

EDITORIAL

From the item to the outcome: the promising prospects of PROMIS

Daniel Aletaha*

See related research by Bruce *et al.*, <http://arthritis-research.com/content/11/6/R191>

Abstract

Evaluation of patient reported outcomes, and in particular physical function, have gained increasing importance in research and therapy of patients with rheumatic diseases. Most instruments that are used for that purpose are rigid and suffer from floor and ceiling effects when used in patients whose physical function differs from the average. A new approach to the assessment of physical function uses computerised adaptive testing, by which precision and reliability of the measurement can be achieved for most patients, while even requiring less time for the assessment. Well calibrated and tested item and large item data banks are a prerequisite for this purpose, a process that is summarised in the present report by Bruce and colleagues.

Bonnie Bruce and colleagues [1] report on the development process of the PROMIS item data bank. In their article the authors describe the stepwise process by which they systematically searched for items reflecting physical function, and then refined and evaluated a subset of these items for further application in the functional assessment of patients with rheumatic diseases by new instruments or by computerised adaptive testing (CAT).

Many patient reported outcomes that focus on functional capacity have been published for use in rheumatic diseases, and rheumatoid arthritis (RA) usually serves as a paradigmatic example of such chronic disease [2]. Physical function has always been a core outcome in RA, even surpassing more 'objective' outcomes, such as radiographically observed damage. The measurement of physical function has been revolutionised by the

development of the Health Assessment Questionnaire (HAQ) Disability Index (HAQ-DI), which was introduced by Fries and colleagues in the early 1980s [3]. The original HAQ or its modifications are by now the most commonly used functional measures reported in the various clinical databases and in clinical trials of RA. However, they might not be appropriate for all patients.

In their current report, Bruce and colleagues [1] elegantly show that attempts to improve functional assessment must start at the level of the individual items that an outcome measure will eventually be composed of. They conclude that items work better if their wording is in the present tense rather than the past tense, if they focus on the ability to do activities rather than actual performance, if they are simple, and if they have four to five response options rather than only two to three.

The most promising application of such items is in CAT. For this purpose, items that are well calibrated on a respective metric (for example, physical function) are needed. The great advantage of CAT in medical assessment in general, and functional assessment in rheumatic disease in particular, is its ability to provide uniformly precise scores for most patients [4]. The computer usually selects a starting item (question) that is of average difficulty. Based on the patient's response, the computer will update its estimation of, for example, the patient's functional ability and accordingly select an easier or harder question. This will be repeated until some level of precision is achieved (for example, by the width of the standard deviation of the estimate or similar), that is, when a termination criterion is reached. This is a major advantage over standard fixed tests, such as the HAQ, which are usually most precise for patients of average functional capacity, and decreasingly less precise for those with more extreme test scores. Thus, CAT can diminish the floor and ceiling effects seen in standard instruments, as the authors also discuss.

Is this relevant for functional assessments in RA? Yes, in terms of the expected trajectories of the patients' disease activity and function; in other words, patients with RA experience great changes in disease activity and function given the new treatment strategies and new and

*Correspondence: daniel.aletaha@medunwien.ac.at
Division of Rheumatology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria

effective drug regimens [5]. They are more than ever prone to make transitions across large ranges of the latent metric of functional capacity. Patients and investigators/rheumatologists will therefore benefit from instruments that provide comparably precise estimates at the beginning of therapy and during follow-up.

Likewise, it will also be relevant regarding several aspects of physical function that have enjoyed much less attention in the past. These include the question of reversibility of functional limitations - that is, the highly pertinent question of 'What is the best possible functional capacity in an individual patient?' - as well as the recently estimated translation of joint damage (by radiographic evaluation) to functional disability [6]. More precise measures will also allow even more accurate functional outcomes research.

Will it work in the respective relevant age group of the middle-aged to the elderly? Yes, probably also in terms of the ability of the elderly population to follow and perform a computer test, especially when they receive a training session in advance [7]. Even to the contrary, elderly patients in particular are often overwhelmed with the increasing number of surveys and questionnaires, and an assessment approach that employs CAT will save their time and energy by providing similar or better precision with fewer test items. As the authors say, with regard to educational testing this aspect of CAT is especially important.

The downsides of CAT are the considerably higher costs of testing and calibrating the items in large numbers of patients. This has already been partly achieved in the present study, even if there is still a need to expand it to a wider range of rheumatic diseases. However, some remaining issues will need to be addressed, including the fact that numerous new tools can now be developed by individual researchers who wish to conduct a specific study [8]. It still needs to be communicated how we can interpret 'custom-made' adaptive functional scales with, for example, traditional HAQ scores as we know them. As the literature is filled with scores of standard HAQ or

SF-36 surveys, it might be wise to encourage investigators to collect data on the traditional tools in addition to the new scales. This will facilitate the interpretation of the results. It is likely, however, that the CAT results will not be very different from traditional functional scales, but that their standard deviations will be smaller. This is relevant, as functional outcomes - and patient reported outcomes in general - will be increasingly used as endpoints in clinical trials. The increased statistical power of CAT analyses will decrease sample size requirements for studies, and accordingly, smaller number of patients will be put on placebo (or comparator therapy). Finally, such studies will be much less costly.

Abbreviations

CAT = computerised adaptive testing; HAQ = Health Assessment Questionnaire; RA = rheumatoid arthritis.

Competing interests

The author declares that they he has no competing interests.

Published: 1 February 2010

References

1. Bruce B, Fries JF, Ambrosini D, Lingala B, Gandek B, Rose M, Ware JE Jr: **Better assessment of physical function: item improvement is neglected but essential.** *Arthritis Res Ther* 2009, **11**:R191.
2. Kalyoncu U, Dougados M, Daures JP, Gossec L: **Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review.** *Ann Rheum Dis* 2009, **68**:183-190.
3. Fries JF, Spitz P, Kraines RG, Holman HR: **Measurement of patient outcome in arthritis.** *Arthritis Rheum* 1980, **23**:137-145.
4. Thissen D, Mislevy R: **Testing algorithms.** In *Computerized Adaptive Testing: A Primer*. 2nd edn. Edited by Wainer H. Mahwah, NJ: Lawrence Erlbaum Associates; 2000:101-135.
5. Smolen JS, Aletaha D, Koeller M, Weisman MH, Emery P: **New therapies for treatment of rheumatoid arthritis.** *Lancet* 2007, **370**:1861-1874.
6. Smolen JS, Aletaha D, Grisar JC, Stamm TA, Sharp JT: **Estimation of a numerical value for joint damage-related physical disability in rheumatoid arthritis clinical trials.** *Ann Rheum Dis* 2009 [Epub ahead of print].
7. Johnsin D, White C: **Effects of training on computerized test performance in the elderly.** *J Appl Psychol* 1980, **65**:357-358.
8. PROMIS [<http://www.nihpromis.org/default.aspx>]04

doi:10.1186/ar2910

Cite this article as: Aletaha D: From the item to the outcome: the promising prospects of PROMIS. *Arthritis Research & Therapy* 2010, **12**:104.