NIH Collaboratory Rethinking Clinical Trials®

Health Care Systems Research Collaboratory

Patient-Reported Outcomes

Table of Contents

Prepared by: NIH Collaboratory Coordinating Center with contributions from Tracie Locklear, PhD; Kevin P. Weinfurt, PhD; Amy Abernethy, MD, PhD; Kathryn E. Flynn, PhD; William Riley, PhD; Laura Lee Johnson, PhD

Version: January 2014

Introduction

This resource white paper was developed in 2014 to introduce clinical researchers to the definition and use of patient-reported outcomes (PROs). To incorporate PRO data collection in the clinical process, providers must be engaged, and the process used to collect PROs must improve clinical productivity and not create undue burdens on clinicians or patients. When PRO collection is aligned with clinical care, the information collected can be used in real time to triage patients, for quality monitoring, to trigger interventions and education, or for research. These uses should be transparent to patients, clinicians, researchers, and other stakeholders and may help engage patients in their own health care over time and ultimately inform and improve evidence-based patient care.

Defining Patient-Reported Outcomes

The U.S. Food and Drug Administration (FDA) defines a patient-reported outcome (PRO) as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else [1]." PROs typically include information about health-related quality of life (HRQOL), symptoms, function, satisfaction with care or symptoms, adherence to prescribed medications or other therapy, and perceived value of treatment [2]. PRO data are used to inform and guide patient-centered care, clinical decision-making, and health policy decisions and are an important component in learning healthcare systems [1].

> HRQOL is broadly defined as an individual's or a group's perceived mental and physical health over time. HRQOL includes the patient's report of the way a disease or its treatment affects physical, emotional and social well-being [2].

PROs are also used to measure risks and benefits of treatments. National surveys of patient experience are a feature of the United Kingdom's National Health Service (NHS) regulations [3], and NHS statistics on patient-reported outcome measures are available to the public. More recently, the Patient-Centered Outcomes Research Institute (PCORI), a private nonprofit institution funded through Medicare and from contributions from private insurers, has awarded funding to nearly 200 recipients conducting patient-centered comparative effectiveness research, including research based on PRO data, and has hosted workshops on building infrastructure for increasing the use of PROs in clinical and research settings.

A key goal of collecting PRO data is to improve clinical decision-making within the context of data-driven care. The successful integration of PRO measures within this context requires continuous collection of accurate, valid, accessible, and reusable data in real time to support patient care, clinical research, quality improvement, and comparative effectiveness research (CER).

How Are PROs Used?

Measuring research study endpoints

PROs play a significant role as study endpoints in the development and evaluation of new therapies [4]. The Patient-Centered Outcomes Research Institute (PCORI) recommends that the outcomes measured should include ones that patients notice and care about (e.g., survival, function, symptoms, health-related quality of life) [5]. If a PRO is used as a primary or secondary endpoint, it must be clearly defined and specified in the research protocol [1].

> Primary endpoints are used to test treatment effect, calculate sample size, and are statistically and rigorously tested. Secondary endpoints are of clinical interest, and it is common to test secondary endpoints only after a treatment effect is shown for the primary endpoint [1].

Although various possible endpoints may be of clinical interest, analysis of multiple endpoints can inflate the likelihood of Type I (false positive) error, and researchers must be careful when selecting the most appropriate endpoints to measure [1]. Primary endpoints that assess signs and symptoms have facilitated labeling claims and positive regulatory review [6], and well-defined and reliable PRO data-collection instruments can be used to support labeling claims, provided that the study is appropriately designed and the instrument has been validated [1].

Monitoring adverse events in clinical research

Patient-reported outcome instruments are also used to measure the adverse effects of a therapy separately from its effectiveness [1]. Exploratory work from the field of oncology suggests that PROs can be used for adverse event reporting with a high degree of patient engagement and compliance [7], and the Patient-Reported Outcomes Safety Event Reporting (PROSPER) Consortium developed a guidance to improve safety reporting by incorporating the perspective of the patient more fully [8]. The guidance for PRO-adverse events (PRO-AEs) includes definitions and suitable taxonomies, the range of datasets that could be used, data collection mechanisms, and suitable analytic methods [8]. For all clinical data, the Joint Commission recommends a common taxonomy for standardized terminology and classification schema for near misses and adverse events [9]. Taxonomies used for safety reasons must be easily understood, able to be reliably and consistently recorded, and allow for detection of patterns and trends from diverse settings [10].

Monitoring symptoms, patient satisfaction, and health care performance

The Effectiveness Guidance Document (EGD) for incorporating PROs into comparative effectiveness research in adult oncology encourages the assessment of patient-reported symptoms as well as health-related quality of life in all prospective clinical CER studies [11]. Although PRO measures were originally designed to support clinical research, they are being used in clinical practice as well. The importance of direct patient reporting has

been highlighted by both FDA and the European Medicines Agency (EMA) [12]. Because some of the effects of illness are known only to patients, FDA recommends that PRO outcomes instruments be used to assess efficacy outcomes in clinical trials [1]. The EMA is working to increase interaction with patient and consumer organizations, and is revising their "Framework of Interaction" to incorporate the role of patients on scientific committees, their involvement in benefit/risk evaluation, and a strategy for training and support [13].

Ideally, a PRO instrument will not only be a valid and reliable way to collect data, it will also make a positive contribution to clinical care [12]. Data collected from a PRO instrument can be used in longitudinal reporting at the point of care and as part of clinical decision-making and review of systems. In addition, PRO data can be used to trigger patient education and interventions and as a means to triage patients to receive other services, helping the patient understand that the information they are reporting is meaningful to their care. A recent systematic review of 27 studies in the cancer setting suggests that PROs improved patient-provider communication and patient satisfaction, in part, because clinicians talked to patients about their feelings and health status and were able to develop a shared view of treatment goals, health status, or reason for the visit [14]. Online patient self-reporting of toxicity symptoms during chemotherapy has been shown to be feasible, even among patients with advanced cancer and high symptom burdens [15]. PROs can also be used as reliable measures of healthcare performance, for example, the National Quality Forum (NQF) endorses the use of PRO-based performance measures (PRO-PM) for the purposes of performance improvement and accountability [16].

Example from the NIH Collaboratory

One of the demonstration projects within the NIH Collaboratory provides a good example of how PROs can be incorporated into clinical care. The Collaborative Care for Chronic Pain in Primary Care project is a mixed-methods, cluster-randomized pragmatic clinical trial that will evaluate the integration of psychosocial services into the primary care of patients with chronic pain. The proposed intervention, the Pain Program for Active Coping and Training (PPACT), involves a number of therapies designed to engage patients in their own care and help them manage their pain. The study will compare usual care to the effects of the intervention on a number of measures, including patients' pain symptoms, functional ability, satisfaction with health care services, and receipt of opiate medication. As part of the project, the patient completes a brief pain inventory (online or on paper), and this information is interfaced with an outside vendor who compiles the data and generates a report. The report incorporates real-time analysis and scoring of the data and presents the information in clinical context; as a result, it provides the physician with easily interpretable, actionable information derived from PRO data in order to promote discussion with the patient, trigger educational interventions, and aid in clinical decision-making. The data are also sent to the electronic health record (EHR) as a PDF in order to support clinical documentation. Discrete data are maintained in a database (with or without integration with the clinical data warehouse) and are available to be retrieved for subsequent research. This project offers an example of how clinicians, researchers, and patients all use the same PRO data points, but with differing purposes.

The Health Care Systems Research Collaboratory program was initiated by the NIH Common Fund in 2012. It engages healthcare systems as partners in discussing and promoting activities, tools, and strategies for supporting active participation in pragmatic clinical trials [17].

Measuring PROs: Instruments, Item Banks, and Devices

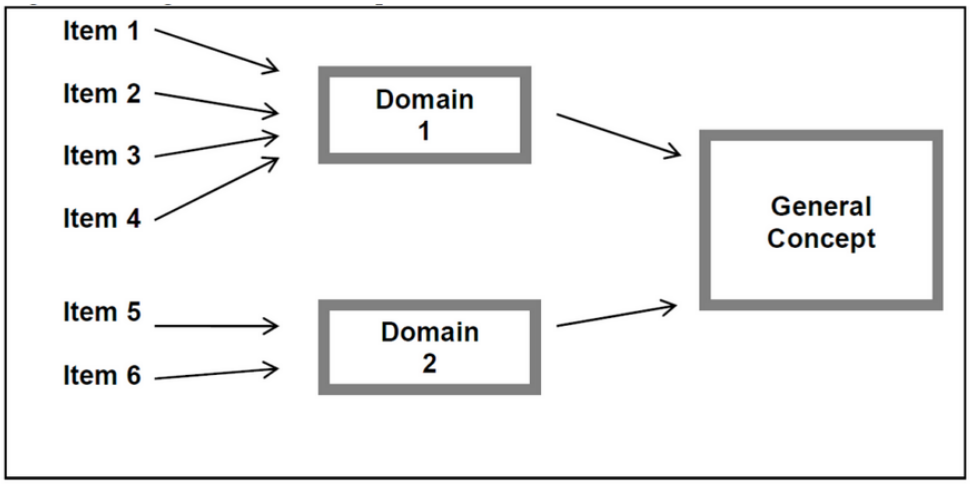
PRO Instruments

The FDA advises using a PRO instrument when the concept being measured is best known by the patient or best measured from the patient's perspective [1]. One simple and widely used example of a PRO instrument that may be familiar to many people is the Wong-Baker FACES scale, which allows patients (particularly children) to communicate a self-assessed measure of discomfort or pain to a healthcare provider.

Because using the best available measurement for an important concept is paramount, the FDA encourages investigators to determine whether an adequate instrument exists to address and measure the concepts of interest, or whether an existing instrument could be modified appropriately [1]. Crafting a measurement strategy may involve combining previously developed and validated instruments, modifying or adapting existing instruments for new purposes, or developing new questions or instruments [18]. When new PRO instruments are created, sponsors and investigators must provide documentation of patient input during the development process, as well as evidence of the instrument's performance in the specific application for which it was intended.

The first step is to hypothesize a conceptual framework based on expert knowledge and literature review for a concept of interest [1]. The conceptual framework should be closely aligned with research goals and should include a rationale for the outcomes of interest, the population of interest, and the particular outcome or treatment decision involved [19]. The framework consists of measurable items that collectively describe a domain—the specific feeling, function, or perception being measured.

Diagram of the conceptual framework of PRO instrument (adapted from FDA Guidance for Industry: Patient-Reported Outcome Measures)

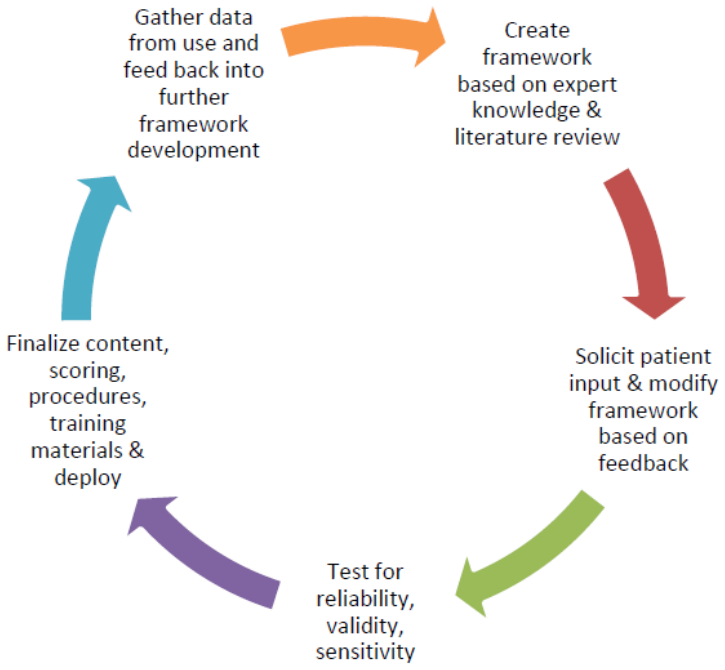


The conceptual framework may also be multidimensional and complex, requiring multiple domains to adequately describe a concept of interest. For example, health-related quality of life (HRQOL) is a multi-domain concept that represents the patient's perception of the effect of illness and treatment on physical, psychological, and social aspects of life. An inadequate conceptual framework can affect the grouping and scoring of items into domains and can also affect the analysis and interpretation of PRO scores [19].

In the clinical setting, an item is a question posed to a patient along with a set of possible answers. A domain is the specific feeling, function, or perception being measured; i.e., pain, fatigue, emotional distress, physical functioning, and social role participation.

After the conceptual framework is hypothesized, it should be adjusted based on patient input. The instruments measuring the domains within the conceptual framework are then tested for reliability, validity, and ability to detect change. Only after this testing is complete are the instrument content, scoring, procedures, formats, and training materials finalized and ready for full use in the research environment. Because the development of an instrument is an iterative process, after data are collected, analyzed, and interpreted, the instrument may be further modified and the cycle repeated [1].

Development and improvement cycle for a PRO instrument



A traditional method of measuring a health domain uses a static questionnaire in which the number and order of the items are set. These instruments are intended to be used in their original validated forms; thus, altering or shortening the instrument would render it invalid. Some of the problems encountered with traditional questionnaires include excessive length, likelihood that data would be missing from completed questionnaires, and difficulties in comparing the results of different questionnaires that evaluated similar concepts.

More recently, efforts to develop PRO measures have relied on item response theory (IRT), which uses information about the items themselves in addition to the answers to the items that a patient provides to estimate the patient's most likely score on an underlying continuum. IRT-based measures need additional psychometric evaluation beyond what is used for traditional questionnaires. Using a measure derived from a rigorously developed IRT-based model allows questionnaires to be tailored to the researcher's needs or to the individual patient's experience of the domain in question while still providing reliable and valid measurement [20]. In other words, we can ask different patients different questions, yet obtain comparable estimates of the domain being measured.

The NIH's Patient Reported Outcomes Information System, or PROMIS, has created a series of six brief instructional videos that provide an introduction to the concepts underlying Item Response Theory.

Item Banks

Many new measures of PROs are based on item banks, which comprise psychometrically calibrated items that measure a single domain. Any and all items can be used to provide a score for that domain; the items help define and quantify a common domain, providing an operational definition of the concept being measured. Because all items in the bank are calibrated based on IRT and reported using a common metric, investigators can compare scores on a domain among diverse groups of patients who may be responding to different items [21]. An item bank is dynamic, meaning that investigators may add more items to the bank over time and use only a subset of the bank's items to estimate the domain value.

There are two primary ways to derive measures from an item bank. In a fixed-length measure, every participant receives the same items in a common questionnaire or instrument. In a computerized adaptive test (CAT), all items in the bank are available to be administered, but subsequent items depend on answers to previous items. This maximizes the information obtained about the respondent while using a minimal number of items, because only the items most likely to provide additional information about the patient are administered. The CAT algorithm is programmed to scan the full set of available questions and determine those that have the least error at the current estimate of the respondent's level of the domain being measured.

Computer Adaptive Testing Approach to PROs Research Tool

For example, item banks and CAT testing have been used to improve, shorten, and computerize self-reported fatigue from the FACIT-Fatigue scale [21]. The investigators used 13 items from the Functional Assessment of Chronic Illness Therapy Fatigue Subscale (FACIT-F) as the basis for the IRT-based rating scale model. The investigators illustrate how CAT may work by using nine core items to measure level of fatigue. Using this illustration, a fatigue measure using four items was found to be as reliable as its full-length 13-item scale administration.

FACIT Measuring Questionnaire

Devices

Patient-reported outcomes can be collected using a wide array of electronic devices and interfaces available at home or in the clinic. These may include electronic devices such as cell phones, smart phones, tablet computers, personal computers, and wearable medical devices, as well as web-based portals accessed by either the patient or by clinic personnel.

The decision regarding the best location for collecting data and the most appropriate device(s) to use will ultimately depend on the clinical scenario. The use of tablet computers to collect symptom and HRQOL data has been shown to be feasible and acceptable to patients [22], who may perceive this as a "safer" means of providing private or highly personal answers about sensitive topics compared with paper questionnaires [23]. However, such devices must be easy to use and patients should understand how answering the questions can affect their care. Advances in technology and the rapid proliferation of mobile devices have liberated patients from needing to always answer questions at the office visit. Instead, patients are now able to provide responses whenever convenient for them and with greater frequency between visits. These technologies also liberate clinicians and staff from coding and scoring instruments by automating this process and making results immediately available. Despite this, clinicians and researchers must remain aware of the ethical and practical issues posed by electronic collection of PRO data. For example: how will the information be used? What happens when a patient reports a severe problem needing attention? How will privacy be maintained? Alignment of data collection with health care processes is a paramount task for researchers and clinicians working together to support optimal PRO data collection. The simplicity of use of new technology also creates a tendency to collect much more information than is needed, but it is important that researchers and clinicians confine data collection only to appropriate items for the issues in question [24].

Mobile Health (mHealth) and PROs: As the use of PROs in both research and routine patient care has grown, so has the use of mobile technologies (such as smartphones, tablet computers, and portable or wearable medical devices). These technologies, known collectively as "mobile health" or "mHealth" devices present a wide array of challenges and opportunities for medical research and quality improvement efforts.

National Institutes of Health (NIH)–sponsored PRO Measurement Information Systems (PROMIS®)

In 2004, a group of scientists from seven institutions and the NIH formed PROMIS, which was funded under the NIH Roadmap for Medical Research Initiative, now part of the NIH Common Fund. In 2009 the Network was expanded to 12 institutions, three coordinating centers, and the NIH.

PROMIS provides adult and pediatric item banks that measure health-related PRO domains. The measures have been standardized to provide common domains and metrics across a wide range of conditions and diseases.

The measures are flexible (allowing optional use of interchangeable items), efficient (minimizing number of items) and precise (minimal error in estimates of scores) [25]. PROMIS measures have been rigorously tested and validated and are freely available from HealthMeasures.

NIH Toolbox for the Assessment of Neurological and Behavioral Function

The NIH Toolbox is a multidimensional set of brief measures assessing cognitive, emotional, motor, and sensory function in patients ranging from 3 to 85 years of age. The toolbox standardizes the measures for use across diverse study designs and settings and uses multiple constructs of each domain, enabling longitudinal monitoring of cognitive and behavioral function and domain constructs across developmental stages [26]. Supported through the NIH Blueprint for Neuroscience Research and built by a development team of more than 250 scientists from almost 100 academic institutions [27], the Toolbox is intended to provide consistent measurement across studies and a scientific basis for identifying evidence-based best practices [26].

Neuro-QOL

Neuro-QOL is a set of PRO measures that assess the HRQOL of adults and children with neurological disorders such as stroke, multiple sclerosis, amyotrophic lateral sclerosis, Parkinson disease, epilepsy, and muscular dystrophy. Neuro-QOL includes item banks and scales that evaluate symptoms, concerns, and issues, thus enabling within- and across-disease comparisons to inform both clinical research and practice. The measures, which were developed through a collaborative, multisite research initiative sponsored by the National Institute of Neurological Disorders and Stroke (NINDS), provide brief, reliable, valid, and standardized QOL assessments [28].

The Adult Sickle Cell Quality of Life Measurement Information System (ASCQ-Me)

ASCQ-Me is a HRQOL instrument for adults with sickle cell disease (SCD) that was developed under a National Institutes of Health (NIH)/National Heart, Lung, and Blood Institute (NHLBI) initiative. It is designed to add specificity to the PROMIS HRQOL instrument in adults with SCD [29].

Interpreting Scores from PRO Instruments

The FDA's PRO guidance [1] provides recommendations for interpreting a change in PRO scores as evidence of treatment efficacy, and recommends establishing "responder definition" at the outset of a study. FDA defines *responder definition* as "a score change in measure, experienced by an individual patient over a predetermined time period that has been demonstrated in the target population to have a significant effect [1]." In other words, the responder definition describes the individual treatment effect, or meaningful change on a scale. This is an important guide, but the approach must balance the scenario and need. For example: not all assessments of responders will have optimal methods (such as anchor-based methods) in place to support the assessment; in these cases, less precise approaches are needed and the results should be put into context. Use of PROs for clinical care or comparative effectiveness research may not require the same level of precision as determining the balance of efficacy vs. harm for a new drug undergoing review for approval by the FDA.

Anchor-based methods

The FDA recommends using an empirical anchor-based method for defining a responder. This method estimates meaningful change using external anchors—usually patient- or clinician-reported global ratings in the same construct using a balanced ordinal change scale [30]. The PRO measures are compared with the anchor measures, and the FDA recommends that the anchors chosen should be easier to interpret than the PRO measure itself [1]. FDA uses this example: the number of incontinence episodes collected in incontinence diaries (anchor) has been used to define a meaningful score change (responder definition) for PRO instruments assessing the annoyance of incontinence [1].

Distribution-based methods

A distribution-based method for responder definition may use the between-person standard deviation or the standard error of the measure to define a meaningful change [1]. Although distribution-based methods can be used to categorize change as small, medium, or large and can be combined with anchor-based methods, the FDA warns that the clinical significance of score changes in a distribution-based method are considered supportive and are not appropriate as primary evidence of meaningful change. That being said, this is often a practical approach to interpreting PRO data in the clinical setting and for purposes of comparative effectiveness research.

Cumulative distribution functions

A cumulative distribution function (CDF) represents the entire distribution of responses for treatment and control group [1]. Essentially, a CDF is a plot that shows the proportion of patients at each point along a scale, i.e., the percentage of responders at each value of the PRO change score [30]. This cumulative display shows the change from baseline on the X-axis and the percentage of patients experiencing the change on the Y axis [1]. It has recently been argued that while CDFs are important, they are not a substitute for careful, anchor-based investigation of a PRO's responder definition [31].

Implementing PRO Measures

After a PRO and a device have been chosen; shown to be valid, reliable, and able to detect change [1]; and the goals for collecting PROs have been aligned, further work is needed to implement the PRO measure. First, the patients, setting, and timing of the assessments should be determined [32]. Whenever possible, PRO data should be collected electronically and the proportion of responders and cumulative distribution of responses in addition to mean changes in scores should be reported [11]. Methods for administering, scoring, and reporting the questionnaire and its results should be determined ahead of time [32], and the results of the PRO analyses should be published simultaneously with other clinical outcomes [11]. In addition, strategies for responding to issues raised by the questionnaire and for reducing missing data should be developed [32].

Reporting and Analysis

A recent review of 794 randomized controlled trials suggests that the quality of reporting, data analysis, presentation of HRQOL outcomes remains highly variable, and that consistent and interpretable PRO reporting practices are needed [33]. A CONSORT (Consolidated Standards of Reporting Trials) Statement extension regarding PROs [2] recommends five checklist items for RCTs in which PROs are primary or secondary endpoints:

1. Identify PROs as primary or secondary outcomes in the abstract.
2. Describe the hypothesis of the PROs and relevant domains (i.e., if a multidimensional PRO tool has been used).
3. Provide or cite evidence of the PRO instrument's validity and reliability.
4. Explicitly state statistical approaches for dealing with missing data.
5. Discuss PRO-specific limitations of study findings and generalizability of results to other populations and clinical practice.

Statistical Analysis

The International Council on Harmonisation's (ICH) Guidance for Industry on the Statistical Principles for Clinical Trials describes a set of best statistical practices that are broadly applicable to many clinical trials [34].

However, there are special considerations for research involving PROs. Importantly, the study protocol should identify and prespecify PROs that will be measured as part of the investigation [1]. For a PRO instrument with multiple domains, combining scores to create a composite endpoint must be justified by substantial empirical evidence that the components are of similar importance to patients, the frequency of the more and less important components are similar, and the components are likely to have similar treatment effects [1].

CONSORT PRO Extension
Description of CONSORT PRO Extension
E9 Statistical Principles for Clinical Trials

Missing Data

PRO data can be difficult to collect, and missing data are a frequent occurrence. For this reason, researchers should seek to minimize missing data and should also be familiar with appropriate methods for analyzing and reporting missing data [11]. Historically, when data collection is research-driven, PRO data are initially collected in abundance, but as the research progresses toward its end and research participants drop out of a study or decline to participate further, the frequency of missing data grows. However, when PRO collection is used to guide and improve clinical care there is less missing data over time. The PCORI Methodology Standards document (PDF) [5] recommends the following techniques for handing missing data:

- Describe in the protocol the methods to prevent and monitor missing data
- Describe in the protocol the statistical methods that will be used to handle missing data
- Use validated methods to deal with missing data that properly account for statistical uncertainty due to missingness
- Record and report all reasons for dropout and missing data, and account for all patients in reports
- Examine the sensitivity of inferences to missing data methods and assumptions, and incorporate into interpretation

Future Directions for PROs

Efforts are currently under way through the NIH Collaboratory to create guidelines and define best practices with respect to (1) selecting, compiling and curating the most appropriate PRO measures (and stimulating the development of new instruments when new solutions are needed); (2) providing guidance for the creation of efficient, high-quality PRO data collection systems compatible with electronic health records (EHRs) and registries; and (3) conducting statistical analyses of PRO endpoints.

One product of these efforts is an Effectiveness Guidance Document for incorporating PROs into CER in adult oncology [11] that includes the following key recommendations:

- Include assessment of patient-reported symptoms as well as health-related quality of life in all prospective clinical CER studies in adult oncology
- Identify symptoms relevant to a particular study population and context based on literature review and/or qualitative and quantitative methods

- Assure that PRO measures used are valid, reliable, and sensitive in a comparable population (measures particularly recommended include EORTC QLQ-C30, FACT, MDASI, PRO-CTCAE, and PROMIS)
- Collect PRO data electronically whenever possible
- Employ methods that minimize missing patient reports and include a plan for analyzing and reporting missing PRO data
- Report the proportion of responders and cumulative distribution of responses in addition to mean changes in scores
- Publish results of PRO analyses simultaneously with other clinical outcomes

Another important addition is a chapter on including PROs in registries included in an Agency for Healthcare Research and Quality (AHRQ) handbook on registries [35]. Read the full text of *Registries for Evaluating Patient Outcomes: A User's Guide*.

With the proliferation of devices for logging personal health data and the ability of investigators to turn big data into meaningful, clinically important information, the use of PROs may usher in a new era of clinical research.

Bibliography

1. FDA Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. Available at: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. Accessed September 25, 2013.

2. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013;309:814–822. PMID: 23443445. doi: 10.1001/jama.2013.879.

3. Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ 2010;340:c2921. PMID: 20566597. doi: http://dx.doi.org/10.1136/bmj.c2921.

4. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. Control Clin Trials 2004;25:535–552. PMID: 15588741. doi: 10.1016/j.cct.2004.09.003.

5. PCORI Methodology Committee. PCORI Methodology Standards. 2012. Available at: http://www.pcori.org/assets/PCORI-Methodology-Standards1.pdf. Accessed October 8, 2013.

6. Gnanasakthy A, Lewis S, Clark M, Mordin M, DeMuro C. Potential of patient-reported outcomes as nonprimary endpoints in clinical trials. Health Qual Life Outcomes 2013;11:83. PMID: 23675876. doi: 10.1186/1477-7525-11-83.

7. 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. PMID: 17991931. doi: 10.1200/JCO.2007.12.4784.

8. Banerjee AK, Okun S, Edwards IR, et al. Patient-Reported Outcome Measures in Safety Event Reporting: PROSPER Consortium Guidance. Drug Saf 2013. PMID: 24092596. doi 10.1007/s40264-013-0113-z.

9. Chang A, Schyve PM, Croteau RJ, O'Leary DS, Loeb JM. The JCAHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events. Int J Qual Health Care 2005;17:95–105. PMID: 15723817. doi: 10.1093/intqhc/mzi021.

10. Weingart SN. Beyond Babel: prospects for a universal patient safety taxonomy. Int J Qual Health Care 2005;17:93–94. PMID: 15772256. doi: 10.1093/intqhc/mzi029.

11. Basch E, Abernethy AP, Mullins CD, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol 2012;30:4249–4255. PMID: 23071244. doi:10.1200/JCO.2012.42.5967

12. Farnik M, Pierzchała WA. Instrument development and evaluation for patient-related outcomes assessments. Patient Relat Outcome Meas 2012;3:1–7. PMID: 22915979. doi: 10.2147/PROM.S14405.

13. European Medicines Agency. Fourth report on the progress of the interaction with patients' and consumers' organisations (2010). 2011. Available here. Accessed December 6, 2013.

14. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013;13:211. PMID: 23758898. doi: 10.1186/1472-6963-13-211.

15. Basch E, Iasonos A, Barz A, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol 2007;25:5374–5380.PMID: 18048818. doi: 10.1200/JCO.2007.11.2243.

16. National Quality Forum. Patient Reported Outcomes (PROS) in Performance Measurement. 2013. Available at: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed December 5, 2013.

17. Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc 2013 20:e226-31. PMID: 23956018. doi: 10.1136/amiajnl-2013-001926.

18. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: instrument development and selection issues. Value Health 2007;10 Suppl 2:S86–93. PMID: 17995478. doi: 10.1111/j.1524-4733.2007.00271.x.

19. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: conceptual issues. Value Health 2007;10 Suppl 2:S66–75.PMID: 17995476. doi:10.1111/j.1524-4733.2007.00269.x.

20. van der Linden WJ. Handbook of modern item response theory. New York: Springer; 1997.

21. Lai J, Cella D, Chang C-H, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. Qual Life Res 2003;12:485–501. PMID: 13677494. doi:10.1023/A:1025014509626.

22. Abernethy AP, Herndon JE 2nd, Wheeler JL, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. J Pain Symptom Manage 2009;37:1027–1038. PMID: 19394793. doi:10.1016/j.jpainsymman.2008.07.011.

23. Dupont A, Wheeler J, Herndon JE 2nd, et al. Use of tablet personal computers for sensitive patient-reported information. J Support Oncol 2009;7:91–97. PMID: 19507456.

24. Kyte D, Draper H, Calvert M. Patient-reported outcome alerts: ethical and logistical considerations in clinical trials. JAMA 2013;310:1229–1230. PMID: 24065005. doi:10.1001/jama.2013.277222.

25. 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. PMID: 20685078. doi: 10.1016/j.jclinepi.2010.04.011.

26. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. Neurology 2013;80:S2–6. PMID: 23479538. doi: 10.1212/WNL.0b013e3182872e5f.

27. Hodes RJ, Insel TR, Landis SC, NIH Blueprint for Neuroscience Research. The NIH toolbox: setting a standard for biomedical research. Neurology 2013;80:S1. PMID: 23479536. doi: 10.1212/WNL.0b013e3182872e90.

28. Cella D, Lai J-S, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology 2012;78:1860–1867. PMID: 22573626. doi: 10.1212/WNL.0b013e318258f744

29. Panepinto JA. Health-related quality of life in patients with hemoglobinopathies. Hematology Am Soc Hematol Educ Program 2012;2012:284–289. PMID: 23233593. doi:10.1182/asheducation-2012.1.284.

30. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:163–169. PMID: 21476818. doi: 10.1586/erp.11.12.

31. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S, Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. Qual Life Res 2013;22:475–483.PMID: 22528240. doi: 10.1007/s11136-012-0175-x

32. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012;21:1305–1314. PMID: 22048932. doi: 10.1007/s11136-011-0054-x

33. Brundage M, Bass B, Davidson J, et al. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. Qual Life Res 2011;20:653–664. PMID: 21110123. doi 10.1007/s11136-010-9793-3.

34. International Council on Harmonization. Guidance for Industry: E9 Statistical Principles for clinical Trials. 1998. Available here. Accessed October 10, 2013.

35. Gliklich RE, Dreyer NA eds. Registries for Evaluating Patient Outcomes: A User's Guide. 2nd ed. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010. Available at: http://www.ncbi.nlm.nih.gov/books/NBK49444/. Accessed October 8, 2013.