qual Life Res* 2009; 18(8): 1067-83.
- Hart D L, Deutscher D, Werneke M W, Holder J, Wang Y-C. Implementing computerized adaptive tests in routine clinical practice: experience implementing CATs. *J Appl Meas* 2010a; 11(3): 288-303.
- Hart D L, Werneke M W, Wang Y C, Stratford P W, Mioduski J E. Computerized adaptive test for patients with lumbar spine impairments produced valid and responsive measures of function. *Spine* 2010b; 35(24): 2157-2164.
- Hart D L, Stratford P W, Werneke M W, Deutscher D, Wang Y-C. Lumbar computerized adaptive test and modified Oswestry Low Back Pain Disability Questionnaire: relative validity and important change. *J Orthop Sports Phys Ther* 2012; 42(6): 541-51.
- Porter M E. A strategy for healthcare reform – toward a value-based system. *N Engl J Med* 2009; 361: 109-112.
- Porter M E, Teisberg E O. Redefining healthcare. *Harv Bus Rev* 2004; 82: 64-76.
- PROQOLID, the Clinical Outcome Assessment (COA) Instruments Database - Oswestry Disability Index (ODI) / Instruments / Home - eZ publish [Internet]. [cited 2015 Nov 3]. Available from: http://www.proqolid.org/instruments/oswestry_disability_index_odi
- Resnik L, Liu D, Mor V, Hart DL. Predictors of physical therapy clinic performance in the treatment of patients with low back pain syndromes. *Phys Ther* 2008a; 88(9): 989-1004
- Resnik L, Liu D, Hart DL, Mor V. Benchmarking physical therapy clinic performance: statistical methods to enhance internal validity when using observational data. *Phys Ther* 2008b; 88(9): 1078-1087.
- Wahl I, Löwe B, Bjorner J B, Fischer F, Langs G, Voderholzer U, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol* 2014; 67(1): 73–86.