International Journal of Clinical Trials
Shalhoub H et al. Int J Clin Trials. 2016 Nov;3(4):174-179
http://www.ijclinicaltrials.com
pISSN 2349-3240 | eISSN 2349-3259
DOI: http://dx.doi.org/10.18203/2349-3259.ijct20163953

Review Article

PROMIS® tools as endpoints in clinical trials: what should you know? A review of PROMIS® capabilities and the current regulatory space

Huda Shalhoub*, Matthew Reaney

ERT, 500 Rutherford Ave, Boston, MA, USA

Received: 30 August 2016
Revised: 01 September 2016
Accepted: 29 September 2016

***Correspondence:**
Dr. Huda Shalhoub,
E-mail: huda.shalhoub@ert.com

ABSTRACT

NIH funded PROMIS measures are increasingly at the forefront of discussions in clinical trial endpoint measurement. In the US alone, 242 clinical trials have included PROMIS measures 2016. Regulatory agencies also recently appear to have more interest in the applications and interpretations of PROMIS tools. This paper provides an overview of the PROMIS tools, how and when they can be used, how they are scored, what modalities exist and what considerations one should make before choosing to use a PROMIS instrument. PROMIS spans across a wide range of areas ranging from mental, physical and social health status, from short form to profile instruments, and to electronic IRT scoring based methods. Psychometric properties measurement testing in a few therapeutic areas has also been evident and continues to develop. The regulatory agency and governmental bodies continue to focus their efforts on having a more profound understanding of the application and use of this patient reported toolbox. PROMIS measures are promising for use within the context of clinical trials, but stakeholders should prudently consider their use, thinking about both the pros and cons. It is likely that for endpoint measurement, PROMIS tools may be used on a case by case basis, but that a consideration of additional disease specific instruments may be recommended.

Keywords: Endpoint, PROMIS, Item banks, Patient reported outcomes, Regulatory, Clinical trial

INTRODUCTION

The patient-reported outcomes measurement information system (PROMIS) network was established in 2004 based on an initiative funded by the US National Institutes for Health (NIH).[1] The PROMIS network aims to use "measurement science to create a state-of-the-art assessment system for self-reported health".[2] Under this remit, the PROMIS network have developed item pools to evaluate patient-reported feelings, functions, and perceptions of various aspects of health and healthcare to be used in a multitude of contexts, including clinical trials, academia and healthcare settings. From these item pools, the PROMIS network created a series of domain-specific "short-form" instruments quantifying physical, mental, and/or social well-being within a specified conceptual dimension. These dimensions are collectively called the PROMIS instruments. These instruments are freely available online and accessible to clinicians, scientists, and researchers to use through the website healthmeasures.net. Although the focus of this review is on the PROMIS initiative's item pools, the healthmeasures website includes not only PROMIS, but incorporates other instruments and items for use from Neuro-Qol, ASCO-Me and the NIH toolbox.

Since their inception, PROMIS instruments have been used in all types of settings including clinical, healthcare, and academic settings. For instance, a clinicaltrials.gov (US based) search conducted by the authors that took

place in July 2015 found 147 studies associated with the use, calibration, or validation of a PROMIS instrument across many conditions (these include a variety of cancers, low back pain, chronic pain, parkinson's disease, multiple sclerosis, childhood asthma among many other therapeutic areas and conditions). The same search was repeated in July 2016 and a total output of 242 clinical trials appeared, indicating an increase in the use of PROMIS tools over the last few years. The library of PROMIS instruments can be administered in three different ways:

As short-form instruments (SF)

Mainly developed in 2007 and 2008, short-form instruments measure single domains as shown in Table 1 including mental, physical and social health with lengths ranging from 4 to 10 items, with the exception of the physical function SF that contains up to 20 items. Since their creation, however, multiple newer versions of the short-forms have been formed. New short-forms can also be developed through selecting items from the relevant item bank. This allows for tailoring content to a specific population to achieve greater precision or to address specific research questions based on one's endpoint. To exemplify this, a target population that experiences higher rates of fatigue may be assessed with a more tailored and targeted fatigue SF than those which have been created by the PROMIS group (4, 6, 7 or 8 items). Additionally, with short-forms PROMIS calls for participants to be instructed to respond to all questions (no skipping allowed).

Table 1: PROMIS domains available in 2016.

	Mental Health	Physical Health	Social Health
Profile domains	Depression (A); Depressive symptoms (P)	Physical function (A)	Ability to participate in social roles and activities (A)
	Anxiety (A,P)	Pain intensity (A,P)	Peer relationships (P)
		Pain interference (A,P)	
		Fatigue (A,P)	
		Sleep disturbance (A)	
		Mobility (P)	
		Upper extremity function (P)	
Additional domains	Anger (A,P)	Pain behavior (A,P)	Satisfaction with social roles and activities (A)
	Cognitive function (A,P)	Pain quality (A,P)	Social support (A)
	Alcohol use, consequences and expectancies (A)	Sleep-related impairment (A)	Social isolation (A)
	Smoking (A)	Sexual function (A)	Companionship (A)
	Substance use (A)	Gastro-intestinal symptoms (A)	Family relationships (P)
	Psychosocial illness impact (A)	Dyspnea (A)	
		Asthma impact (P)	
	Self-efficacy (A)	Physical activity (P)	
	Life satisfaction (P)	Physical stress experiences (P)	
	Meaning and Purpose (P)	Strength impact (P)	
	Positive affect (P)		
	Psychological Stress Experiences (P)		

Note: There are also other (including global) and multiple domains available but the above listing forms the majority; A = Adult; P = Paediatric.

As profile instruments

Fixed collections of short-form instruments that measure multiple domains and are universal (i.e. not specific to a disease or therapeutic area) have been developed. These are referred to as PROMIS profiles. There are currently 3 PROMIS profile instruments for adults named PROMIS-29, PROMIS-43, and PROMIS-57. Each profile includes short-forms assessing anxiety, depression, fatigue, pain intensity and interference, physical function, sleep disturbance and ability to participate in social roles/activities; the difference lies in whether they use the 4-item, 6-item or 8-item short form measures of each domain. For pediatric populations, there are PROMIS-25, PROMIS-37, and PROMIS-49, each assessing anxiety, depressive symptoms, fatigue, pain intensity and interference, mobility and peer relationships. They can either be self-completed (ages 8-17) or completed by a

parental proxy (ages 5-17). Distinct from the Profiles, there is also a PROMIS global instrument which serves as a brief (10 item adult, 7/9 item pediatric) measure of physical and mental health.

Using the full item pools through computer adaptive tests (CAT)

Unlike the short-form instruments and profiles, a CAT approach to measurement is variable and adaptive as everyone is tested differently based on their individual responses to items in the item bank. Specifically, CAT uses a targeted algorithm to determine which items from a given item pool should be asked to a participant at time n. The algorithm is based on the participants' responses to questions at time n-1, and the order of items is algorithmically assigned. The number of items used to assess a domain varies, but is usually at least 4-5. The CAT program has "stopping rules" for determining when to stop administering items. CAT therefore, is argued to provide higher precision in the response than the short-form or profile instruments.

Table 2: PROMIS considerations.

Considerations before selecting a PROMIS instrument
How old is your respondent? If a child, do you want them to self report or have a proxy do so?
What domains are relevant to the condition you are testing?
What mode of data collection is preferred? Paper, web, mobile device?
How much time is available to complete an assessment?
What level of precision is necessary? Primary outcome? Secondary?
Do you want all respondents to answer the same items? (e.g., short form) Are there particular items you want included? (e.g., custom short form)
Is there sufficient validation information for the measure?
Are there other symptoms or domains outside of PROMIS that are important to include?

MODALITIES

PROMIS provides its users the opportunity to administer their short-form instruments and Profiles using paper, online data collection software applications (e.g., Assessment Center, REDCap), or electronic medical record systems (e.g., Epic, OBERD). In general, electronic or web data-collection methods are well established for use in clinical trials, with recent meta-analyses supporting their equivalence to paper formats where a process of faithful migration has been followed.[3,4] For PROMIS, the different versions were developed simultaneously, and research supports equivalence between the modes of administration.[5-7] The use of smartphones and similar technologies are becoming more ubiquitous around the world and there is an increasing interest in using app-based technology to collect patient-reported data in clinical trials. Although PROMIS has developed some electronic applications for some short-form instruments (e.g. NIH Toolbox app), there is little published research to date that examines equivalence between different electronic modalities by screen size or mobile application type, and thus this bring your own device (BYOD) model has not yet been fully endorsed.

SCORING

Item response theory (IRT), is the mode for scoring, the PROMIS SFs, Profiles and CAT administrations with response pattern scoring preferred (vs. raw score lookup tables). The preferred way to score short-forms is within the assessment center scoring service website database where services can be provided free by an expert.[8] The total raw score of SF items is generally the sum of the values if the responses per domain. After the raw sum is calculated, the sum is divided by the number of items that were answered. Once that number is calculated, there is a conversion table that translates the raw scores into a T-score for each short-form. Standard errors (SE) are also provided in the same lookup table.

In terms of the CAT-based scoring, where a varying number of items from an item pool are administered between and within participants over time based on relevance, the participant's score is also calibrated to a T-score at each time point, allowing for a standardized metric for comparisons.

In addition to being scored as stand-alone instruments, algorithms have been developed to map PROMIS instrument scores to many commonly used, or "legacy," patient reported outcomes (PRO) instruments, including the Medical Outcome Survey 36-item Short Form (SF-36), Brief Pain Inventory (BPI), Functional Assessment of Chronic Illness Therapy (FACIT) questionnaires, and the 9-item Patient Health Questionnaire (PHQ)-9. This initiative, which is called PROsetta Stone also uses T-scores to facilitate comparability of PRO instruments through standardized metrics. Other NIH-funded measurement initiatives including Neuro-QOL and the NIH toolbox are also included on the PROsetta Stone website.[9] All linking tables are available to the public for free. The PROsetta Stone website additionally assists in explaining the crosswalking methodology.

PSYCHOMETRIC SUPPORT FOR USE IN THERAPEUTIC AREAS

Validation studies have been undertaken to demonstrate that some of the PROMIS short-form instruments are well-defined and reliable (i.e. "validated") in certain contexts of use. For example, the PROMIS depression item bank and an 8-item static short-form were tested in a

Spanish clinical sample by Vilagut et al.[10] The results demonstrated sound psychometric properties to monitor depression in a clinical setting with adequate fit scores (CFI = 0.97, RMSEA = 0.08) and reliability scores of over 0.90, with high responsivity to change (d > 0.7) and the PROMIS tools detected depression disorder with great accuracy (AUC = 0.89). In another study with rheumatoid arthritis (RA) patients, validity and measurement precision were evaluated for the PROMIS physical function item bank and a 20-item short-form in patients with RA in comparison to the health assessment questionnaire disability index (HAQ-DI) and 36-item short form health survey (SF-36) physical functioning scale (PF-10).[11] High correlations were reported between the physical function instruments (r= 0.74-0.84).

In light of these findings, other disease states where the domain may be equally relevant may require further validatory support and testing. A few short-form measures have in fact been validated in more than one therapeutic area: for example, the fatigue item bank has been validated for cancer chemotherapy, osteoarthritis, heart failure, pain management and joint pain.[12-17] One must consider that if an item bank is chosen for use in a clinical trial, reliability and validity of the selected items need be established according to the FDA 'fit for purpose' criteria for that specific context of use based on the 2009 PRO guidance.[18]

WHY USE PROMIS?

There are both advantages and disadvantages to using PROMIS instruments in clinical and health outcomes research. It is noteworthy that PROMIS instruments and profiles are available online in multiple data collection platforms or as downloadable print-ready PDFs, translated into multiple languages, and are psychometrically validated in many chronic conditions. They are also available for immediate use without any additional overheads. Another benefit of using a PROMIS instrument is that it may reduce patient burden as it requires fewer questions than some other commonly-used instruments to measure the same concept. This especially applies to the short-form versions and CAT.

However, it should be noted that most PROMIS items/instruments (with the exception of cancer-specific instruments) were developed for diverse samples of patients with chronic conditions. Most item banks are centred on a sample that represents the US general population (as a rule of thumb, for PROMIS, the mean in the US general population is T=50). Patients with specific conditions (heart disease, cancer, rheumatoid arthritis, osteoarthritis, psychiatric illness, chronic obstructive pulmonary disease, and spinal cord injury among others) have participated in validation studies, yet the content validity of the item banks (i.e. their comprehensiveness and understandability) within a specific disease-state should not be assumed. Additional qualitative work may be needed to ensure the applicability and relevance of the tool for a specific population or rare disease. Additionally, one should consider that the PROMIS item banks do not assess the majority of illness signs, symptoms, or adverse events that are used as primary or secondary endpoints in clinical trials. For that reason, one may want to add other instruments that measure other health outcomes needed to be addressed in the trial to ensure accuracy in measurement of efficacy of all primary, secondary and exploratory endpoints.

PROMIS AND US REGULATING BODIES

The PROMIS measures have been used in clinical practice and research since their inception. Within pivotal clinical-trials for drug development, however, PROMIS measures have been less widely adopted (for example, only 16 of the 147 clinical trials identified in a July 2015 search were Phase 3 trials).

The US Food and Drug Administration (FDA) and European Medicines Agency (EMA) positioning on the use of the PROMIS measures is not formally documented or well-known but there is some discussion and support.[19] However, the FDA recently released a Clinical Outcomes Assessment (COA) Compendium which includes a table of all COAs that have been used to support drug labeling claims or submitted for qualification as well-defined and reliable tools for use in drug development programs.

The PROMIS physical function item bank (not any defined short-form or CAT instrument) was listed for qualification for sarcopenia, haematology and oncology.[20] At this point, it is still to be seen how the COA Compendium will create more confidence in labeling approvals, and thus further communication is encouraged with the FDA or EMA in these instances.

In addition to regulatory bodies, the centers for medicare and medicaid services (CMS) in the United States have cited the PROMIS 10-item for functional and global health status outcomes in their health care innovations award self-monitoring measures listing.[21] The National quality forum (NQF) is also in the process of assessing and dissecting the PROMIS measures as they relate to patient reported outcomes in performance measurement to understand their potential use and application.[22,23]

Generally speaking, there are several notable issues to be discussed before considerations are made to include a PROMIS instrument into a pivotal clinical trial:

1) The universal nature of the PROMIS item banks has raised questions about the validity, reliability, and generalizability of these concepts to specific populations and conditions. According to the FDA PRO Guidance for labeling-claims support and as previously noted, a PRO should demonstrate content validity in a specific context of use (i.e. in a specific population, for a specific objective, in a specific setting).[18] It is unclear as to

whether the development of the PROMIS item banks is considered adequate by FDA at this point in time.[24]

2) With regards to recall periods, many of the PROMIS items use seven-day recall periods (such as for anxiety, anger, depression, and pain interference) and a few use a 30-day period when the content being assessed may occur less frequently (sexual activity as an example). Consequently, special care should be made when choosing the instrument and its recall period as the FDA currently often encourages the use of daily symptom reporting (although their perspective on this appears to be changing and if a solid rational is presented for less frequent measurement, other periods may be acceptable).

3) The FDA further often encourages uni-dimensional measures with unequivocal interpretations. Although the item banks and certain short-form instruments have demonstrated uni-dimensionality using traditional and modern psychometric techniques, the profile instruments are in nature multidimensional and may cause further complications and concerns about the interpretations of clinically important differences.

4) Lastly, the regulatory perspective on the use of CAT to determine study endpoints is currently ambiguous. Although a recent publication states that "the FDA encourages demonstrations of the application and added-value of IRT-based instruments and CAT in the clinical trial setting to determine how to apply established measurement principles to endpoints", the following need to be discussed with the regulators prior to embracing CAT in pivotal studies:[19] (1) does the lack of consistency in items administered (and the order of those items) cause concern?; (2) how should the sponsor demonstrate equivalency of scores over time?; (3) how should the sponsor analyse data to compare cohorts and derive an estimate of treatment benefit (particularly related to derivation of responder definitions)?; (4) how can reliability of the item banks, administered via CAT be assessed?

DISCUSSION

Considerations for the use of universal PRO instruments in compliance with the FDA's PRO guidance have been advocated in a discussion meeting on the validity of measures.[18] In this meeting published by Magasi, PROMIS instruments were encouraged to be used in clinical trials via a working PROMIS initiative committee, as add-ons to disease- specific tools.[25] The initiative also supported the PROMIS group for more qualitative and quantitative research to take place for PROMIS measures to support the content validity and other measurement properties in a specific context of use. Such data is a prerequisite for reliable and valid assessment in a drug development trial and to meet the evidentiary expectations of the FDA. Some short-form PROMIS instruments have adequate empirical data to this regard; others do not.

The PROMIS II NIH common fund initiative ended in 2014, and the new funding scheme was modified to the Health Measures services instead (NIH grant U2C CA186878). More recently, the patient-centred outcomes research institute (PCORI) approved 5 million dollars of collaborative research funding in 2014 with PROMIS, focusing on individual projects no longer than 2 years. Bridging the gap between the NIH PROMIS network goal (to develop instruments that are appropriate across therapeutic areas) and the FDA PRO guidance (which requires that instruments have documentation of content validity specific to the target patient population) will be key for the use of these instruments in drug development. The PROMIS Physical Functioning item banks are currently undergoing consideration for qualification by FDA as endpoint measures of physical functioning in sarcopenia, haematology and oncology. This will provide welcome insight into the view of FDA on the PROMIS network initiative to use "measurement science to create a state-of-the-art assessment system for self-reported health".

ACKNOWLEDGEMENTS

We wish to sincerely thank Dr. Colleen McHorney, Dr. Arthur Stone, Dr. Leighann Litcher- Kelly and Dr. Susan Dallabrida for their invaluable input and editing of the information presented in this article.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. National Institute for Health. PROMIS Overview, 2016. Available at http://www.nihpromis.org/about/overview. Accessed on 3 March 2016.
2. Khanna D, Krishnan E, DeWitt E, Khanna P, Spiegel, Hays R. Patient-Reported Outcomes Measurement Information System (PROMIS) - The future of measuring patient reported outcomes in rheumatology. Arth Care and Res. 2011;63(11):486-90.
3. Gwaltney C, Shields A, Shiffman S. Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures: A Meta-Analytic Review. Value Health. 2008;11(2):322-33.
4. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual of Life Out. 2015;13:167.
5. Magnus B, Liu Y, He J, Quinn H, Thissen D, Gross H, et al. Mode effects between computer self-administration and telephone interviewer-administration of the PROMIS pediatric measures, self- and proxy report. Qual Life Res. 2016;25(7):1655-65.

6. Bjorner J, Rose M, Gandek B, Stone A, Junghaenel D, Ware J. Difference in Method of Administration Did Not Significantly Impact Item Response: An IRT-Based Analysis from the Patient-Reported Outcomes Measurement Information System (Promis) Initiative. Qual Life Res. 2014;23(1):217-27.
7. Bjorner J, Rose M, Gandek B, Stone A, Junghaenel D, Ware J. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. J Clin Epidemiol. 2014;67(1):108-13.
8. Health Measures. HealthMeasures Scoring Service powered by Assessment CenterSM. Available from https://www.assessmentcenter.net/ac_scoringservice Accessed on 20 Dec 2015.
9. PROsetta Stone®, Linking Patient- Reported Outcome Measures. What is PROsetta Stone? Available from http://www.prosettastone.org. Accessed on 20 December 2015.
10. Vilagut G, Forero C, Adroher N, Olariu E, Cella D, Alonso J. Testing the PROMIS Depression measures for monitoring depression in a clinical sample outside the US. J Psychiatr Res. 2015;68:140-50.
11. Oude M, Ten P, Glas C, Vonkeman H, Taal E, Krishnan E, et al. Validity and measurement precision of the PROMIS physical function item bank and a content validity-driven 20-item short form in rheumatoid arthritis compared with traditional measures. Rheumatology. 2015;54(12):2221-9.
12. Jensen R, Potosky A, Reeve B, Hahn E, Cella D, Fries J, et al. Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res. 2015;24(10):2333-44.
13. Broderick J, DeWitt E, Rothrock N, Crane P, Forrest C. Advances in Patient-Reported Outcomes: The NIH PROMIS Measures. EGEMS 2013;1:1015.
14. Broderick J, Schneider S, Junghaenel D, Schwartz J, Stone A. Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. Arthritis Care Res. 2013;65(10):1625-33.
15. Flynn K, Dew M, Lin L, Fawzy M, Graham L, Hahn E, et al. Reliability and construct validity of PROMIS measures for patients with heart failure who undergo heart transplant. Qual Life Res. 2015;24(11):2591-9.
16. Crins M, Roorda L, Smits N, de Vet H, Westhovens R, Cella D, et al. Calibration and Validation of the Dutch-Flemish PROMIS Pain Interference Item Bank in Patients with Chronic Pain. PLoS ONE. 2015;10(7):0134094.
17. Hung M, Baumhauer J, Latt L, Saltzman C, SooHoo N, Hunt K. Validation of PROMIS Physical Function computerized adaptive tests for orthopaedic foot and ankle outcome research. Clin Orthop Relat. 2013;471(11):3466-74.
18. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Edited by Food and Drug Administration. Rockville, MD: US Department of Health and Human Services; 2009.
19. Reeve B, Burke L, Chiang Y, Clauser S, Colpe L, Elias J, et al. Enhancing Measurement in Health Outcomes Research Supported by Agencies within the Us Department of Health and Human Services. Qual Life Res. 2007;16(1):175-86.
20. Food and Drug Administration, COA Compendium. Clinical Outcome Assessment Compendium. Available from http://www.fda.gov/downloads/drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM481225.pdf. Accessed on 18 July 2016.
21. Centers for Medicare and Medicare Services. Health Care Innovation Awards: Recommended Awardee Self-Monitoring Measures. Available from https://innovation.cms.gov/files/x/hciatwoawrdmsrs.pdf. Accessed on 25 August 2016.
22. National Quality Forum. Patient-Reported Outcomes. Available from http://www.qualityforum.org/Projects/n-r/Patient Reported_Outcomes/Patient Reported_Outcomes.aspx. Accessed on 26 August 2016.
23. Cella D. Patient Reported Outcomes in Health Care Performance Measurement: Issues Related to Selection and Administration. National Quality Forum: Northwestern University, 2012. Available at http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed on 5 July 2016.
24. Awad G, Scooler, N. Patient Reported Outcomes: New Approaches Using Item Banks and Computerized Adaptive Assessments. 2008. Available from https://isctm.org/public_access/ISCTM_SessionIISummary_PatientReportedOutcomes_DC2008.pdf. Accessed on 30 June 2016.
25. Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, et al. Content Validity of Patient-Reported Outcome Measures: Perspectives from a Promis Meeting. Qual Life Res. 2012;21(5):739-46.

Cite this article as: Shalhoub H, Reaney M. PROMIS® tools as endpoints in clinical trials: what should you know? A review of PROMIS® capabilities and the current regulatory space. Int J Clin Trials 2016;3(4):174-9.