

News from the NIH: Person-centered outcomes measurement: NIH-supported measurement systems to evaluate self-assessed health, functional performance, and symptomatic toxicity

Ashley Wilder Smith, PhD, MPH,¹ Sandra A. Mitchell, PhD, CRNP,¹ Cheryl K. De Aguiar, MPH,¹ Claudia Moy, PhD,² William T. Riley, PhD,³ Molly V. Wagster, PhD,⁴ Ellen M. Werner, PhD, MA⁵

¹Outcomes Research Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH, Rockville, MD, USA

²Office of Clinical Research, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA

³Office of Behavioral and Social Sciences Research, NIH, Bethesda, MD, USA

⁴Behavioral and Systems Neuroscience Branch, Division of Neuroscience, National Institute on Aging, NIH, Bethesda, MD, USA

⁵Blood Epidemiology and Clinical Therapeutics Branch, Division of Blood Diseases and Blood Resources, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA

Correspondence to: A Smith
smithas@mail.nih.gov

Cite this as: *TBM* 2016;6:470–474
doi: 10.1007/s13142-015-0345-9

INTRODUCTION

There is rapidly growing interest in the capture of person-centered outcomes in clinical and population-based research and in healthcare delivery settings. Stakeholders (e.g., patients, clinicians, payers, regulators, researchers) increasingly agree that person-centered outcome measurement can accelerate the development of new knowledge, improve the efficiency and quality of care, and may also contribute to clinician or health system performance metrics and regulatory review of new therapies [1–3]. These outcomes may be incorporated into both observational studies and clinical trials, and provide salient endpoints in trials of preventive or disease-modifying treatments, as well as behavioral or psychosocial interventions. Over the past decade, the National Institutes of Health (NIH) has invested in the development and evaluation of several measurement systems that are now available for research and clinical use. These include the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) [4], the NIH Toolbox for Assessment of Neurological and Behavioral Function (NIH Toolbox[®]) [5], the Quality of Life Outcomes in Neurological Disorders (Neuro-QoL) [6], Adult Sickle Cell Quality of Life Measurement Information System (ASCQ-Me) [7], and the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) [8]. In this paper, we (i) describe each system; (ii) highlight considerations in the design and interpretation of studies that employ one or more of these systems; and (iii) summarize future directions for continued implementation of these systems in clinical practice, population-based research, observational studies, and clinical trials.

OVERVIEW OF FIVE NIH-SPONSORED PERSON-CENTERED MEASUREMENT SYSTEMS

Historically, clinical research has suffered from a lack of comprehensive tools to measure person-centered

outcomes that are brief, highly accurate, and valid for comparisons across the age spectrum, and in healthy populations and disease groups. Data integration across studies has also been limited by the use of different measures of the same construct. PROMIS, the NIH Toolbox, Neuro-QoL, ASCQ-Me, and PRO-CTCAE were designed to address these issues.

All five systems measure a complement of important health outcomes through either self-report (e.g., common disease and treatment-related symptoms, function, health-related quality of life), or via performance-based measures (e.g., cognitive, motor, and sensory function). In combination, these systems cover both the spectrum of health and disease as well as more focused domains relevant within specific diseases.

These measurement systems utilized both modern measurement theory and classical test theory for question development, survey construction, scoring, and validation. For example, several systems used item response theory (IRT) [9] to develop and administer item banks (sets of questions) that measure different health domains. Item banks allow for flexible administration (i.e., any number of questions in any order) and greater precision. To ease interpretation and facilitate comparisons, several of the systems use a standardized *T*-score scoring metric (US population-based mean of 50 and standard deviation of 10). These systems have also made use of other innovative methods, such as computer adaptive testing (CAT) and conditional branching to tailor short forms, thus reducing respondent burden and allowing researchers to obtain precise measurement with a minimal number of items. Measures can be validly administered via multiple modes, including web, tablet, interactive voice response (IVR), and smartphone/handheld devices [10, 11].

Four of the systems (PROMIS, Neuro-QoL, the NIH Toolbox, and ASCQ-Me) are available as a suite of tools under one research resource, *HealthMeasures*. *HealthMeasures* is funded through a trans-NIH cooperative agreement facilitated by the National Cancer

Institute (NCI) and supported by 12 NIH Institutes and Centers. The goals of *HealthMeasures* are to stimulate use of these measurement systems by the research and practice communities, and to transition the systems to long-term sustainability via public/private partnerships. Developed under contract to the NCI, PRO-CTCAE is hosted at the NCI Center for Bioinformatics and Information Technology. It is anticipated that in the future, the PRO-CTCAE data collection system will interface with the NCI's Cancer Therapy Evaluation Program Enterprise System for clinical trials data management. The five measurement systems share many features; however they also have unique attributes, and are designed to measure distinct constructs (Table 1).

PROMIS[®]

PROMIS is a patient-reported outcome (PRO) measurement system comprising item banks that measure child and adult health across physical, mental, and social well-being (e.g., pain intensity, physical function, sleep disturbance, depression, anxiety, ability to participate in social roles and activities). PROMIS measures are not disease-specific and were designed for use across medical conditions in clinical research. The PROMIS system includes both static (fixed item) short forms as well as CAT. Measurement properties of PROMIS item banks, including mode invariance, have been extensively explored [4, 10, 12, 13].

Neuro-QoL

Like PROMIS, Neuro-QoL is a set of PRO tools developed using IRT, that measures health across physical, mental, and social domains for adults and children. However, Neuro-QoL was designed to be psychometrically sound and clinically relevant for individuals with neurological conditions. Neuro-QoL was specifically developed and tested within clinical populations with stroke, multiple sclerosis, amyotrophic lateral sclerosis, Parkinson's disease, epilepsy, and muscular dystrophy. Neuro-QoL enables within-disease as well as cross-disease comparisons and is intended for use in both neurology clinical trials and clinical practice. Validity, reliability, and responsiveness have been evaluated in neurological populations [6, 14, 15].

ASCQ-Me

Developed to complement the disease-agnostic PROMIS system, ASCQ-Me provides systematic, reliable, and valid PROs in adults with Sickle Cell Disease (SCD). ASCQ-Me domains can be assessed using both static and CAT measures and include severity, frequency, and impact of various domains such as pain, stiffness, sleep, SCD symptoms, social, and emotional outcomes for individuals with SCD. Initial psychometric testing of ASCQ-Me has been conducted [7].

NIH Toolbox

The NIH Toolbox is a multidimensional set of measures designed to monitor neurological and behavioral function in four domains: cognition, emotion, motor, and sensation. The NIH Toolbox includes participant self-report for emotional function, but is unique in its use of performance-based measures to evaluate cognition, sensation, and motor function. The NIH Toolbox has been tested for validity and reliability [5] across the age range for which it was developed—3 years to 85 years. The goal of the NIH Toolbox is to support rigorous measurement of functional status across the lifespan using a range of study designs.

PRO-CTCAE

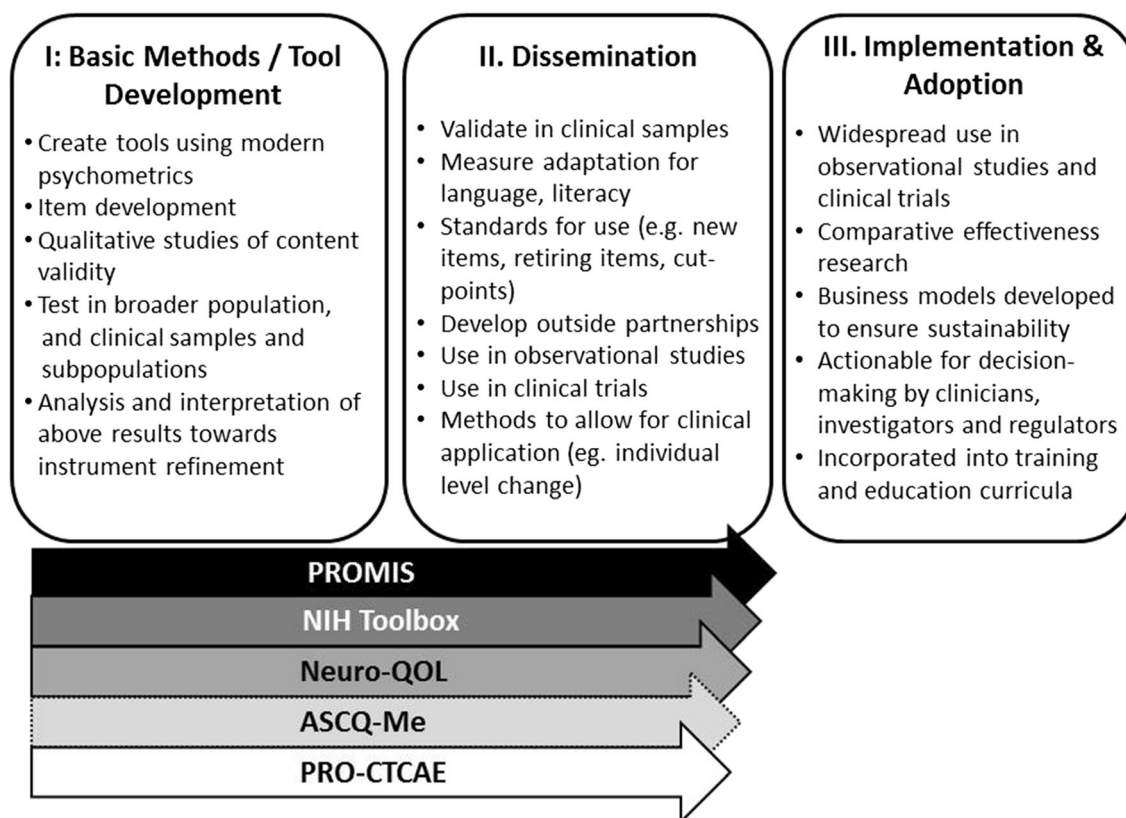
PRO-CTCAE assesses symptomatic toxicities (e.g., nausea, fatigue, neuropathy) experienced during and following cancer treatment in patients on cancer clinical trials. It was developed to complement and extend the Common Terminology Criteria for Adverse Events (CTCAE), NCI's system for clinician grading of treatment-related adverse effects in cancer clinical trials [8, 16]. Approximately 10 % of the adverse effects listed in the CTCAE are subjective and can be best assessed directly from patients [17]. PRO-CTCAE is intended to improve precision and reliability in gauging symptomatic toxicities of cancer treatment. PRO-CTCAE is applicable in selected cancer clinical trials where a precise description of the symptomatic toxicities experienced by patients is needed to better understand treatment tolerability. Based on the anticipated toxicity profile of a given therapy, investigators select a subset of the toxicities (including free-text write-ins), creating a study-specific short form. There is accumulating evidence demonstrating the psychometric properties [11, 18–21], and a pediatric version is being developed [22].

MEASUREMENT DEVELOPMENT AND IMPLEMENTATION STAGES

Each of these five measurement systems is at different stages of maturation along the measurement development and implementation continuum (Fig. 1). PROMIS, the NIH Toolbox, Neuro-QoL, ASCQ-Me, and PRO-CTCAE have completed development and initial evaluation (Stage I) and are progressing through scientific activities designed to enhance our capacity to compare and interpret research findings across multiple study designs and populations. The instruments in most of these systems either have gone through or are currently undergoing validation across the spectrum of health and disease, and in various languages (Stage II) [18, 23]. As NIH continues to expand the capacity for clinical research, the next phase (Stage III), focuses on widespread adoption of these instruments for use in clinical trials of new therapies, healthcare delivery research, and observational studies, as well as to improve the quality and patient centeredness of care. The inclusion of these tools in

Table 1 | Comparison of the five measurement systems

	PROMIS®	NIH Toolbox®	Neuro-QoL	ASCO-Me	PRO-CTCAE
Type of Measure	Self-report, proxy-report	Performance-based; self-report, proxy-report	Self-report	Self-report	Self-report
Generic or Disease-Specific	Generic	Generic	Neurological conditions	Sickle Cell Disease	Cancer
Constructs Measured	Physical, mental, social, and global health domains	Emotion, cognition, sensation, and motor function	Effects of neurological conditions on physical, mental, and social health domains	Effects of Sickle Cell Disease on physical, social, and emotional functioning and well-being	Symptomatic toxicity (companion to CTCAE)
Available Instruments	Item banks addressing >75 domains	Item banks and performance-based measures of function organized into four domains: emotion, cognition, sensation, and motor function	Item banks addressing >20 domains	Item banks addressing >7 domains	124 items reflecting 78 symptomatic toxicities
Formats	Static short forms, computerized adaptive tests (CATs)	Static short forms, CATs	Static short forms, CATs	Static short forms, CATs	Static short forms with conditional branching
Mode of Administration	Paper, web, handheld device, interactive voice response system (IVRS)	Paper (for emotion), web, handheld device (for all)	Paper, web, handheld device, IVRS	Paper, web, handheld device, IVRS	Paper, web, IVRS
Age	Pediatric and adult	Pediatric, adult, older adult	Pediatric and adult	Adult	Adult; pediatric measure in development
Available Languages	>40 (including English and Spanish)	English and Spanish	English and Spanish	English and Spanish	English, Spanish, German, and Japanese
Contexts of Use	Population surveillance, clinical research and trials, care delivery	Others in development; Epidemiologic studies, clinical research and trials	Others in development; Population surveillance, clinical research and trials in persons with neurologic disorder	Others in development; Population surveillance, clinical research and trials in persons with Sickle Cell Disease	Others in development; Cancer clinical trials including supportive care trials, clinical studies of cancer patients and survivors
For more information	www.HealthMeasures.net	www.HealthMeasures.net	www.HealthMeasures.net	www.HealthMeasures.net	http://healthcaredelivery.cancer.gov/pro-ctcae/



Adapted from: Mitchell SA, Fisher, C, Hastings C, Silverman, L, Wallen G. (2010). *Nursing Outlook*, 58 (6), 287-300.

Fig. 1 | Measurement development along the translation science continuum

clinical practice provides the opportunity for clinicians to benchmark their outcomes relative to research findings, and the use of harmonized measures across clinical settings supports the conduct of pragmatic clinical trials and accelerates knowledge transformation in learning healthcare systems.

CONSIDERATIONS FOR MEASURE SELECTION—AN EXAMPLE

Investigators select instruments from this suite of measures appropriate to their scientific aims and study design. As an example, an investigator studying the effects of armodafinil on fatigue, cognitive functioning, and depression in patients who have completed treatment for leukemia and are experiencing severe fatigue chooses measures drawn from *HealthMeasures* and PRO-CTCAE. For the efficacy endpoints, she selects both self-report (PROMIS Fatigue, Depression, and Cognitive Function item banks) and performance-based measures (the NIH Toolbox cognitive function measures addressing attention, processing speed, and executive function). These will be gathered at baseline; 1, 3, and 6 months after treatment initiation; and at treatment discontinuation. To capture the tolerability of armodafinil, the clinician-investigator will grade adverse treatment effects using the CTCAE and will

employ selected items reflecting symptomatic toxicity drawn from PRO-CTCAE (specifically anxiety, dizziness, sweating, insomnia, headache, and muscle weakness), administering PRO-CTCAE at baseline, weekly during the first 8 weeks of treatment, and monthly thereafter. Mixed linear models will be used to examine change over time in PROMIS and the NIH Toolbox measures; PRO-CTCAE data will be summarized using descriptive statistics.

OPPORTUNITIES AND CHALLENGES

It is anticipated that the availability of valid, precise, efficient, standardized self-report and performance-based measures will advance scientific discovery, enhance our ability to evaluate the effectiveness of alternative interventions and treatments, strengthen our national capacity to survey and monitor treatment effects over time, and improve patient-provider communication and decision-making in care delivery. Given that these tools are developed for use across diseases, they are also well-suited to capture the unique burden of illness and treatment that is added in the setting of multiple chronic conditions. However, continued research using these measures is needed to address current limitations and hurdles. These include incomplete coverage of all relevant PRO domains,

psychometric challenges with IRT (e.g., dimensionality), sparse research on cut-points, and population representativeness (low literacy, low educational attainment, minorities) in validation studies. Further, efforts are also needed to sustain these systems over the long-term to support increased accessibility and adoption.

The availability of these rigorously developed measurement systems creates a common currency for the evaluation of person-centered health outcomes. These systems support data harmonization across studies and settings, ease of interpretation, and reduced patient/participant burden. Adoption of these measurement systems enables economies of scale and enhanced efficiency and accelerates the knowledge generation/knowledge application cycle.

1. Perfetto EM, Burke L, Oehlein EM, Epstein RS. Patient-focused drug development: a new direction for collaboration. *Med Care*. 2015; 53: 9-17.
2. Jensen RE, Rothrock NE, DeWitt EM, et al. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Med Care*. 2015; 53: 153-159.
3. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, et al. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q*. 2014; 92: 754-775.
4. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010; 63: 1179-1194.
5. Gershon RC, Wagster MV, Hendrie HC, et al. NIH Toolbox for assessment of neurological and behavioral function. *Neurology*. 2013; 80: S2-6.
6. Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology*. 2012; 78: 1860-1867.
7. Keller SD, Yang M, Treadwell MJ, Werner EM, Hassell KL. Patient reports of health outcome for adults living with sickle cell disease: development and testing of the ASCQ-Me item banks. *Health Qual Life Outcomes*. 2014; 12: 125.
8. Basch E, Reeve BB, Mitchell SA, et al.: Development of the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *J Natl Cancer Inst*. 2014, 106.
9. Fries JF, Witter J, Rose M, et al. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J Rheumatol*. 2014; 41: 153-158.
10. Bjorner JB, Rose M, Gandek B, et al. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol*. 2014; 67: 108-113.
11. Bennett AV, Dueck AC, Mitchell SA, et al.: Mode equivalence and acceptability of Web, interactive voice response system, and paper-based administration of US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) Health Qual Life Outcomes. TBD.
12. Rothrock NE, Hays RD, Spritzer K, et al. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2010; 63: 1195-1204.
13. Liu H, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol*. 2010; 63: 1169-1178.
14. Lai JS, Nowinski C, Victorson D, et al. Quality-of-life measures in children with neurological conditions: pediatric Neuro-QOL. *Neurorehabil Neural Repair*. 2012; 26: 36-47.
15. Gershon RC, Lai JS, Bode R, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res*. 2012; 21: 475-486.
16. Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol*. 2007; 25: 5121-5127.
17. Xiao C, Polomano R, Bruner DW. Comparison between patient-reported and clinician-observed symptoms in oncology. *Cancer Nurs*. 2013; 36: E1-e16.
18. Arnold B, Mitchell SA, Lent L, et al.: Linguistic validation of the Spanish translation of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Health Qual Life Outcomes*. TBD.
19. Dueck AC, Mendoza TR, Mitchell SA, et al.: Validity and reliability of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *JAMA Oncology*. TBD.
20. Hay JL, Atkinson TM, Reeve BB, et al. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Qual Life Res*. 2014; 23: 257-269.
21. Kirsch M, Mitchell SA, Dobbels F, et al. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. *Eur J Oncol Nurs*. 2015; 19: 66-74.
22. Reeve BB, Withycombe JS, Baker JN, et al. The first step to integrating the child's voice in adverse event reporting in oncology trials: a content validation study among pediatric oncology clinicians. *Pediatr Blood Cancer*. 2013; 60: 1231-1236.
23. Alonso J, Bartlett SJ, Rose M, et al. The case for an international patient-reported outcomes measurement information system (PROMIS (R)) initiative. *Health Qual Life Outcomes*. 2013; 11: 210.