The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK

Jonathan P Evans1
Alexander Smith2
Chris Gibbons3
Jordi Alonso4-6
Jose M Valderas1

1Health Services and Policy Research Group, University of Exeter Medical School, Exeter, UK; 2Aneurin Bevan University Health Board, Clinical Research and Innovation Centre, St Woolos Hospital, Newport, UK; 3Patient-Reported Outcomes, Value and Experience (PROVE) Center Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA; 4IMIM-Hospital del Mar Medical Research Institute, Barcelona, Spain; 5Pompeu Fabra University (UPF), Barcelona, Spain; 6CiBeR en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

Abstract: The interest in patient-reported outcome measures (PROMs) continues to increase as recognition of their potential utility rises in an effort to make health systems more patient-centered. The US National Institutes of Health (NIH) Patient-Reported Outcomes Measurement Information System® (PROMIS®) has used state of the art psychometric and statistical techniques to create a universal PROMs language, with potential application across the whole spectrum of health conditions, languages, and geographic locations. PROMIS offers a versatile platform where specific health domains are assessed using both standardized short forms and computerized adaptive tests, which are automatically tailored to individual patients. The scores of each health domain or a standardized profile of multiple domains are all scored on a common metric scale. PROMIS is increasingly recognized as the international gold standard for patient-centered assessment, although the use of these tools in the UK is limited. In this review, the developmental methodology of the PROMIS is described with discussion of its relevant strengths and limitations for use in the UK. We provide a case study of the largest application of the PROMIS tools in the UK as an example of straightforward integration into health-care research. Barriers to the uptake of PROMIS in the UK include the technology requirement, measurement tradition, and lack of a clear understanding of its benefits, and although potential stakeholders should cautiously consider its use, its impressive potential and increasing international utilization should be recognized.

Keywords: outcomes, patient reported outcomes, quality of life, health-related quality of life, PROMIS

Introduction

Patient centeredness is an essential feature of high-performing health-care systems worldwide.1,2 Patient-reported outcome measures (PROMs) capture patients’ own evaluation of their health status3 and have become an essential tool in health-care evaluation.4,5 PROMs facilitate the quantification of the holistic health-care experience, by capturing aspects of the patient’s health that are truly relevant to themselves, their family, and their associated support network.

In the constant strive to improve the function of PROMs, modern psychometric and statistical techniques have elevated our expectations of measurement tools beyond compliance with the baseline metric standards of validity, reliability, and sensitivity to change. The evolution of relevant techniques allows the functionality of PROMs to expand, allowing accurate comparison across conditions, health-care systems, and geographical locations. A central tenet of the Patient-Reported Outcomes Measurement
Information System® (PROMIS®) is the development of common metrics that facilitate such comparisons. Developed as a health-focused rather than disease-focused measurement system, its domain structure can be utilized in the assessment of a single disease and single therapy or in patients with multimorbidity with their associated polytherapy. Developed in 2004 and sponsored by the US National Institutes of Health (NIH) Common Fund, PROMIS was conceived in an effort to address five major concerns about the current status of patient-centered outcome measurement: 1) the numerous PROMs available that measure the same concept, 2) the significant burden of many of those PROMs, 3) the variation in their psychometric quality, 4) the difficulty in comparing or combining data from different studies or populations, and 5) the difficulty in incorporating the PROMs into clinical practice. Although its use has risen rapidly, the current application in UK population remains limited. This review discusses the development and emergence of the PROMIS, its potential utility in the UK, and the current barriers to its implementation.

Discussion
The current state of play of PROMs in the UK
In many respects, the UK has been at the forefront of PROMs development and utilization. Although there is some adoption of PROMs for clinical care, the main driver of PROMs use in the UK has been to create, compare, and aggregate scores at the national level to inform policy. Led by the National PROMs program, these tools have been routinely collected for all National Health Service (NHS) patients undergoing hip replacement, knee replacement, groin hernia repair, and varicose vein surgery since 2009. This highly ambitious venture used fixed-length paper-based PROMs covering disease-specific and generic health domains. The NHS PROMs program has highlighted that it is possible to collect PRO data as part of routine delivery of care and then to collect follow-up data to track changes in health status. However, the routine use of PROMs has struggled to progress beyond these four conditions, and a clear route toward presenting the data to patients themselves has not been achieved.

What is PROMIS?
PROMIS is a set of measures covering different domains (eg, feelings, functions, or perceptions experienced by the patient) of physical, mental, and social health. These health domains are felt to be relevant across all health conditions, can be measured in both adult and pediatric populations, and are comparable across populations (Figure 1).

At its core, PROMIS relies on large collections of items (known as item banks) for each individual health domain such as pain interference, fatigue, and depression. Item banks are calibrated using modern psychometric techniques including item response theory (IRT), which ensures that they can be administered flexibly (ie, using different combinations of items from the bank) while remaining directly comparable. This flexibly means that PROMIS measures can be delivered as a fixed-length short form, either electronically or on paper, or as an individually tailored computerized adaptive test (CAT). In a CAT, questions from the item bank are sequentially presented to the patients using an algorithm that ensures only the more relevant and informative questions are asked. The PROMIS calibration method also allows scores to be directly compared with reference populations as well as with other PROMs, which measure the same domains; a technique known as a crosswalk.

In excess of 1,000 individual items (questions) have been psychometrically assessed for adults, producing item banks ranging from 12 to 124 items, which assess 51 distinct health domains. In the pediatric population, over 150 items have now been assessed, producing 18 distinct item banks; the delivery of these item banks in the pediatric population is through self-report or proxy-reported health by a parent or carer. A key aim of PROMIS was to standardize the outcome measurement in clinical practice and research, analogous to more commonly regarded health measurements such as a full blood count or blood chemistry panel, thereby facilitating comparability of data across studies and settings.

Essential components of PROMIS
The PROMIS item banks have been developed through a multisite research collaborative, applying current best practice techniques of mixed method item generation. Initially starting with a comprehensive collection and review of existing items found in legacy measures, as well as the development of new and modified items, over 8,000 items were collated. These were then reduced through expert content review, focus groups, cognitive testing, and secondary data analysis. The resultant items were then administered to large representative population samples in longitudinal validation studies.
**Figure 1** PROMIS® health domains for adult and pediatric populations.


**Abbreviation:** PROMIS, Patient-Reported Outcomes Measurement Information System.

**Abbreviation:** PROMIS, Patient-Reported Outcomes Measurement Information System.
The results from these samples were then analyzed using IRT techniques, resulting in calibrated item banks that were reviewed for reliability, validity, and sensitivity.

IRT is a family of mathematical models that express the probability of a particular response to an item as a function of the variable being measured. Thereby, it quantifies the relationship between a person’s response to a question and their individual position on the continuum of what is being measured (e.g., pain or physical function). This modern psychometric technique contrasts with classical test theory (CTT) from which many of the fixed-length PROMs in use today are developed; here, a person’s raw score on a questionnaire is a function of the true score and random error, where the error is the same for every person. The advantages of IRT are that scales can be reduced, resulting in precision with less items and thereby decreasing the questionnaire length and consequently the patient burden; the measures are more flexible as different items measure the same trait but with a known weight on the scale metric; the scores can be compared with different scales, allowing crosswalking of the score to other measures.

By delivering PROMIS electronically, advanced techniques such as CAT can be utilized. This system uses an algorithmic approach to search the whole available item bank to choose the optimal item to ask the participant based on their previous response. This process is repeated until a termination criterion has been met; a commonly used criterion is to stop the assessment when a sufficiently reliable estimate of the individual’s true score for the health domain has been reached. This minimizes respondent burden, improves reliability, maximizes the administration process, and can tailor the item content to the patient preferences.

Understanding PROMIS scores

The output from a PROMIS score is represented as a T-score. A T-score is a standardized score that is developed using a representative sample of the entire population, much like a z-score in an IQ test. Just as a z-score is centered around an average IQ score of 100, the PROMIS T-score is centered around a score of 50 with a score of and 10 being equivalent to 1 SD of that population. High scores equal more of the domain being measured (e.g., more fatigue, more pain). The average population scores have generally been derived from large US population samples. Thus, a patient who scores 70 on any PROMIS measure is 2 SDs above the average for that population on the domain which was assessed.

The scores are presented on a T-score map, either as item-level scores or as composite domain scores. The color-coded scale (blue = better to red = worse) was developed to facilitate interpretation by the clinician and patient. If multiple domains are being measured, these can easily be represented on a T-score map and the comparison with the population average scaled for each domain (Figure 2).

Operationalizing PROMIS

The delivery of PROMIS is adaptable to the health-care setting. The optimal situation is the use of electronic interfaces to deliver CATs, either within a clinical environment or in the patient’s home. This provides the opportunity to gain the most individualized data, with reduced patient burden (shorter questionnaires), instant feedback, and streamlined data storage and management. The portals through which the PROMIS measures are delivered use proprietary algorithms. Many of the platforms require subscription payments for use, though free and open-source software, such as REDCap, is available. Importantly, for UK users, the storage of data on non-UK servers requires careful data protection assessments, and Caldicott guardian review for all NHS patients and organizations will be required.

The collection of electronic data in UK health service settings may be limited by availability of computers or Wi-Fi-enabled devices in many NHS hospitals, as well as the staffing resources to assist in implementation. Should electronic assessment prove to be unfeasible, the alternative is the collection of paper-based short forms, either in clinical areas or using postal services. Although paper-based assessment also entails substantial implementation issues including scoring, storing, and presenting the collected data, all tasks which would be automated on an electronic system.

For those wishing to implement PROMIS, numerous short forms covering a wide range of domains are available free of charge from the healthmeasures.org website. Furthermore, custom short forms can be created which match the clinical features and symptom severity in the target population. As the items are scored on a common metric, each individual score, regardless of the items selected, remains comparable with the population norms. The item scores can then be tabulated on a preformatted spreadsheet using anonymized patient identifiers; this is uploaded onto the HealthMeasures scoring service, powered by the assessment center (https://www.assessmentcenter.net/ac_scoringservice) and the relevant T-scores are returned.

The flexibility of the PROMIS allows a number of routes to implementation, which may even include stepwise sequential adoption in which institutions begin with the method.
which best fits their current operational structure and scale up to more complex PROMIS versions.

An example of the operationalization of the PROMIS in a UK health-care setting is presented in the case study.

**PROMIS in the UK**

Four main PROMIS initiatives are currently developed in the UK. Three of them are research oriented, while the fourth one focuses on the routine measurement of PROMIS as part of poststroke routine care (see Case study). As part of an international collaborative effort with researchers in Germany and France, a web-based survey was simultaneously conducted in all three countries using the PROMIS Profile 29. Multigroup confirmatory factor analysis confirmed that scores derived from the PROMIS Profile 29 are largely comparable across UK, France, and Germany. General population reference data were also obtained, which can be compared with data collected with other PROMIS short forms or CATs.13

The National Institutes for Health Research (NIHR) has also supported efforts to validate PROMIS mental health and physical functioning in the UK. This research is now underway and will calibrate PROMIS domains on a representative sample of general population to provide the necessary information to use these domains in the UK, including calibrated item banks and T-scores. Once completed, these domains will be available as both short forms and CATs for use in clinical practice and research.

PROMIS has recently been used in a study aimed at assessing patient activation and the relationship between it and the outcome of upper limb injury with 775 UK adults who completed PROMIS upper extremity CATs through the assessment center portal.14

**What does it offer for the future?**

**Advantages and challenges**

The PROMIS international development team has gone to impressive lengths in their determination to attain the goal of creating a universal language for measuring health concepts. The translation of this aim into enhanced research, improved clinical decision-making, and facilitation of health policy planning are yet to be evidenced in the UK, but the roadmap toward these goals has been created.

The journey toward these aspirations will undoubtedly encounter some barriers. What may be the biggest obstacle is our ability to let go of our attachment to CTT measures and legacy PROMs.15 Within the UK, though PROMs are...
championed by the NIHR and the National Institute of Health and Care Excellence (NICE), the vast majority of research studies and clinical applications use both disease-specific and generic-health measures. The use of disease-specific measures, each with their own scoring metric, is a huge hindrance to evidence synthesis and meta-analysis, significantly impairing the ability to drive evidenced-based treatment and policy decisions. Any shift in use toward PROMIS becoming an international standard will require not only evidence of its utility in research studies but also buy in from national bodies, funders, and industry. The International Consortium for Health Outcomes Measurement (ICHOM), the mission of which is to unlock the potential of value-based health care by defining global standard sets of outcome measures, has integrated PROMIS into their recent stroke and pregnancy and childbirth outcome sets; this endorsement is likely to have a significant impact as we progress toward harmonizing outcome measurement.

The benefits of IRT have also been exploited to address the continued use of legacy measures. The ability to crosswalk the score of a legacy PROM to the universal metric of PROMIS permits comparison between studies using different PROMs to measure the same health outcome. The facility to crosswalk a PROMIS score to a legacy measure score would allow us to continue to utilize the data from seminal studies while also performing comparisons between contemporary research that either may or may not be applying PROMIS scores. Owing to the application of modern psychometric methods, this incredible flexibility within the PROMIS could have significant impact. Simple linking tables outlining the equivalent legacy PROM and PROMIS score are available via the PROsetta stone website (www.prosettastone.org).

Expertise in the application of IRT within the UK is significantly limited compared with CTT; knowledge translation and educational efforts will be required to increase confidence in the added value of using IRT methodologies. Training, dissemination, and the enthusiasm of early adopters in the UK will be required to convince potential users that IRT-derived measures are feasible to administer within an NHS, easy to interpret by clinicians and patients, and likely to increase the international impact of their research.

There is no shortage of evidence of the utility of PROMIS to feed enthusiasm in the UK. The University of Rochester has administered PROMIS CATs to nearly every orthopedic patient since 2015 and has over 1,000,000 evaluations collected. Each patient completes scores on three PROMIS domains: physical function, pain interference, and depression, and in-clinic testing takes an average of 2.4 minutes per patient. Analysis of these data has produced internationally award winning research articles. Furthermore, the American Academy of Orthopedic Surgeons recommend PROMIS among its instruments for the collection of orthopedic quality data. In New South Wales, Australia, the Integrated Care Strategy administers the PROMIS Global Health scale to patients for direct, timely feedback about their health. The US Food and Drug Administration includes PROMIS physical function scores in their Clinical Outcome Assessment Compendium for use in sarcopenia and oncology. The ability of PROMIS to provide information on value in health care can also be evidenced by its portfolio of successful funding.

The PROMIS Health Organization (PHO)

To inspire and support international use of PROMIS, the PHO was created with the mission to improve health outcomes by developing, maintaining, improving, and encouraging the application of PROMIS in research and clinical practice. This vibrant community of international collaborators has four main goals: 1) to advance the science of health outcomes assessment, 2) to disseminate standardized and validated questionnaires, 3) to foster the development of new patient-reported health outcomes for diverse populations, and 4) to educate the scientific and clinical communities on the science of PROs. Ultimately, they aspire to achieve widespread adoption of PROMIS internationally, developing PROMIS into the gold standard outcome metric that becomes part of routine clinical practice across multiple specialities (http://www.promishealth.com). Their growing community of researchers deliver educational programs, direct the PROMIS National Center contact lists (spanning 17 countries), and hold annual conferences, which in 2018 is intentionally directed at the UK audience and will be held in Dublin (October 28 and 29, 2018).

Conclusion

As the application of PROMs continues to rise, the adoption of a universal outcome measurement system is both feasible and desirable. By applying modern psychometric techniques and domain-specific measurement metrics, PROMIS stands as the most advanced step toward a common outcome language. The open-source ethos of its founders also represents a forward thinking approach that may level the playing field for PROMs application. Although there will be reticence to change from the currently applied outcome evaluations and technical barriers that impede implementation in the UK, these are far from insurmountable. By engaging with this system, assessing its feasibility, and confirming its UK popu-
lation validity, we will begin to benefit from its vast potential; if established it may herald a new era in international research collaboration, directly comparable health technology research and novel PROMs application.

Case study – PROMIS in the UK – 6-month poststroke review

The PROMIS-10 Global Health short form has been recommended by ICHOM for collection of patient-reported outcomes with stroke survivors as part of the ICHOM Stroke Standard Set of Outcome Data. The Welsh Government and Wales’ Stroke Implementation Group (SIG) have supported PROMs collection using PROMIS Global in the SIG’s National Stroke Plan.

Currently, the Stroke Sentinel National Audit Programme (SSNAP), overseen by the Royal College of Physicians (RCP), monitors and audits the quality and organization of stroke care in the NHS across three devolved NHS organizations: England, Northern Ireland, and Wales. All stroke units participate in the SSNAP. The unit level results are available in the public domain with specific domain scores color coded using a traffic light green to red system and scored from A–E. Within the SSNAP data set, there is a requirement for a 6-month poststroke follow-up, which is supported by the NICE and RCP guidelines. However, the data required for collection by SSNAP do not currently feature any data collected from the patient’s perspective.

The opportunity to supplement the SSNAP 6-month poststroke assessment with a PROM presented itself as an ideal research opportunity. The natural candidate for inclusion in the study was the PROMIS-10, due to its inclusion in the ICHOM Standard Set for Stroke and the similarity in ICHOM’s recommendation for PROMIS-10 collection at 90–120 days poststroke and SSNAP’s recommendation of a 6-month follow-up 120–240 days poststroke. In the current routine practice, SSNAP data recognize four methods of delivering the 6-month poststroke review: face-to-face, online, postal, and telephone. The acceptability of the PROMIS-10 via the four methods of administration used by SSNAP has been assessed as part of an RCT (ClinicalTrials.gov Identifier: NCT03177161) using a noninferiority design, with face-to-face (the most common method of 6-month follow-up) being the method against which three of the other methods’ acceptability would be assessed. A pragmatic design was adopted for the study, and it was embedded within routine stroke care as much as tolerable to replicate PROM implementation in as near to real world conditions as possible.

A number of disease-specific and generic challenges to PROM implementation have been identified. Previously stated was the challenge of NHS IT infrastructure, which resulted in the study not adopting CAT for the “online” arm. Moreover, many rural populations in the UK and especially Wales have limited internet connectivity, thus hindering mass participation in PROM collection from home. Leaving aside generic challenges, the prevalence of disability and the sudden onset of disabling symptoms poststroke present unique challenges to PROM completion poststroke. The prevalence of communication difficulties (aphasia), visual deficits (hemianopia), weakness, or inattention to one side (hemiparesis and hemi-inattention) all pose specific challenges to the acceptability of the PROMIS-10 with stroke survivors. This being the case, the recent validation study of the PROMIS-10 with stroke survivors by Katzan and Lapin described the minor disability levels in the study cohort as a limitation of the study. This in many respects leads to the continued use of well-established clinician- assessed outcome assessment in stroke, such as the Modified Rankin Scale.

Looking to the future, it will take both time and effort to overcome the challenges posed by poststroke disability and to enable stroke survivors to complete PROMs such as the PROMIS-10. Having said this, the growing PROMs research community in stroke is committed to meeting this challenge and there is good reason to be optimistic about the future of PROMs in routine stroke care and stroke research in the UK.

Acknowledgment

The present study was developed based on a presentation given by JMV at the Annual UK and Ireland PROMs Research Conference, Oxford, UK, on June 8, 2017.

Disclosure

JA, CG, and JMV are members of the PROMIS Health Organization. CG and JMV are PROMIS representatives in the UK. The authors report no conflicts of interest in this work.

References


