



Item banking and computer-adaptive testing in clinical trials: Standing in sight of the PROMISed land

Adam B. Smith*, Andria Hanbury, Jenny Retzler

York Health Economics Consortium, Enterprise House, University of York, Hesling, York, YO105NQ, UK

ABSTRACT

The use of patient-reported outcome instruments (PRO) in clinical trials in order to capture the impacts of treatment on patients is widespread. However, regulatory agencies have over the past decade highlighted the need for PROs that are fit for purpose and target relevant aspects of the patient's condition. Many legacy PROs were developed with little patient input, are lengthy, and may lack relevance having not been modified or adapted as medical treatments have advanced. Computer-adaptive test (CAT) systems provide the possibility of targeted approaches to capturing patient-centric data, while minimising patient burden. Coupled with greater patient input in the development of PROs, CAT offers the opportunity of overcoming the shortcomings of the previous generation of PROs. This paper describes the some of the issues facing legacy PROs, current regulatory guidance, and initiatives, such as the Patient-Reported Outcome Measurement Information System (PROMIS), as well as the early signs of use of CAT to capture PRO data in clinical trials.

1. Introduction

Patients, their concerns and experiences of treatment and beyond are taking centre stage in clinical trials with patient-centricity increasingly forming the focus for many trial sponsors [1]. The common currency for capturing the views of the patient (concerning their symptoms, functioning, quality of life) is patient-reported outcomes (PROs). PROs are provided directly by patients with no mediation from others, such as family, friends and medical practitioners [2].

These PRO instruments used to collect patient-centred data are standardised, validated tools that have been demonstrated to reliably capture patient responses and changes in health status/symptoms/quality of life consistently over time.

Many PROs have been used in clinical trials for numbers of years, however, developing guidance from healthcare regulatory agencies [2,3] has placed a higher hurdle on criteria PROs are required to fulfil in order to support product claims. This guidance has placed a significant emphasis on patient involvement in the early development of PRO and has highlighted the potential shortcomings of “legacy” instruments – those PROs developed prior to the regulatory guidance, many of which were created with little input from patients.

While the intention was to enhance the ability of PROs to capture patient-centric data the guidance has had some unintended consequences. For instance, there have been fewer PRO-based labelling claims granted in recent years [4] with issues around the content validity (i.e. whether a PRO measure all the major aspects relevant to the patient's condition) being cited as one of the major reasons for claims

being rejected [5].

Although it may be argued that the frequent and widespread use of legacy PROs is, in itself, evidence of instruments being fit for purpose [6], it is clear that with ever faster moving advances in medicines and treatments, and within an era of personalised medicine, new PROs are required that make use of new approaches and developments in technology.

This idea is by no means a new one, earlier commentators [7] foresaw a future where PROs in clinical trials would be based on or drawn from item banks – large sets of individual questions covering the breadth and depth of symptom domains far beyond legacy instruments and where computer-adaptive data collection (i.e. systems that present questions from PROs contingent on previous patient responses) would enable PROs to be tailored specifically to the individual patient.

These computer-adaptive tests (CAT) would be shorter, better-targeted, and consequently more efficient instruments for obtaining patient-centric data. These CAT systems are based on statistical techniques known collectively as item-response theory (IRT) models. These models differ from traditional approaches to PRO development by focussing on individual items in the measurement of latent traits, rather than the PRO in its entirety.

Latent traits are unobservable constructs, such as pain, that may be viewed as a spectrum. PRO responses by patients aim to measure where on the spectrum a particular patient falls with respect to a latent trait, and some traits may be measured using multiple complementary items. The focus on items used in approaches such as IRT modelling lends itself to a number of potential benefits, allowing us to pinpoint where

* Corresponding author.

E-mail address: adam.smith@york.ac.uk (A.B. Smith).

<https://doi.org/10.1016/j.conctc.2018.11.005>

Received 8 August 2018; Received in revised form 7 November 2018; Accepted 9 November 2018

Available online 17 November 2018

2451-8654/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the item falls along the latent trait in relation to others, thus enabling PRO instruments to be developed that cover the full range of latent trait using the minimal number of items.

What IRT models allow then is to determine where items fall along this latent trait. This means for instance, that items from existing PROs can be assessed using IRT to determine where they fall along the latent trait, additional items can be added from other measures or through further input from patients and experts to fill gaps and ensure that items cover the entire latent trait, and in such a way large items banks are created.

Once these items banks have been created where the location of each individual item is known along the latent trait, PROs can be developed into multiple equivalents (parallel) fixed or static versions (i.e. the same combination of items is presented to patients) or into computer-adaptive tests (CAT), where the next item to be drawn from the item bank and presented is dependent on the patient's previous response.

PROs and CAT systems developed in this way using IRT models are estimated to at least halve the number of items required to determine patients' point or "score" along the latent trait [8]. Typically, the static versions of PROs developed based on item banks only comprise 4 to 10 items (contrast this, as an extreme example with, for instance, legacy instruments such as the Sickness Impact Profile [9] consisting of more than 100 items).

As these instruments cover the whole spectrum of the latent trait researchers can avoid overestimating (floor effects) and underestimating (ceiling effects) patients' health status, while reducing questionnaire burden (through fewer, but more relevant items) and increasing specificity (health domains tailored to individual patients).

The Patient-reported Outcome Measurement Information System (PROMIS) was established in 2004 with funding from US National Institutes of Health (NIH) [10–12]. The aim of this initiative is to develop reliable, valid PROs with increased precision, using IRT to overcome the shortcomings of legacy instruments.

The PROMIS instrument development process includes use of focus groups and cognitive (in-depth) interviews with patients, community samples and experts to ensure that the items cover the area or concept of interest in sufficient depth and breadth [(www.healthmeasures.net)]. This means that regulatory criteria for content validity are met and ensures that along with the quantitative ("psychometric") validation that the instruments are fit for purpose.

The PROMIS initiative to-date has primarily, but not exclusively, focussed on domains, rather than specifically on disease or health conditions. The approach by PROMIS then, is a combination of generic set of tools, not specific to any particular medical condition, but falling within the broad themes of Physical, Mental and Social Health. For instance, the PRO domains for adult respondents are specific to functions, behaviours, emotion and symptoms. However, there are sets of instruments with the PROMIS system that are disease-specific. For instance, PROMIS-Cancer incorporate items for Physical Activity, Fatigue and Anxiety associated with cancer. There are additional suites of items and instruments, including the Neuro-QOL, assessing quality-of-life in neurological disorders, and ASCQ-Me, designed for use with adults with sickle cell disease.

In general, therefore, PROMIS is a move away from disease-specific PROs towards more generic – but domain-specific PROs. Item banks have been developed, validated and calibrated for these domains, typically comprising fewer than 100 items for each domain for use in computer-adaptive tests, as well as static versions of 4, 5 and 6 items that have been selected from the item banks to provide the greatest range of coverage of the latent trait with the fewest items. The PROMIS system utilises standardised scores, i.e. the T-score metric (mean of 50, and standard deviation of 10), which has been referenced against large sample of the general population in the United States. The system also comprises guidelines on meaningful change, all of which facilitates interpretation and comparisons across patient groups.

In terms of the current state of affairs regarding the use of PROMIS within clinical trials, a search on [ClinicalTrials.gov](https://clinicaltrials.gov) (accessed July 2018) showed just under 900 registered trials including the PROMIS system with 27 trials recorded as including CAT versions of the system. Similarly, the number of published studies with computer-adaptive testing is in excess of 100. Many of these studies include the (further) development and validation of the PROMIS system, as well as other PRO instruments, and cover a wide range of medical conditions and surgical interventions. However, less than a handful of these published studies involved randomized controlled trial (RCT) designs [13,14].

While it may seem that there is still some way to go before computer-adaptive systems are fully incorporated within RCTs, the number of upcoming and ongoing trials ([ClinicalTrials.gov](https://clinicaltrials.gov)) suggests that the tide may be turning towards more CAT in clinical trials. Recognition of the PROMIS system and the various associated instruments by the FDA will certainly facilitate and potentially expedite further adoption in clinical trials. An example of the latter is the fact that the PROMIS Physical Function scale has been included in the FDA Clinical Outcomes Assessment (COA) Compendium¹ – a list of those PRO instruments that have been used in clinical trials and may be used to potentially support labelling claims – for use in patients with sarcopenia, as well as in oncology.

The promise that computer-adaptive testing potentially holds in terms of providing shorter, more accurate and relevant computer-adaptive tests is, therefore, in sight.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.conctc.2018.11.005>.

References

- [1] G. Yeoman, P. Furlong, M. Seres, H. Binder, H. Chung, V. Garzya, R.R.M. Jones, Defining patient centrality with patients for patients and caregivers: a collaborative endeavour, *BMJ Innov.* 3 (2017) 76–83.
- [2] Food and Drug Administration, Guidance for Industry. Patient-reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims, (2009) <https://www.fda.gov/downloads/drugs/guidances/UCM193282.pdf>, Accessed date: 22 July 2018.
- [3] European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP): Reflection Paper on the Regulatory Guidance for the Use of Health-related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products, (2005) http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf, Accessed date: 27 July 2018.
- [4] A. Gnanasakthy, M. Mordin, E. Evans, L. Doward, C. DeMuro, A review of patient-reported outcome labeling in the United States (2011–2015), *Value Health* 20 (3) (2017) 420–429.
- [5] C. DeMuro, M. Clark, M. Mordin, S. Fehnel, C. Copley-Merriman, A. Gnanasakthy, Reasons for rejection of patient-reported outcome label claims: a compilation based on a review of patient-reported outcome use among new molecular entities and biologic license applications, 2006–2010, *Value Health* 15 (3) (2012) 443–448.
- [6] A.B. Smith, K. Cocks, Content validity and legacy patient-reported outcome measures in cancer, *Qual. Life Res.* 24 (7) (2015) 1585–1586.
- [7] D.A. Revicki, D.F. Cella, Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing, *Qual. Life Res.* 6 (6) (1997) 595–600.
- [8] J.E. Ware Jr., M. Kosinski, J.B. Bjorner, M.S. Bayliss, A. Batenhorst, C.G. Dahlöf, S. Tepper, A. Dowson, Applications of computerized adaptive testing (CAT) to the assessment of headache impact, *Qual. Life Res.* 12 (8) (2003) 935–952.
- [9] M. Bergner, R.A. Bobbitt, W.B. Carter, B.S. Gilson, The Sickness Impact Profile: development and final revision of a health status measure, *Med. Care* 19 (8) (1981) 787–805.
- [10] J. Alonso, S. Bartlett, M. Rose, N. Aaronson, J. Chaplin, F. Efficace, A. Lepège, A. Lu, D.S. Tulsky, H. Raat, U. Ravens-Sieberer, D. Revicki, C.B. Terwee, J.M. Valderas, D. Cella, C. Forrest, The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative, *Health Qual. Life Outcome* 11 (1) (2013) 210.
- [11] J.F. Fries, B. Bruce, D. Cella, The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes, *Clin. Exp. Rheumatol.* 23 (5 Suppl 39) (2005) S53–S57.

¹ <https://www.fda.gov/drugs/developmentapprovalprocess/developmentresources/ucm459231.htm>.

- [12] D. Cella, S. Yount, N. Rothrock, R. Gershon, K. Cook, B. Reeve, D. Ader, J.F. Fries, B. Bruce, M. Rose, PROMIS Cooperative Group, The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap co-operative group during its first two years, *Med. Care* 45 (5 Suppl 1) (2007 May) S3–S11.
- [13] M.H.F. Keulen, T. Teunis, G.A. Vagner, D. Ring, L.M. Reichel, The effect of the content of patient-reported outcome measures on patient perceived empathy and satisfaction: a randomized controlled trial, *J. Hand Surg. Am.* 43 (12) (2018), <https://doi.org/10.1016/j.jhssa.2018.04.020> (epub ahead of publication).
- [14] J.J. Mellema, C.M. O'Connor, C.L. Overbeek, M.G. Hageman, D. Ring, The effect of feedback regarding coping strategies and illness behavior on hand surgery patient satisfaction and communication: a randomized controlled trial, *Hand (N Y)*. 10 (3) (2015 Sep) 503–511.