National Quality Forum
Patient Reported Outcomes (PROs) Workshop #1
July 30-31, 2012

Workshop Summary

Table of Contents

National Quality Forum
Patient Reported Outcomes (PROs) Workshop #1
July 30-31, 2012
Workshop Summary

I. Introduction

The increasing integration of health care delivery systems provides an opportunity to manage the entire patient-focused episode of care and to assess the impact of care on patient outcomes, including patient-reported outcomes (PROs). PROs are defined here as "any report of the status of a patient's (or person's) health condition, health behavior, or experience with health care that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."[1]

Various tools (i.e., instruments, scales, single-item measures) enable assessment of patient–reported health status for physical, mental, and social well–being are referred to PRO measures (PROMs). It is important to distinguish between PROMs and aggregate-level performance measures. A PRO-based performance measure (or PRO-PM) can be defined as a performance measure that is based on patient-reported outcome data aggregated for an accountable healthcare entity such as a hospital, a physician practice, an accountable care organization, etc. NQF endorses performance measures (PRO-PMs); NQF does not endorse tools to measure PROs (PROMs).

Distinction Between PROM vs. PRO-based Performance Measure (PRO-PM)
The following example illustrates the distinction between the PROM and PRO-PM. The Patient Health Questionnaire (PHQ-9) is a widely accepted, standardized *tool* to assess depression [Copyright © 2005 Pfizer, Inc. All rights reserved]. NQF has not endorsed the PHQ-9 *tool*, but has endorsed three performance measures from Minnesota Community Measurement based on the PHQ-9. The first is a process performance measure — percentage of patients with a diagnosis of major depression or dysrhythmia administered the PHQ-9 in a 4-month measurement period (NQF #0712). The process measure is paired with two outcome performance measures of depression remission — percentage of patients with diagnosis of major depression or dysrhythmia and initial PHQ-9 score >9 with a follow-up PHQ-9 score <5 at six months (NQF #0711) and at twelve months (NQF #0710).

[1] U.S. FOOD AND DRUG ADMINISTRATION. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Federal Register 2009;74(35):65132-133. Available here.

Terms Used in this Report

Patient-reported outcome (PRO): The concept of 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.

PRO domains included in this project encompass:

- functional status/health-related quality of life;
- symptom and symptom burden;
- experience with care; and
- health-related behaviors.

PRO measure (PROM): Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).

Performance measure: Numeric quantification of healthcare quality for a designated accountable healthcare entity, such as hospital, health plan, nursing home, clinician, etc.

PRO-based performance measure (PRO-PM): A performance measure that is based on PRO data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

There are two major challenges to using PRO-PMs for purposes of accountability and performance improvement: 1) PROMs have not been widely adopted in clinical use, and are therefore not familiar to many providers and payers; and 2) little is known about aggregating patient data on PROMs for measuring performance of the healthcare entity delivering care. While there has been great interest in using PROMs in performance measurement, foundational work needs to be done to address the methodological and data challenges. Efforts are currently underway to develop and test mechanisms for collecting and using patient-reported data such as Dartmouth Spine Center and Partners Healthcare in Massachusetts. Accordingly, this is an opportune time to also consider the methodological issues surrounding use of such data when available in performance measurement. These issues include collecting patient-reported outcome data in the clinical environment and the aggregation of the data to assess performance of a healthcare entity.

NQF Direction for Performance Measurement

NQF is a voluntary consensus standard setting organization. It endorses performance measures to assess the quality of healthcare for use in accountability applications such as public reporting and payment as well as performance improvement. NQF is a neutral evaluator of performance measures and a neutral convener of groups such as the PRO Expert Panel; NQF is not a measure developer.

The NQF work informs and is aligned with the National Quality Strategy, which places priority on engaging patients and families in their care. Patient-reported outcomes, patient engagement and self-management, and shared decision-making have been identified as important gap areas for measure development by the NQF- convened National Priorities Partnership and the Measure Applications Partnership recognizing the patient is a valuable and often untapped resource in the assessment of the efficiency (defined as quality and costs) of care.

The quality performance measurement is evolving and the direction for NQF-endorsed performance measure includes:

- a drive toward higher performance reflected in more outcome measures rather than very basic processes such as assessment;
- measuring disparities in all we do;
- a shift toward composite measures that summarize multiple aspects of care;
- harmonization of measures across sites and providers; and
- measurement across longitudinal patient-focused episodes including outcome measures (including PRO-PMs), process measures with direct evidence of impact on desired outcomes; appropriateness measures; and cost/resource use measures coupled with quality measures, including overuse.

Figure 1 depicts the relationship between structure, process, and outcome. For NQF endorsement, there is a hierarchical preference for performance measures of health outcomes that are linked to evidence-based processes or structures; or outcomes of substantial importance with a plausible link to healthcare processes. Next in the preferred hierarchy are measures of intermediate outcomes and processes closely linked to desired outcomes. Measures of processes that are distal to desired outcomes (e.g., assess patient) and those that are satisfied by a "checkbox" are considered to have the least impact on the goal of improving healthcare and health.

Figure 1. Structure-Process-Outcome

PRO Project Goals and Description

The goals of the NQF PRO project are to:

- identify key characteristics for selecting PROMs to be used in PRO-PMs;
- identify any unique considerations for evaluating PRO-PMs for NQF-endorsement and use in accountability applications; and
- lay out the pathway to move from PROM to NQF-endorsed PRO-PM.

The PRO project was designed to bring together the stakeholders necessary to facilitate the groundwork for the development, testing, endorsement, and implementation of PRO-PMs. Those stakeholders included researchers, clinicians, performance measure developers, and consumer and purchaser representatives. Two workshops with an expert panel and two commissioned papers – the first focused on the individual-level PROMs and the second on the aggregate-level PRO-PMs – will be used to achieve the goals of the project.

Workshop #1

NQF hosted the first workshop on July 30-31, 2012. Objectives for the workshop included:

1. Identify best practices and lessons learned from initiatives that have implemented individual-level PROMs in performance measurement;
2. Discuss the major methodological issues related to the selection, administration and use of individual-level PROMs in performance measures;
3. Discuss key considerations for inclusion of PROMs into EHRs and policy implications;
4. Identify the characteristics of individual-level PROMs suitable for potential use in performance measures; and
5. Identify an initial set of PROMs most suitable for development and testing of performance measures.

A key input to Workshop #1 was a commissioned paper on the methodological issues in the selection, administration, and use of patient-reported outcome measures in performance measurement (located in Appendix C).

II. Key Themes from the Workshop Discussions

Person-Centeredness

The resounding overarching theme that arose from the workshop discussions was "person-centeredness". In this context, PROMs are seen as an important step towards engaging patients and providers in creating a person-centered health system. Several principles were identified to more

authentically engage patients in performance measure development include: (1) identifying meaningful measures; (2) capturing the measurement target; (3) communicating and using results; and (4) relating performance measures to patient goals. Figure 2, presented by Lori Frank of the Patient-Centered Outcomes Research Institute (PCORI), illustrates a cycle of engagement that represents truly engaging patients at all phases.

Figure 2. Guiding Principles for Stakeholder Engagement in Performance Measurement

The expert panel found the term "patient-reported" is too limiting because it does not convey inclusion of persons receiving supportive services, such as those with disabilities. For example PROs in the experience with care category also need to address whether needs are met and appropriate linkages to community-based services are made.

Importantly, as patients become more activated in their care by providing systematic feedback on their functional or health status, for example, it is critical for the flow of information between provider and patient to be bi- directional so patients do not view this as overly burdensome for which they are not receiving benefit.

Accountability

Another key theme of the workshop discussions was accountability. Specifically, what is the state of readiness of using PROs for accountability purposes? Some of the issues raised were similar to those about other outcome-based performance measures: outcomes are influenced by multiple factors, including some outside the control of providers; questioning whether there is clear evidence that the outcome is influenced by healthcare; and outcome measures do not provide the level of detail needed so providers know what to target to improve. Conversely, outcomes are what people receiving services care about the most and are the goal of providing services. Another strong point of outcome-based measures is they are integrative, reflecting multiple processes and clinicians involved in care, and thus

promote shared accountability for healthcare and health. NQF guidance on evidence recognizes that once measured and reported many outcomes that were not thought to be modifiable tend to improve. This suggests that measurement stimulates identification and adoption of effective practices. NQF criteria on validity recognize the role of risk adjustment when using outcome-based performance measures.

To begin to address the state of readiness of the field for PRO-based performance measures, several suggestions were made regarding the pathway from PROM to PRO-PM presented below. These are in a formative stage and will be discussed in more depth at Workshop #2 with the end goal of laying out a detailed pathway or roadmap to guide the field.

- Begin with process measures focused on use of PROMs in clinical practice to promote their use and gain experience before using outcome measures. Using PROMs may also be an indicator for patient engagement and person-centered care. The challenge is to construct process measures that are meaningful and represent more than checking that a PROM was given to a patient.
- For outcome performance measures based on PROMs, as a starting point select PROMs that are actionable or mutable by the healthcare entity whose performance is being measured. This relates to the concept of responsiveness in Table 3.
- A related suggestion is to begin with PROs that are treatment-specific such as function after hip or knee replacement or revascularization where the treatments are evidence–based and there is some experience in the field.
- An alternative approach, however, is to begin with more global PROMs that have applicability to broader populations. This breadth will need to be balanced with emerging but not as widely applied system-based interventions.

Methodological Issues

There was general agreement that the commissioned paper addressed the methodological issues and key characteristics that need to be considered when selecting PROMs for performance measurement. Several issues were emphasized in the context of using PROMs in creating performance measures. Missing data is an important consideration when aggregating PROM data for performance measurement. This issue encompasses missing responses on a multi-item scale; missing responses from eligible people and its impact on potential response bias; missing information due to exclusions; and using proxies in the face of missing responses. Processes must be in place to safeguard against these exclusions and biases, and more robust engagement strategies are needed over time to prevent these gaps in response rates.

There are multiple modes and methods for collecting data and also multiple validated tools for the same PRO concept. The need for flexibility in implementation needs to be balanced with the need for standardization or equivalence for performance measurement to ensure comparability. Both legacy tools and PROMs built on Item Response Theory (IRT) have advantages and disadvantages in their use, and an approach that bridges the two is needed in order to mitigate their shortcomings.

Information Technology

Another theme that developed throughout the workshop was the usefulness of information technology in regards to usability and widespread use of PRO-based performance measures. Technology can increase response rates by allowing responses from home or waiting room, via computer tablet or telephone. Technology allows for quick scoring and feedback to the people providing the responses or scanning of paper and pencil responses. Computers facilitate the real-time application of item response theory in computer adaptive testing, which allows more efficient administration of PROMs and calibration of multiple instruments to a standard scale.

Integration of PROMs into electronic health records (EHRs) could facilitate their use for patient-centered care management as well as provide data for performance improvement. However, integration into EHRs has some important considerations. Data standards for EHRs are needed to address incorporation of PROM data. Aspects capturing PROM data such as the source (e.g., self or proxy), specific PROM instrument, method and mode of data collection, date, capturing items or computed score; and using PROM data in clinical practice such as display of results, alerts, and decision-logic need to be considered. Some PROMs such as experience with care may not be appropriate for inclusion in EHRs because current tools and approaches are based on the premise of anonymity. Incorporating data provided by patients into the health record may increase their sense of ownership of the record and demands for extracting information as well as providing data. This is an opportune time to include PROMs in EHRs and leverage the resources being directed to adoption of electronic health records through the Medicare EHR Incentive Program referred to as Meaningful Use.

III. Characteristics for Selecting Individual-level PROMs

There was general agreement that the characteristics identified in the commissioned paper provide a sound basis for selecting PROMs that would potentially be suitable for use in performance measurement. Below is a list of the characteristics from Table 3, but the entire table with full details can be found in Table 3 in Appendix A.

Characteristics for Selecting PROMs Identified in Commissioned Paper

1. Conceptual and Measurement Model Documented
2. Reliability
    - 2a. *Internal consistency (multi-item scales)*
    - 2b. *Reproducibility (stability over time)*
3. Validity
    - *3a. Content Validity*
    - *3b. Construct and Criterion-related Validity*
    - *3c. Responsiveness*
4. Interpretability of Scores

5. Burden
6. Alternatives modes and methods of administration
7. Cultural and language adaptations
8. Electronic health record (EHR) capability

Additional Characteristics for Selecting PROMs

In addition to the key psychometric properties and characteristics listed above, the Expert Panel identified the following other characteristics to serve as guideposts for identifying PROMs most suitable for development and testing of performance measures. The Expert panel will be asked to review for potential refinement through a survey. Further exploration of these characteristics will take place at the second workshop and their relationship to the NQF endorsement evaluation criteria for which they are mutually reinforcing.

The preliminary characteristics put forth include:

- Meaningful to persons and their families – as well as clinicians and other health professionals who serve them. Meaningfulness encompasses the degree of importance to adequately capture the impact of health related quality of life (including functional status), symptom and symptom burden, experience with care, or of a health-related behavior on the patient.
- Actionable with evidence-based justification for selection that leads to improvement by key end users including persons, providers and systems.
- Able to facilitate shared decision-making—including engaging patients in their own self-management and goal attainment aligned with their preferences (e.g., identifying outcomes important to them; involving in measure development and testing; assessing cultural/ linguistic adaptability) while being flexible enough to aggregate or roll up for a population or accountable entity.
- Implementable taking into account burden to the person, provider, and system including but not limited to: cost barriers to the use of proprietary measures; potential for unintended consequences (e.g., gaming or adverse selection); shown to be successfully integrated into routine clinical practice; measures that are disparities sensitive; and adaptability to electronic or other alternate formats.

IV. Next Steps

The second workshop will be held at NQF offices in Washington, D.C., on September 11-12. This workshop will build off of the first workshop and the first commissioned paper; however a second commissioned paper will help inform next steps regarding developing reliable and valid performance measures eligible for NQF endorsement that can be used for accountability and to inform quality improvement. The second workshop will focus on the aggregation of the individual patient-reported

outcome data to measure the performance of an accountable entity providing healthcare (e.g., hospital, physician, accountable care organization). The actual workshop objectives include:

1. Discuss the major methodological issues related to reliability and validity when aggregating PROM data into a performance measure;
2. Identify unique considerations in relation to the NQF endorsement criteria for PRO-based performance measures (PRO-PM) (as compared to other quality outcome performance measures); and
3. Lay out the critical path from PROM to PRO-based performance measure (PRO-PM) endorsed by NQF for use in accountability and performance improvement.

Following the workshop #2, a draft report will be prepared and reviewed by the Expert Panel and then posted for comment. The Consensus Standards Approval Committee and NQF Board of Directors will approve the report and any recommendations with implications for the NQF measure endorsement criteria.

V. Appendices

A. Table 3 from commissioned paper
B. Expert Panel Roster
C. Commissioned paper
D. Workshop participant list

Appendix A – Table 3 from Commissioned Paper

Table 3[2]. Important characteristics and best practices to evaluate and select PROs for use in performance measures[260,265]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
1.	**Conceptual and Measurement Model**		
	A PRO measure should have documentation defining and describing the concept(s) included and the intended population(s) for use.	Target PRO concept should be a high priority for the health care system.	Factorial validity of the physical function and pain subscales has been inadequate.[331]
	There should be documentation of how the concept(s) are organized into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts.		
2.	**Reliability**		
	The degree to which an instrument is free from random error.		
2a.	***Internal consistency*** *(multi-item scales)*	▪ reliability estimate ≥ 0.70 for group-level purposes ▪ reliability estimate ≥ 0.90 for individual-level purposes	Cronbach alphas for the three subscales range from 0.86 to 0.98.[332-334]
2b.	***Reproducibility*** *(stability over time)* ▪ type of test-retest estimate depends on the response scale (dichotomous,		Test-retest reliability has been adequate for the

[2] This table is adapted from recommendations contained within a report from the Scientific Advisory Committee of the Medical Outcomes Trust and a report submitted to the PCORI Methodology Committee. The recommendations from these sources have been adapted to enhance relevance to PRO selection for performance measurement.

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
	nominal ordinal, interval, ratio)		pain and physical function subscales, but less adequate for the stiffness subscale.[334]
3.	**Validity**		
	The degree to which the instrument reflects what it is supposed to measure.	There are a limited number of PRO instruments that have been validated for performance measurement.	
3a.	***Content Validity***		
	The extent to which a measure samples a representative range of the content.		
	A PRO measure should have evidence supporting its content validity, including evidence that patients and/or experts consider the content of the PRO measure relevant and comprehensive for the concept, population, and aim of the measurement application.		Development involved expert clinician input, and survey input from patients,[335] as well as a review of existing measures.
	Documentation of qualitative and/or quantitative methods used to solicit and confirm attributes (i.e., concepts measured by the items) of the PRO relevant to the measurement application.		
	Documentation of the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, socio-economic status, literacy).		
	Documentation of sources from which items were derived, modified, and prioritized during the PRO measure development process.		

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
	Justification for the recall period for the measurement application.		
3b.	***Construct and Criterion-related Validity***		
	A PRO measure should have evidence supporting its construct validity, including: • documentation of empirical findings that support predefined hypotheses on the expected associations among measures similar or dissimilar to the measured PRO • documentation of empirical findings that support predefined hypotheses of the expected differences in scores between "known" groups		Patient ratings of satisfaction with arthroplasty were correlated with WOMAC scores in the expected direction.[22,336,337]
	A PRO measure should have evidence that shows the extent to which scores of the instrument are related to a criterion measure.		
3c.	***Responsiveness***		
	A PRO measure for use in longitudinal initiatives should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the target population.	If a PRO measure has cross-sectional data that provides sufficient evidence in regard to the reliability (internal consistency), content validity, and construct validity but has no data yet on responsiveness over time (i.e., ability of a PRO measure to detect changes in the construct being measured over time), would you accept use of the PRO measure to provide valid data over time in a longitudinal	Demonstrates adequate responsiveness and ability to detect change in response to clinical intervention.[338]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
		study if no other PRO measure was available?	
		Important to emphasize responsiveness because there is an expectation of consequences. Need to be able to demonstrate responsiveness if action is to be taken.	
4.	**Interpretability of Scores**		
	A PRO measure should have documentation to support interpretation of scores, including: • what low and high scores represent for the measured concept • representative mean(s) and standard deviation(s) in the reference population • guidance on the minimally important difference in scores between groups and/or over time that can be considered meaningful from the patient and/or clinical perspective	▪ If different PROs are used, it is important to establish a link or cross-walk between them. ▪ Because the criteria for assessing clinically important change in individuals does not directly translate to evaluating clinically important group differences, [306] a useful strategy is to calculate the proportion of patients who experience a clinically significant change[252,306]	Availability of population-based, age- and gender-normative values[339] Availability of minimal clinically important improvement values[340] Can be translated into a utility score for use in economic and accountability evaluations[341]
5.	**Burden**		
	The time, effort, and other demands on the respondent and the administrator.	In a busy clinic setting, PRO assessment should be as brief as possible,	Short form available[342]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
		and reporting should be done in real-time.	Average time to complete mobile phone WOMAC = 4.8 minutes[343]
6.	**Alternatives modes and methods of administration**	The use of multiple modes and methods can be useful for diverse populations. However, there should be evidence regarding their equivalence.	Validated mobile phone and touchscreen based platforms[344,345]
7.	**Cultural and language adaptations**	The mode, method and question wording must yield equivalent estimates of PRO measures.	Available in over 65 languages[346]
8.	**Electronic health records (EHR)**	Critical features: ▪ interoperability ▪ automated, real-time measurement and reporting ▪ sophisticated analytic capacities	Electronic data capture may allow for integration within EHR[343]

Appendix B – Expert Panel Roster

Dr. Richard Bankowitz, MBA, MD, FACP
Chief Medical Officer, Premier healthcare alliance

Dr. Ethan Basch, MD, MSc
Associate Attending Physician and Outcomes Research Scientist- Memorial Sloan-Kettering Cancer Center

Dr. Jim Bellows, PhD, MPH
Senior Director, Evaluation and Analytics- Kaiser Permanente Care Management Institute (CMI)

Dr. Patricia Flatley Brennan, RN, PhD
Professor, School of Nursing and College of Engineering, University of Wisconsin

Ms. Laurie Burke, RN
Associate Director for Study Endpoints and Labeling in the Center for Drug Evaluation and Research, Food and Drug Administration

Ms. Joyce Dubow, MUP
Senior Director, Health Care Reform- AARP

Ms. Jennifer Eames-Huff, MPH
Director, Consumer-Purchaser Disclosure Project- Pacific Business Group on Health

Dr. Stephan Fihn, MD, MPH
Director, Office of Analytics and Business Intelligence, Veterans Health Administration

Dr. Floyd Jackson Fowler, Jr., PhD
Senior Scientific Advisor and Past President- Foundation for Informed Medical Decision Making

Dr. Lori Frank, PhD
Director, Engagement Research- Patient Centered Outcomes Research Institute

Dr. Theodore Ganiats, MD
Professor- University of California San Diego

Dr. Kate Goodrich, MD
Senior Technical Advisor to the Director of the Office of Clinical Standards and Quality and Chief Medical Officer, Centers for Medicare and Medicaid Services

Dr. Judith Hibbard, DrPH
Professor Emerita and Senior Researcher, Institute for Policy Research and Innovation, University of Oregon

Dr. Dennis Kaldenberg, PhD
Chief Scientist, Senior Vice President- Press Ganey Associates

Dr. Irene L. Katzan, MD, MS
Director, Neurological Institute Center for Outcomes Research & Evaluation- Cleveland Clinic

Dr. Lewis Kazis, Sc.D
Professor, Health Policy and Management- Boston University School of Public Health

Dr. Uma Kotagal, M.B.B.S, MSc
Senior Vice President for Safety, Quality and Transformation and Executive Director of the James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center

Dr. Kevin Larsen, MD
Medical Director of Meaningful Use, Office of the National Coordinator

Dr. Kathleen Lohr, PhD
Distinguished Fellow, RTI International

Dr. Elizabeth Mort, MD
Associate Chief Medical Officer, Senior Vice President Quality and Safety, Massachusetts General Hospital

Dr. Charles Moseley, Ed.D.
Associate Executive Director- National Association of State Directors of Developmental Disabilities Services

Dr. Eugene C. Nelson, DSc, MPH
Director, Population Health Measurement Program, The Dartmouth Institute; Director, Population Health and Measurement, Dartmouth-Hitchcock Medical Center; Professor, Community & Family Medicine and of The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Medical School

Dr. Kenneth Ottenbacher, PhD, OTR
Russell Shearn Moody Distinguished Chair in Rehabilitation, University of Texas Medical Branch at Galveston

Dr. Greg Pawlson, MD, MPH
Executive Director, Quality Innovations- BlueCross BlueShield Association Office of Policy and Representation

Dr. Eleanor M. Perfetto, PhD
Senior Director- Pfizer

Ms. Collette M. Pitzen, BSN, RN, CPHQ
Manager, Measure & Program Development- Minnesota Community Measurement

Ms. Cheryl Powell, MS
Deputy Director, Federal Coordinated Health Care Office- Centers for Medicare and Medicaid Services

Dr. David Radley, PhD, MPH
Senior Policy Analyst and Project Director, Institute for Healthcare Improvement

Mr. Ted Rooney, RN, MPH
Project Leader, Maine Quality Counts

Dr. Debra Saliba, MD, MPH
Senior Natural Scientist, The RAND Corporation

Dr. Marcel Salive, MD, MPH
Health Scientist Administrator, Division of Geriatrics & Clinical Gerontology, National Institutes of Health

Dr. Barbara L. Summers, PhD, RN, FAAN
VP, Nursing Practice and Chief Nursing Officer- University of Texas-MD Anderson Cancer Center

Dr. Kalahn A. Taylor-Clark, PhD, MPH
Director, Health Policy- The National Partnership for Women & Families

Dr. Mary Tinetti, MD
Professor of Medicine and Epidemiology, Yale School of Medicine

Ms. Phyllis Torda, MA
Vice President, Quality Solutions Group- National Committee for Quality Assurance

Dr. John Wasson, MD
Emeritus Professor, Dartmouth Medical School

Dr. Robert Weech-Maldonado, PhD
Professor and Chair- University of Alabama at Birmingham

Ms. Linda Wilkinson, MBA
Coordinator of Patient and Family Centered Care, Dartmouth Hitchcock Medical Center

Dr. Albert Wu, MD, MPH
Professor- Johns Hopkins Bloomberg School of Public Health

Appendix C – Commissioned Paper

Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings

Draft manuscript, July 17, 2012
Prepared for NQF PRO Workshop #1 – July 30-31, 2012

David Cella, Elizabeth A. Hahn, Sally E. Jensen, Zeeshan Butt, Cindy J. Nowinski, Nan Rothrock

Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University

I. Introduction

The increasing integration of health care delivery systems provides an opportunity to manage the entire patient-focused episode of care[3] and to assess the impact of care on patient outcomes, including patient-reported outcomes. This paper reviews issues to consider when evaluating patient-reported outcomes (PROs)[5] as candidate performance measures in health care settings.

PROs are defined here as "any report of the status of a patient's health condition, health behavior, or experience with health care that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else (See Figure 1)." In other words, PRO tools measure what patients are able to do and how they feel by direct, unfiltered inquiry. The use of PROs is supported by a large literature that provides cogent evidence suggesting that clinical providers are limited in accurately estimating outcomes for patients.[6-10] PRO tools enable assessment of patient–reported health status domains (e.g., health status; physical, mental, and social functioning; health behavior; experience with health care). A wide variety of patient-level instruments to measure PROs have been used for clinical research purposes and to guide clinical care; many have been evaluated and catalogued in the work conducted by the NIH Patient-Reported Outcomes Measurement Information System (PROMIS; www.nihpromis.org) cooperative group. The PROMIS system itself has not yet been used for performance measurement; however, components of it have been used in the past. There are two major challenges to using PROs for purposes of accountability and performance improvement: 1) they have not been widely adopted in clinical use, and are therefore not familiar to many providers and payers; and 2) little is known about the best set of responsive questions to aggregate for the purpose of measuring *performance* of the health care entity.

While there has been great interest in moving toward use of PROs, foundational work needs to be undertaken to address methodological and data challenges. Efforts are currently underway to develop and test mechanisms for collecting patient-reported data, so this is an opportune time to also consider methodological issues, including collection of PRO data in the clinical environment and the aggregation of the data to assess organization/provider-level performance.

The purpose of this white paper is to address the major methodological issues related to the selection, administration and use of PROs for individual patients in clinical practice settings.

This will inform the selection of PROs for use in performance measures. This paper will also identify best practices in the context of identifying and using PROs in performance measures. A separate white paper will outline the path to developing reliable and valid performance measures eligible for NQF endorsement that can be used for accountability and to inform quality improvement.

Figure 1. Definitions and key concepts that are central to the purpose of this paper

Patient-reported outcome (PRO): Any report of the status of a patient's health condition, health behavior, or experience with health care that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.

PRO measure/instrument: A standardized tool to assess health condition (e.g., health status; physical, mental, and social functioning), health behavior, or experience with health care).

Performance measure: Numeric quantification of health care quality for a designated accountable health care entity, such as hospital, health plan, nursing home, clinician, etc.

PRO-based performance measure: A performance measure that is based on patient-reported outcome data aggregated for an accountable health care entity (e.g., percentage of patients in an accountable care organization with an improved depression score as measured by a standardized tool).

e-health[1,2]**:** Health-related Internet applications that deliver a range of content, connectivity and clinical care. Examples include: online formularies, prescription refills, test results, physician-patient communication.

Patient-Centered E-Health (PCEH)[4]**:** Combination of three themes: (1) Patient-focus (developed primarily based on needs and perspectives of patients); (2) Patient-activity (application designs in which patients can participate meaningfully to provide and consume information about, and of interest to, them); and (3) Patient-empowerment (applications assume that patient want to, and are able to, control far-ranging aspects of their health care via a PCEH application).

Patient-Centered Outcomes Research (PCOR): PCOR is the integration of patient perspectives and experiences with clinical and biological data collected from the patient to evaluate the safety and efficacy of an intervention.

Reliability and Validity: A measure may be reliable (always yields the same score for the same state), but it may not be valid, in that it may be consistently measuring the wrong thing (not measuring what it is supposed to measure). Reliability and Validity are not static characteristics. Demonstrating reliability is essentially accumulating evidence about the stability of the measure, whereas demonstrating validity involves accumulating evidence of many different types which indicate the degree to which the measure denotes what it was intended to represent.

II. Types of PROs

PROs can be used to assess a wide variety of health-relevant concepts, including health-related quality of life, functional status, symptoms and symptom burden, health behaviors, and patient satisfaction. These concepts are neither mutually exclusive nor exhaustive. Table 1 summarizes the main characteristics of these types of PROs.

Health-Related Quality of Life

One type of PRO is health-related quality of life (HRQL). HRQL is a multi-dimensional[11] construct encompassing physical, social, and emotional well-being associated with illness and its treatment.[12] Different types of HRQL measures[13,14] are useful for different purposes.[15] There are a number of generic health status measures such as the SF-36 and Sickness Impact Profile.[16-19] This type of HRQL PRO is useful in assessing both individuals with and without a health condition. This allows for comparisons of groups with and without a specific condition as well as identifying population norms. A health utility or preference measure is also not disease-specific. It provides a score ranging from 0 (death) to 1 (perfect health) that represents the patient's value on his or her own health.[20] This score can be used to calculate quality adjusted life years or compared to population norms.

Many PROs are intended for use in populations with chronic illness.[21-23] Recently, the Patient-Reported Outcome Measurement Information System (PROMIS) has developed a number of PROs in physical, mental, and social health for adults and pediatric samples with chronic conditions.[24,25] Neuro-QOL is another measurement effort focused on capturing important areas of functioning and well-being in neurologic diseases.[26] Each of these measurement efforts do not reference a specific disease in the items and allows for comparisons across conditions. Other PROs are targeted to focus on a specific disease (e.g., spinal cord injury) or treatment (e.g., chemotherapy).[27,28] Often these instruments are developed to be able to demonstrate responsiveness to treatment in a clinical trial rather than compare to population norms or other conditions.[29] Disease-specific PROs often provide additional, complementary information about a patient's HRQL when compared with generic instruments.[22,30-32]

Functional Status

Another type of PRO is a functional status measure. Functional status refers to a patient's ability to perform both basic and more advanced (instrumental) activities of daily life.[33] Examples of functional status include physical function, cognitive function and sexual function. As with HRQL instruments, there are a large number of functional status measure that vary widely in quality.[34] Some may address a very specific type of function (e.g., Upper Limb Functional Index), be developed for use in a specific disease population (e.g., multiple sclerosis), or appropriate for use across chronic conditions.[35-41]

Symptoms and Symptom Burden

Symptoms such as fatigue and pain intensity are also best assessed by PRO measures. Symptoms are typically negative and best assessed through patient report.[42] Scales are

focused on severity. The impact of symptoms such as the degree to which pain interferes with usual functioning, is also a common focus of PROs. Symptom burden captures the combination of both symptom severity and impact experienced with a specific disease or treatment.[42] Common symptom and symptom burden measures include the Functional Assessment of Chronic Illness Therapy – Fatigue scale[43-46] and disease-focused symptom indices.[43-46] The PROMIS initiative developed the PROMIS Pain Interference measure that quantifies the impact of pain on functioning.[43-46]

Health behaviors

Another category of PROs assesses health behaviors. While health behaviors may be considered predictors of health outcomes, they are also health outcomes in their own right in the sense that they may be impacted by health care interventions. The information obtained from health behavior PROs serves several important clinical purposes. Health behavior PROs can be used to monitor risk behaviors with potentially deleterious health consequences. This information enables the identification of areas for risk reduction and health promotion intervention. Health behavior PROs can also be used to assess patients' response to health promotion intervention and for monitoring of health behaviors over time.

The increasing recognition of the impact of preventable unhealthy behaviors on the rising incidence of costly chronic health conditions strengthens the rationale for more widespread use of health behavior PROs. Health behavior PROs are increasingly viewed as important metrics of quality improvement and health outcomes in the clinical setting.[47] Moreover, with the introduction of legislation emphasizing the role of electronic health records (EHRs) in the promotion of patient-centered care, health behavior PROs will constitute an important aspect of future stages of "meaningful use" EHRs.[48,49] This increasing emphasis on health behavior PROs reflects initiatives to shift from a "response to disease" model to a "prevention of disease" model.[50]

As the emphasis on the importance of health behaviors has increased, so has the number of available PROs developed to assess health behaviors across multiple domains. Although many of the available health behavior PROs were originally developed for use in research, they are increasingly being implemented in the clinical setting. PROs measuring aspects of substance use constitute one important category of health behavior PRO tools. A number of substance use PROs have been identified as candidates for use in the clinical setting. Several examples include: the health risk survey, an interactive computer-based health risk survey assessing alcohol consumption and smoking;[51] the CAGE-Adapted to Include Drugs (CAGE-AID), a self-reported screening measure of substance use disorder among treatment-seeking adolescents;[52] the Methadone Treatment Index (MTI), a measure of recent substance abuse, social/behavioral functioning and physical and psychological health for use in methadone maintenance clinics;[47] the alcohol use screener;[53] and the tobacco use screener.[54] In addition to substance use PROs, several PROs have been created to assess other types of risky compulsive behaviors. For example, the Compulsive Internet Use Scale (CIUS)[55] and the Compulsive Sexual Behavior Inventory[56] measure problematic internet use and problematic sexual behavior, respectively. A subset of health behavior PROs also assesses health-promoting behaviors. "Starting the conversation," a brief measure of dietary intake,[57] "Exercise as the fifth vital sign," a brief measure of physical activity,[58] School Health Action, Planning and Evaluation System (SHAPES), a school-based self-report physical activity measure,[59] and the Morisky Medication

Adherence Scale (8 item)[60] constitute several examples of PROs assessing health-promoting behaviors. Sleep quality has emerged as another clinically-relevant health behavior, with several PROs available, including the PROMIS sleep disturbance short form.[61]

Patient Experience of Care

Patient ratings of health care are an integral component of patient-centered care. In its definition of the essential dimensions of patient-centered care, the Institute of Medicine includes shared decision-making among clinicians, patients and families; self-efficacy and self-management skills for patients; and the patient's experience of care.[62,63] Conceptually, measurement of patient ratings is a complex concept that is related to perceived needs, expectations of care, and experience of care.[64-71] Patient ratings can cover the spectrum of patient engagement, from experience to shared decision-making to self-management to full activation. Recognition of patient preferences can help to tailor treatments based on informed decisions. In fact, improving decision quality is one of the most important things that the nation can do to improve quality and outcomes and value. Thus, patient ratings have policy implications and are also of great importance to patients and their families. Each safe practice in the updated NQF consensus report includes a section titled "Opportunities for Patient and Family Involvement."[72]

There are two major types of patient health care ratings: 1) patient satisfaction, and 2) patient reports of their actual experiences. Patient satisfaction is a multidimensional construct that includes patient concerns about the disease and its treatment, issues of treatment affordability and financial burden for the patient, communication with health care providers, access to services, satisfaction with treatment explanations, and confidence in the physician.[73] [74-79] Shikiar and Rentz[69] proposed a three-level hierarchy of satisfaction: 1) satisfaction with health care delivery, including issues of accessibility, clinician-patient communication, quality of facilities; 2) satisfaction with treatment, including medication and other aspects of treatment, e.g., dietary and exercise recommendations; and 3) satisfaction with the medication itself, rather than the broader treatment. Patient satisfaction has important implications for clinical decision-making and improvement in the delivery of health care services, and is increasingly the focus of research and evaluation of medical treatments, services and interventions.[80] It has been shown to be an important indicator of future adherence to treatment. [68,81-86] Satisfaction has a long history of measurement and there are numerous available instruments.[66,71,87-95]

There is a newer focus on measuring patient reports of their actual experiences with health care services.[96] Reports about care are often regarded as more specific, actionable, understandable, and objective than general ratings alone.[97,98] The Consumer Assessment of Healthcare Providers and Systems (CAHPS) program is a multi-year initiative of the Agency for Healthcare Research and Quality (AHRQ) to support and promote the assessment of consumers' experiences with health care (www.cahps.ahrq.gov/About-CAHPS/CAHPS-Program.aspx) The goals of the CAHPS program are: 1) to develop standardized patient questionnaires that can be used to compare results across sponsors and over time, and 2) to generate tools and resources that sponsors can use to produce understandable and usable comparative information for both consumers and health care providers. The CAHPS project

has become a leading mechanism for the measurement of patient perspectives on health care access and quality.[96]

Table 1. Main Characteristics of PROs

PRO Category	Main Characteristics	Strengths	Limitations
HRQL	• Multi-dimensional • Can be generic or disease-specific	• Global summary of well-being	• May not be considered a sufficiently specific construct
Functional Status	• Ability to perform specific activities	• Provide patient-reported data that can be used in addition to performance-based measures of function	• Self-reported capability and actual performance of activities may vary
Symptoms and Symptom Burden	• Specific to type of symptom of interest • Capable of measuring symptoms not otherwise captured by medical work-up	• Best assessed through self-report	• May fail to capture general, global aspects of well-being considered important to patients
Health Behaviors	• Specific to type of behavior • Typically measures frequency of behavior • Available for health risk behaviors as well as health promoting behaviors	• Targeted	• Validity may be impacted by social desirability • Potential patient discomfort in reporting socially undesirable behaviors
Patient Experience	• Satisfaction with health care delivery, treatment recommendations,	• Essential component of patient-centered care	• Complex, multidimensional construct • Confidentiality

	and medications • Actual experiences with health care services	• Valued by patients, families and policymakers • Related to treatment adherence	required to ensure patient comfort in disclosing negative experiences

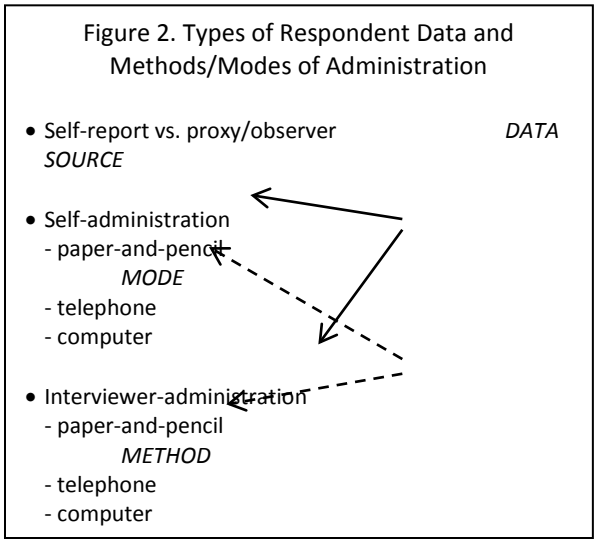
III. Method and mode of administration, data collection and analysis issues

In order to accommodate the needs of patients with diverse linguistic, cultural, educational and functional skills, clinicians and researchers require some flexibility in choosing appropriate methods and modes of questionnaire administration for PROs.[99] There are many issues involved in scoring and analysis of PRO response data. We first describe these methodological issues (see summary in Table 2) and then discuss barriers.

Methodological issues

Administration of PRO instruments requires decisions about three aspects of data collection: the source of the information, the recorder of the information (mode), and the method used to capture the information (see Figure 2). Each of these is described below. These three aspects can also be combined in various ways, e.g., a patient might use the telephone to self-administer a PRO instrument, or an interviewer might use a computer to read questions and record answers.

Figure 2. Types of Respondent Data and Methods/Modes of Administration



Source: Self versus Proxy

The patient's perspective is the focal point of PRO assessment. There are circumstances in which it may be difficult or impossible to directly obtain this perspective. In adults, cognitive and communications deficits and burden of disease, for example, can limit potential subjects' ability to complete PRO questionnaires.[100] This is especially likely to occur with the elderly, with people who suffer from neurological disorders and those with severe disease. Children's participation can be limited by these same factors plus issues specific to age and developmental level.[100-102] Yet, failing to include these populations can result in potentially misleading interpretations of results.

One way to ensure inclusion of the greatest number of patients is to use proxy respondents to obtain PRO information in conjunction with patient reports. Using either significant others (e.g., parents, spouses or other family members, friends) or formal caregivers (physicians, nurses, teachers) as proxies can provide many potential benefits. They not only allow inclusion of a broader and more representative range of patients, they can also help minimize missing data and increase the feasibility of longitudinal assessment. However, the usefulness of proxy responses as substitutes for patient responses depends on the validity and reliability of proxy responses compared to patient responses. When evaluating the quality of proxy responses, proxy responses are usually compared to patient responses. This is a reasonable approach, when proxy responses are being used to replace patient responses. Agreement between proxies and patient pairs is typically assessed at the subscale level via the Intraclass Correlation Coefficient (ICC) or at the item level by the kappa statistic, although other types of analyses have been advocated.[103] Patient and proxy responses are also often compared at the group level by comparing mean scores. Group comparisons help detect the magnitude and direction of any systematic bias that might be present.

Both the adult and pediatric literature suggests that there is greater agreement between proxy and patient ratings when rating observable functioning or HRQL dimensions such as physical and instrumental activities of daily living, physical health and motor function and less for more subjective dimensions such as social functioning, pain, cognitive status/function and psychological or and emotional well-being.[102,104-108] Using continuous rather than dichotomous ratings improves agreement.[109] Extent of disagreement increases with increasing age of adolescents,[110] and as the severity of patient illness, cognitive impairment or disability increases[111-114] Type of proxy (e.g., parent versus caregiver), and proxy characteristics such as age, education, and level of stress may also affect agreement.[115,116] In terms of direction of disagreement, proxies for adults tend to rate them as having more symptoms, functional difficulties, emotional distress and negative quality of life with the exception of pain; where proxies tend to under-report.[104] There is no consistent pattern of disagreement for child versus proxy reported outcomes.[117] Even when there is disagreement for children or adults, differences tend to be small.[117,118]

Proxy assessment may substitute for patient assessment where needed, but may also complement it. Proxies can be asked to assess the patient as they think the patient would respond (i.e. proxy-patient perspective) or for the proxy to provide their own perspective on the patient's functioning or HRQL. This type of rating may be better described as either external- or other-ratings[119] for the sake of clarity. It is important that the measure makes clear which perspective is desired.[117] The external or other perspective may provide particularly relevant information when the person is unable to self-assess, but can be important even when they can. In such cases, patient-other agreement may not be necessarily desirable. This point can be

illustrated in Alzheimer's disease when patients in the earlier stages of dementia fail to recognize the extent of their impaired well-being and physical role functioning compared to family members around them. In such cases, next-of-kin caregivers such as a spouse could provide a different ("external") assessment that indicates the patient has a lot of problems getting the groceries from A to B, or a lot of problems with being comfortable in a social setting, thereby introducing clinically important information.

Mode: Self-administration versus Interviewer-administration

Self-administration of PRO questionnaires is neither expensive nor influenced by interviewer effects, and therefore has traditionally been preferred. However, self-administration is not feasible for some patient populations, such as those who may be too ill to self-administer a questionnaire. In these cases, interviewer-administration is often required. Until recently, interviewer-administration was also required for those with low literacy; however, new multimedia methods are now available to overcome this issue (see below).

Advantages and disadvantages of different modes of administration were summarized by Fowler[120] and Naughton[121] (see Table 2). Self-administered instruments are more cost-effective from a staffing perspective, and may yield more participant disclosure, especially when collecting sensitive information.[122] Disadvantages include the potential for more missing data, and the inability to clarify any misunderstandings. Interviewer-administered instruments allow for probes and clarification, and permit more complexity in survey design (e.g., the use of skip patterns). This mode is also useful for respondents with reading, writing or vision difficulties, and for culturally diverse populations. Disadvantages include the costs required to hire, train and supervise interviewers, and the potential pressure on respondents to answer quickly, without letting them proceed at their own pace. There is also a potential for interviewer bias, resulting in systematic differences from interviewer to interviewer, or, occasionally, systematic errors on the part of many or even all interviewers.[123] Other sources of bias for both administration modes include social desirability (the tendency to give a favorable picture of oneself) and acquiescent response sets (the tendency to agree/disagree with statements regardless of their content).[124,125]

There has been some concern about the potential biasing effects of mode of administration on data quality and interpretation.[126] Overall, there is evidence of high reliability for instruments administered with different modes, but response effects have varied and have not been consistently in the same direction.[121-124] For example, some studies found evidence of more favorable reports of well-being on self-administered questionnaires,[127] while others found the opposite effect.[128-130] Still other studies reported mixed results[131] or found no important differences due to mode of administration, after adjusting for other factors.[120,132,133] Fortunately, many types of error and bias can be overcome by appropriate selection and training of interviewers. Effects of different modes can also be evaluated with various psychometric and statistical techniques and models to determine the potential impact of response effects.[134-138]

Method of Administration

Advances in technology have changed the face of PRO assessment, increasing the number of administration options available. Multiple methods of self-report administration currently exist,

and the different methods may have different effects on the quality of the data.[126] While the different administration methods provide more options for researchers and clinicians, the different methods of administration require different skills and resources of the participant and consequently may result in differing levels of participant burden.[126] A number of factors may account for differences in data quality across methods of administration, including the impersonality of the method, cognitive burden on the participant, ability to establish the legitimacy of the study, control over the questionnaires, and communication style.[126] Thus, these factors must be considered when deciding upon the appropriate method of administration for a given PRO.

Historically, paper-and-pencil administration served as the primary method of PRO assessment. As such, many PROs were originally developed with the intention of paper-based administration, but may be amenable to an electronic-based administration.[139] Paper-and-pencil remains a widely used PRO administration method, with its primary advantage being cost-effectiveness. However, the paper-and-pencil method is not without its disadvantages. For example, it typically requires that a participant's responses be manually entered into a database for scoring purposes, raising the possibility of data entry errors that threaten the integrity of the results. Similarly, the need for manual data entry and scoring can also be time-intensive. This may limit the acceptability of paper-and-pencil administration for purposes in which timely scoring and interpretation is of importance. Finally, paper-based PROs are less likely to provide structured data in EHRs, limiting their usefulness in tracking patient progress over time or influencing change in care plans and, consequently, health outcomes.

Advances in technology and the increasingly widespread availability of electronic resources have provided a number of alternatives to the paper-and-pencil administration method. Advances in telephone technology have enabled the use of interactive voice response (IVR) to administer PROs. IVR involves a computer audio recording of PRO questions administered via telephone to which participants indicate their response by selecting the appropriate key.[126,139] In addition, a number of computer-based administration methods have emerged as feasible alternatives to paper-and-pencil, such as web-based platforms, touchscreen computers, and multimedia platforms that can accommodate people with a range of literacy and computer skills (e.g., Talking Touchscreen/Pantalla Parlanchina, audiovisual computer-assisted self-interviewing[126,139-141]). Newer mobile forms of technology such as tablet computers and smartphones also offer promise as newer generation methods of PRO administration. The electronic administration methods have a number of advantages that contribute to their increasingly widespread adoption. For example, because the participant enters the data themselves, there is minimal chance for data entry errors. These electronic methods also typically allow for immediate scoring and feedback, which lends well to purposes requiring timely results. Furthermore, electronic PRO administration is interactive and has been demonstrated to be practical, acceptable, and cost-effective.[51] Electronic methods may also provide participants with increased comfort when responding to questions about socially undesirable behaviors.[142] However, these advantages must be considered in light of several important disadvantages. First, while electronic PRO administration methods may be cost effective, the cost of purchasing technology-based platforms may exceed that of traditional paper-and-pencil methods. Additionally, some participants may experience discomfort with technology or lack the skills necessary to navigate electronic administration methods. Moreover, reliance upon methods such as web-based platforms or smartphones raises questions about participants' access to these technologies, if they are not provided to them as part of the study.

The availability of multiple methods of PRO administration highlights the importance of measurement equivalence across methods.[139] Measurement equivalence is determined by comparing the psychometric properties of the data obtained via paper-based administration and electronic-based administration.[139] It can be assessed via cognitive testing, usability testing, equivalence testing, or psychometric testing.[139] A growing body of research findings support the equivalence of electronic and paper-and-pencil administration of PROs.[143-145] These findings support the viability of electronic PRO administration as an alternative to paper-and-pencil methods.

In addition to measurement equivalence, patient privacy is another concern that cuts across both paper-and-pencil and electronic administration methods, albeit in differing ways. In the case of paper-based PROs, the physical transfer of the PRO measure from patient to provider, as well as the physical existence of the PRO confers threat to the privacy of patients' responses. Privacy also emerges as a concern with electronic-based methods, given the potential security concerns related to transfer of data or unauthorized access to patient-reported data. This underscores the need for reliable and secure electronic platforms in order to protect patients' privacy in the context of PRO assessment.

PROs in the Clinical Setting

The collection of PRO data as part of clinical care has become more common. Advocates for the use of PROs in clinical care propose that the results assist clinical providers management of patients' care,[146] enhance the efficiency of clinical practice,[145,147] improve patient-provider communication,[145,147-149] identify patient needs in a timely manner,[145,150] and facilitate patient-centered care.[145] However, as PROs are used more in clinical practice, a number of methodological issues pertaining to the settings in which they are administered merit consideration.

A growing number of studies have investigated the use of PROs in the clinic setting.[145,147,149-154] When selecting PROs for administration in the clinic setting, it is important to consider the efficiency of the PRO administration, scoring, and interpretation, given the time-sensitive nature of the clinic flow.[145,151] In addition, the acceptability of the PRO measures and data collection process for both patients and clinic staff is essential.[145,151,155] Historically, several barriers have impeded the widespread implementation of PRO data collection in clinics, many of which are inherent to the drawbacks associated with paper-and-pencil administration of PROs. One such barrier involves concerns about the potential disruption to clinic flow if patients are asked to complete PROs.[146] Conversely, concerns arise regarding the impact of clinic flow on the integrity of data collection, given the potential for patients to be interrupted while completing PROs, which could potentially result in missing data.[146] Another potential barrier involves the possibility that patients may experience anxiety in completing PRO measures in clinic prior to their appointments.[146] Similarly, the lack of privacy when completing PROs in-clinic poses another methodological barrier. Finally, in-clinic collection of PRO data may be impeded by staff burden and clinician disengagement.[146] Fortunately, technology advances, and the increased opportunities for methods of PRO administration that they afford, may help to overcome some barriers to in-clinic PRO data collection.[151] For example, findings support the feasibility of using tablet computers[145,150] and touchscreen computers for in-clinic PRO data

collection.[140,141,149,151,152] The use of computers to administer PROs in-clinic may streamline and expedite the process, as well as minimize staff burden and impact on clinic flow.

Given some of the barriers to PRO data collection in-clinic, completion of PRO measures from home prior to or in between medical appointments has been proposed as one strategy to overcoming barriers to collecting PRO data in-clinic.[146,156,157] Both web-based PRO administration and interactive voice response constitute possible methods for at-home PRO data collection.[151,161,162] While the home may serve as a feasible alternative to the clinic for a number of reasons, there are several factors to consider when implementing home-based PRO data collection completed prior to clinic visit.[146,156] First, in order for patients to be able to complete PRO measures at home, they must have access to the type of technology by which the PRO is administered (e.g., internet). Second, the type of PRO data collected from home must be useful in informing clinical care. Third, the completion of PRO measures at home must be acceptable for patients. Finally, there must be a plan in place to address the reporting of critical or acute problems via home-based PROs. This may pose a logistical challenge in comparison to PROs completed in-clinic, where medical providers and access to intervention is readily available. Several additional barriers to home-based collection of PRO data exist. For example, health information privacy is paramount, and therefore one barrier to home-based PROs is availability of secure data collection platforms.[146 158] As noted, patient safety poses another potential barrier to collection of PRO data at home, given the challenges to addressing critical patient-reported health issues. An additional barrier involves clinician acceptability of home-based PRO data collection, given that questions arise regarding clinician reimbursement for clinician time using a website to address patient-reported outcomes, as opposed to meeting directly with patients to discuss the findings from PROs.[146,158]

Implementation of PRO data collection in other settings, such as rehabilitation or skilled nursing facilities, may also yield valuable clinical information and guide interventions. Less research has addressed the methodological issues involved in administering PROs in these settings. However, handheld technology has been proposed as a means to facilitate collection of PRO data in home health care and the rehabilitation setting following orthopedic surgery.[159] Given the varying level of patients' acuity status in these types of settings, potential factors to consider may include patients' cognitive capacity to complete PRO measures, and whether the use of proxy reports may be beneficial.

Scoring: Classical Test Theory versus Modern Test Theory

PROs are "latent (not directly observable) variables." The only way to estimate a person's level on a particular attribute is by asking questions that are representative of that attribute. Most PRO instruments are comprised of multiple items that are aggregated in some way to produce an overall score that best represents the latent attribute. Scoring is based on classical test theory (raw scores) or modern test theory (item response theory; IRT).[160-169] Multiple items are preferred because a response to a single item provides only limited information to distinguish between individuals.[170] In addition, measurement error (the difference between the "true" score and the "observed" score) tends to "average out" when responses to individual items are summed to obtain a total score.[170-172]

Classical test theory (CTT) estimates the level of an attribute as the sum, perhaps weighted, of responses to individual items, i.e., as a linear combination.[170,173-177] This approach requires all

of the items on a particular PRO instrument to be used in every situation in order for it to be considered valid, i.e., the instrument is "test-dependent"[174,177-179] Item response theory (ITT) enables "test-free" measurement, i.e. the latent trait can be estimated using different items as long as their locations (difficulty levels) have been calibrated on the same scale as the patients' ability levels.[170,176-180 170,181,182] IRT allows computer-adaptive testing (CAT), where questions are tailored to the individual patient. This has two advantages: 1) questionnaires can be shorter, and 2) the scale scores can be estimated more precisely for any given test length. This also means that patients do not need to complete the same set of items in every situation.[176]

There are some challenges to be overcome in order to use IRT. It can be difficult to understand the assumptions and the psychometric jargon, e.g., calibration, difficulty levels. The methodology and software are complex. IRT is also not appropriate for causal variables and complex latent traits.[176-178,183] Overall, though, IRT offers a very convenient and efficient framework for PRO measurement.

Linking/Cross-talk Between Different Measures of the Same Construct

A common problem when using an array of health-related outcomes for diverse patient populations and subgroups is establishing the comparability of scales or units on which the outcomes are reported.[184,185] The emphasis has typically been focused on the metric over the measure. "Equating" is a technique that involves the process of converting the system of units of one measure to that of another. This process of deriving equivalent scores has been used successfully in educational testing to compare test scores obtained from parallel or alternate forms that measure the same characteristic with or without having common anchor items. Theoretically (and in practice when certain conditions are met) different age-specific measures could be linked, thus placing child, adult, and geriatric estimates on a common metric. The many items that constitute a disease-specific (e.g., cancer) quality of life scale could be incorporated into a single shared bank and linked through a common-anchor design.[184] The methods of establishing comparable scores (often called "linking") vary substantially depending on the definition of comparability, and therefore, standardization is critical in facilitating comparing PROs across studies. Two measures may be considered linked if they produce scores that match the first two moments (i.e., mean and SD) of their distributions for a specific group of examinees or two randomly equivalent groups. Another definition may involve matching scores with equal percentile ranks based on a single sample of examinees or random samples drawn from the same population.

Table 2. Main characteristics of key PRO methodological issues

Methodological Issue	**Main Characteristics**	**Strengths**	**Limitations**
Source of report			
Self	• Person responds about him/herself	• Expert on own experience	• Not always possible to assess directly e.g., because of cognitive or

Methodological Issue	Main Characteristics	Strengths	Limitations
			communication deficits or age/developmental level
Proxy	• Person responds about someone else	• Useful when target of assessment unable to respond • Can provide complementary information	• May not accurately represent subjective or other experiences
Mode of administration			
Self	• Person self-administers PRO and records the responses	• Cost-effective • May yield more participant disclosure • Proceed at one's own pace	• Potential for missing data • Simple survey design (e.g., minimal skip patterns)
Interviewer	• Interviewer reads questions out loud and records the responses	• More complex survey design (e.g., skip patterns) • Useful for respondents with reading, writing or vision difficulties	• Interviewer costs • Potential for bias (interviewer bias, social desirability bias, acquiescent response sets)
Method of administration			
Paper-and-pencil	• Patients self-administer PRO using a paper and writing utensil	• Cost-effective	• Prone to data entry errors • Data entry, scoring requires more time • Less amenable to incorporation within EHR
Electronic	• Patient self-administers PRO using computer- or telephone-based platform	• Interactive • Practical • Increased comfort for socially	• Cost • Potential discomfort with technology • Accessibility

Methodological Issue	Main Characteristics	Strengths	Limitations
		undesirable behaviors • Minimizes data entry errors • Immediate scoring, feedback • Amenable to incorporation within EHR	• Measurement equivalence
Setting of administration			
Clinic	• Patients complete PROs when they arrive to clinic appointments	• Real-time assessment of outcomes • Feasibility with use of electronic methods of administration	• Impact on clinic flow • Interruptions resulting in missing data • Patient anxiety • Staff burden
Home	• Patients complete PROs at home prior to, or in between clinic visits	• Minimizes impact on clinic flow • Minimizes staff burden	• Accessibility • Health information privacy • Data security • Patient safety
Other	• Patients complete PROs at other types of settings (e.g., skilled nursing, rehabilitation)	• Feasibility with electronic methods of administration	• Cognitive capacity and potential need for proxy
Scoring			
Classical test theory	• Raw scores	• Easy to implement and understand	• All items must be administered
Modern test theory	• Probabilistic approach	• Enables CAT (tailored questions) • Shorter questionnaires with more precision	• Difficult to implement and understand

Addressing Barriers to PRO Measurement

Several barriers to PRO measurement exist, including administering PROs in vulnerable populations, literacy, language and cultural differences, differences in functional abilities, response shift, use of different methods and modes of administration, and the impact of non-responders. These will each be reviewed below, along with best practices and recommendations for addressing these barriers.

Vulnerable Populations

There is growing recognition that some population subgroups are particularly vulnerable to receiving suboptimal health care and achieving poorer health outcomes compared with the general population. [186-188] Vulnerability is multifaceted and may be because of financial circumstances or place of residence; health, functional, or developmental status; ability to communicate effectively; or age, race, ethnicity, or gender.[186] This definition encompasses populations who are vulnerable because of a chronic or terminal illness or disability and those with literacy or language difficulties.[140,187] It also includes people residing in areas with health professional shortages.[168]

Administration of PRO questionnaires is usually performed with paper-and-pencil instruments, and multilingual versions of questionnaires are often not available. Interviewer administration is labor intensive and cost prohibitive in most health care settings. Therefore, patients with low literacy, those with certain functional limitations, or those who do not speak English are typically excluded, either explicitly or implicitly, from any outcome evaluation in a clinical practice setting in which patient-reported data are collected on forms.

As PROs continue to play a greater role in medical decision making and evaluation of the quality of health care, sensitive and efficient methods of measuring those outcomes among underserved populations must be developed and validated. Minority status, language preference, and literacy level may be critical variables in differentiating those who receive and respond well to treatment from those who do not. These patients may experience different health outcomes because of disparities in care or barriers to care. Outcome measurement in these patients may provide new insight into disease or treatment problems that may have gone undetected simply because many studies have not been able to accommodate the special needs of such patients.[187,189]

Literacy

Low literacy is a widespread but neglected problem in the U.S. The 1992 National Adult Literacy Survey (NALS)[190] and the 2003 National Assessment of Adult Literacy (NAAL)[191] measured three kinds of English language literacy tasks that adults encounter in daily life (prose literacy, document literacy, quantitative literacy). Almost half of the adult population experiences difficulty in using reading, speaking, writing, and computational skills in everyday life situations. An additional seven million adults in the U.S. population were estimated to be non-literate in English. “Health literacy,” the constellation of skills required to function in the health care environment, may be significantly worse than functional literacy because of the unfamiliar context and vocabulary of the health care system.[192]

Contributing to poor understanding of the importance of literacy skills is the fact that low literacy is often underreported. The NALS reported that 66% to 75% of adults in the lowest reading level and 93% to 97% in the second-lowest reading level described themselves as being able to read or write English "well" or "very well."[190] In addition, many low literate individuals are ashamed of their reading difficulties and try to hide the problem, even from their family.[193,194] Lack of recognition and denial of reading problems creates a barrier to health care. Because they are ashamed of their reading difficulties, low literacy patients have acknowledged avoidance of medical care.[193,194] And because there are generally only moderate reading demands in everyday life, individuals may not be aware of their reading problems until a literacy-challenging event (e.g., reviewing treatment options, reading a consent document, completing health assessment forms).[193,194]

A reader's comprehension of text is dependent on the purpose for reading, the ability of the reader, and the text that is being read. Two important factors in the readability of text are word frequency (semantic difficulty) and sentence length (syntactic complexity).[195] Unfamiliar words are difficult when first encountered. Long sentences are likely to contain more clauses, which communicates more information and more ideas. Longer sentences may also require the reader to retain more information in short-term memory.[196-199]

Addressing health literacy is now recognized as critical to delivering person-centered health care.[200] It is an important component of providing quality health care to diverse populations, and will be incorporated into the National Standards for Culturally and Linguistically Appropriate Services.[201] Health literacy practices are also included in the National Quality Forum updated set of safe practices.[72] A recent discussion paper summarized 10 attributes that exemplify a "health literate health care organization."[200] These attributes cover practical strategies across all aspects of health care, from leadership planning and evaluation, to workforce training, to clear communication practices for patients.

Language and Culture

The availability of multiple language versions of PRO questionnaires has enabled them to be routinely measured in diverse research and practice settings. It is often desirable to perform analyses on data that have been pooled across all patients. Yet concern is often voiced regarding combining data from different cultures or languages.[5] In some research and practice-based initiatives, there is interest in evaluating cross-cultural differences in PROs. In all of these applications, it is important to use unbiased questionnaires that can detect important differences between patients.[187,202,203]

Possible cultural differences in interpreting questions and in response styles may limit data pooling or may limit comparisons between members of different cultural groups.[204-206] Similarly, poor quality translations could result in non-comparable language versions of PRO questionnaires.[205,207 205,208] The extent to which items in a questionnaire perform similarly across different groups (e.g., the extent to which they are cross-culturally or cross-linguistically equivalent) is of critical interest when determining whether the questionnaire can be used as an unbiased measure of a PRO.[209-220 203] Without assurances that the PRO questionnaire is culturally and linguistically "fair," detected treatment differences caused by items that function differently across groups could incorrectly be interpreted to reflect real treatment differences. Similarly, true treatment differences may be masked by differences in questionnaire

performance, especially if there is imbalance of language or cultural groups across treatment arms. These possible unwanted effects of cultural or linguistic differences on PRO measurement and outcomes are therefore important at the most basic level.

Functional Abilities

Ideally, PRO instruments that are intended to be used in performance measures are capable of being completed by all patients in the target population. Otherwise, if a significant proportion of the population is excluded, the sample may be unrepresentative and the validity of the performance measure can be compromised. Functional limitations associated with disability are one type of potential barrier to PRO assessment that could affect PRO use in performance measures. The prevalence of disability, defined as specific functional or sensory limitations, is estimated as 47.5 million Americans, or 22% of the U.S. population.[221] People with disability are more likely to develop health conditions and be consumers of health care. Thus, they are an important group to include when evaluating health care but one that is frequently excluded in health research.[222,223]

Common disabilities that can affect PRO assessment include vision (e.g., decreased visual acuity, color-blindness), hearing, motor (e.g. upper extremity limitations) and cognitive deficits (e.g., impaired comprehension, reading). Fortunately, many of these barriers can be addressed by choice of method and mode of data collection, by enabling the use of Assistive Devices/Technology, and by using principles of universal design when developing instruments [201-202]. Universal Design refers to designing products and environments in such a way as to be usable by all people, to the greatest extent possible, without adaptation or specialization.[224,225] A well-known example of universal design is the use of curb cuts. Initially intended to facilitate the use of wheelchairs, they have also benefited bicycle riders and children in strollers, amongst others. An exhaustive examination of how the principles of universal design can be applied to PRO assessment is beyond the scope of this paper, and those developing or modifying measures according to the principles of universal design are encouraged to consult with relevant experts. Also ,if developing an information-technology based instrument, using the standards included in Section 508 of the Rehabilitation Act Amendments of 1998 can maximize flexibility.[226] While we cannot list all potential ways to address functional limitations, in the next paragraph we identify some common ways to do so. Harniss and colleagues provide a description on how PROMIS is taking a systematic approach to enhancing accessibility.[227]

In general, it is important to provide multiple means of understanding and responding to measures including visual, voiced and tactile. The specific means may differ depending on the method and mode of administration. Thus, for people with impaired vision one might consider using in-person or telephone interviews (advantages and disadvantages discussed in an earlier section), an Integrated Voice Response system, Braille responses for Braille users, or touchscreen with tactile or audio cues. Information technology-based systems should enable assistive devices such as screen readers and screen-enlargement software. For the hearing impaired, options include providing visual presentation of words or images, use of TTY or a Video Relay Service, and allowing the user to adjust the sound level. For those with motor limitations, response modes that are easier to manipulate (track ball) or non-motoric (e.g. using voice recognition software) can be helpful. For those with certain types of cognitive deficits (e.g., limited reading comprehension) the methods to address literacy described earlier should be

considered. However, if cognitive deficits are severe it may be more appropriate to use a proxy respondent.

Concerns have been raised that allowing for multiple response modes or methods may lead to measurement error. In a later section, we discuss the potential impact of different methods and modes on response rate, reliability and validity. The risk of introducing measurement error seems outweighed by the risk of excluding a significant segment of the population.

Response Shift, Adaptation, and Other Challenges to Detecting True Change

The ability to detect true change over time in PROs poses another barrier to the integrity of PRO assessment. Often, the ability to detect true change is attributable to the phenomenon of response shift, which has been defined as, "A change in the meaning of one's self-evaluation of a target construct as a result of (a) a change in the respondent's internal standards of measurement (i.e., scale recalibration); (b) a change in the respondent's values (i.e., the importance of component domains constituting the target construct); or (c) a redefinition of the target construct (i.e., reconceptualization)."[228] A change in perspective over time may result in participants attending to PROs in a systematically different way from one time point to another.[229]

Response shift serves as a barrier to PRO assessment for several important reasons. For example, it threatens longitudinal PRO assessment validity, reliability, and responsiveness.[229-232] Response shift can complicate the interpretation of PRO outcomes, since a change in PRO outcome may occur due to a response shift, an effect of treatment, or both.[233]

Monitoring for response shift can aid in interpretation of longitudinal PRO data.[231] A number of strategies have been proposed to identify response shift, although each has limitations. The "then test" compares an actual pre-test rating and a retrospective pre-test rating to assess for shift, but is less robust than other methods of detecting response shift,[229] and is confounded with recall bias.[232] Structural equation modeling (SEM) has also been proposed as a method to identify response shift; however, it is sensitive only if most of the sample is likely to make response shifts.[234] Finally, growth modeling represents another potential strategy for identifying response shift. Growth modeling creates a predictive growth curve model to investigate patterns in discrepancies between expected and observed scores, thus assessing response shift at the individual level.[235] Although growth modeling enables detection of both the timing and shape of response shift,[231] it cannot differentiate between random error and response shift.[232]

Implications of the Different Methods and Modes on Response Rate, Reliability and Validity

Decisions About the Choice of Data Collection Methods

Decisions must be made related to the data collection method and the implications of those decisions on costs and errors in surveys.[122] Two basic issues underlie these decisions: (1) "What is the most appropriate method to choose for a particular question?" and (2) "What is the impact of a particular method on survey errors and costs?" Different methods differ along a variety of dimensions,[122] including the degree of interviewer involvement, the level of interaction with the respondent, the channels of communication used (sight, sound, touch; various

combinations may yield different issues of comprehension, memory stimulation, social influence affecting judgment, and response hurdles), and the degree of technology use.

Implications of Using a Different Method/Mode than Originally Validated

It is also necessary to consider the implications of using a different method or mode than with which the PRO was originally validated. Many existing PROs were initially validated in paper-and-pencil form. However, potential differences exist between paper-and-pencil and electronic-based PRO administration,[143] ranging from differences in how items/responses presented (e.g., items presented one at a time, size of text) to differences in participant comfort level in responding (e.g., ability to interact with electronic –based platform).[143] As noted earlier, a growing body of research suggests measurement equivalence between paper- and computer-administered PROs.[143,236] However, the effect of a particular data collection method on a particular source of error may depend on the specific combination of methods used.[122] Thus, as new methods are developed, studies comparing them to the methods they may replace must be done. Theory is important to inform our expectations about the likely effect of a particular approach. Theory is informed by past mode-effects literature as well as by an understanding of the features or elements of a particular design.[122] Similarly, mode choices involve trade-offs and compromises. As such, the choice of a particular approach must be made within the context of the particular objectives of the survey and the resources available.[122]

Implications of Using Multiple Methods/Modes

The implications of using multiple methods and modes also warrant consideration. There are a number of reasons why one might choose to blend methods, such as cost reduction, faster data collection, and optimization of response rates.[122] When combining methods/modes, it is critical to ensure that any effects of the method/mode can be disentangled from other sample characteristics. This is especially true when respondents choose which method/mode they prefer, or when access issues determine the choice of method/mode.[122] As in the case of using a different method/mode than originally validated, instruments and procedures should be designed to ensure equivalence across methods/modes.[237]

Impact of Non-responders

Difficulties with data collection and compliance are major barriers to the successful implementation of PRO assessment. The principal problem is that bias may be introduced through data that are missing.[176] The choice of mode and method of questionnaire administration can affect nonresponse rates and nonresponse bias.[122] In addition, retrospective collection of PRO data is rarely possible, and often the timing of the assessment is important, e.g., prior to or just after surgery.

Missing data may be classified as either item non-response (one or more missing items within a questionnaire), or unit non-response (the whole questionnaire is missing for a patient). It is important to evaluate the amount, reasons and patterns of missing data.[238-241] Some common strategies to evaluate non-response bias are listed here:

- Conduct an abbreviated follow-up survey with initial non-respondents[122]
- Compare characteristics of respondents and non-respondents[242,243]
- Compare respondent data with comparable information from other sources[244]

- Compare early vs. late respondents[245]

When dealing with missing data, there are various statistical methods of adjustment. For item non-response in multi-item scales, several techniques are useful and tend to yield unbiased estimates of scores, e.g., simple mean imputation, regression imputation, and IRT models. For both item and unit non-response it is important to determine whether missing data are considered to be missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR).[238,239] For unit non-response, there is a range of statistical techniques that could be implemented, depending on the reason for missing data.[246-250]

IV. Selection of patient-level PRO measures

Patient-Centered Outcomes Research

An essential aspect of patient-centered outcomes research (PCOR) is the integration of patient perspectives and experiences with clinical and biological data collected from the patient to evaluate the safety and efficacy of an intervention. Such integration recognizes that while traditional clinical endpoints such as laboratory values or survival are still very important, we also need to look at how patients' health-related quality of life (HRQL) is affected by the disease and treatment. For such HRQL endpoints, in most cases, the patient is the best source for reporting what they are experiencing. The challenge is how to best capture patient data in a way that maximizes our ability to inform decision making in the research, healthcare delivery, and policy settings.

Access to psychometrically sound and decision-relevant PRO will allow investigators to collect the empirical evidence on the differential benefits of a study intervention.[251-254] These data can then be disseminated to patients, providers and policy makers to provide a richer perspective on the impact of interventions on patients' lives using endpoints that are meaningful to the patients.[255] Increasingly, longitudinal observational and experimental studies have included PRO measures. In order to optimize decision making in clinical care, these PROs must be measured in a standardized way using questionnaires that demonstrate specific measurement properties.[251,254,256-259] Our group recently identified minimum standards for the design or selection of a PRO for use in patient-centered outcomes research.[260] Central to this work was an understanding of the critical attributes for which a PRO is judged to be appropriate or inappropriate for such purposes. We identified these standards through two complementary approaches. The first was an extensive review of the literature including both published and unpublished guidance documents. The second was to assemble a group of international experts in PRO and patient-centered outcomes research to seek consensus on the minimum standards.[260]

Common Themes and Lessons Learned

There exist many documents summarizing attributes of a good HRQL measure, including guidance documents from the FDA; [261-264] the 2002 Medical Outcomes Trust guidelines on attributes of a good HRQL measure;[265] the extensive, international expert-driven recommendations from COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments);[257,266-270] the European Organization for Research and Treatment of Cancer (EORTC) guidelines for developing questionnaires; [271] the Functional Assessment of

Chronic Illness Therapy (FACIT) approach;[28] the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) task force recommendation documents;[139,213,272,273] and several others.[217,256,274-276] There is also a standards documents just released by the NIH Patient-Reported Outcomes Measurement Information System® (PROMIS®) network, which we considered useful for informing the minimal and optimal standards for designing PRO measures. In addition, the ISOQOL recently completed two guidance documents on use of PROs in comparative effectiveness research and on integrating PROs in healthcare delivery settings that were relevant for this landscape review.

The selection of PROs for use in performance measurement raises the question of what are the key differences, if any, when selecting PROs for research purposes as opposed to performance measurement. Generally speaking, the factors to consider when selecting PROs for research versus performance measurement are more similar than different. One key difference to consider involves the length of the PRO. Although longer PROs with more items may be better tolerated in the context of research, the feasibility and acceptability of using PROs for performance measurement demands shorter instrument length to facilitate widespread adoption. The need for shortened PRO measure length for use in performance measurement may compromise other important measurement characteristics, such as measurement precision. Another key difference in factors to consider when selecting PROs for performance measurement versus research is the implication or consequence of the PRO data. Specifically, the use of PROs for performance measurement carries the expectation that there will be consequences in terms of public reporting and accountability for the clinical providers or clinical setting. Therefore, the stakes of PROs are higher in the performance measurement context, but there may be constraints to the quality of the measurement level due to factors unique to performance measurement, such as length. This highlights the importance of emphasizing responsiveness/sensitivity to change when considering PROs for use in performance measurement.

In selecting a PRO for performance measurement purposes, a logical first step involves a review of what measures have been used successfully previously. While the use of PROs in performance measurement remains an under-studied area, several examples of PROs used as indexes of performance measurement provide an initial foundation upon which the field can expand. This is most clearly illustrated by the Veterans Health Study, which was developed to assess patient-reported outcomes within the VA system.[277] In response to the Veterans Health Administration's incorporation of patient-reported functional status as a domain of interest in their performance measurement system, the Veterans RAND 36 Item Health Survey (VR-36) and the Veterans Rand 12 Item Health Survey (VR-12) have been administered within the VA system to evaluate veterans' needs as well as to assess outcomes of clinical care at the hospital, regional, and healthcare system levels.[277,278] These methods have also been applied for performance measurement in the Medicare Advantage Program[279] and the VR-12 has also been designated as the principal outcomes of the Medicare Health Outcomes Survey (HOS).[280]

While the research examining the VR-36 and SF-36 in patient-reported performance measures informs the selection of PROs for performance measurement, there are limitations to the use of these measures as indexes of performance and accountability. These limitations include their "static" nature which requires all items to be administered in order to receive a score, even if some items add little to the precision of measurement. In addition, content is fixed by the composition of the scale. As such, attention has turned to alternative PROs with the

potential for use as patient-reported performance measures. The PROMIS measurement system constitutes one example of a future direction of PROs acceptable for use in performance measurement. Developed using IRT methodology, PROMIS offers a new generation of PRO measures with improved reliability, validity, precision, and shortened length.[167] PROMIS PRO measures form a hybrid between static generic PROs and more flexible adaptive measures that are comprised of items specific to measure content, but applicable across the diverse spectrum of health status. Although a growing body of literature provides preliminary evidence supporting the psychometric adequacy of the PROMIS measures, future work is needed to explore the application of PROMIS measures as performance measure PROs. Nevertheless, the PROMIS system provides a model by which the use of PROs as performance measures can be expanded and elaborated upon, owing to its rigorous methodological characteristics.

(Please see Table 3 for characteristics and best practices to evaluate and select PROs for use in performance measurement.)

Documentation, in peer-reviewed literature and/or on publically accessible websites, of the evidence of a PRO to reflect these measurement properties will result in greater acceptance of the PRO for use as performance measures. To the extent the evidence was obtained from populations similar to the studies' target population, the more confidence the investigator will have in the PRO to capture patient's experiences and perspectives.

There are a number of considerations when applying any set of selection standards for PROs. The populations participating in research will likely be quite heterogeneous. This population heterogeneity should be reflected in the samples that participate in the evaluation of the measurement properties for the PRO. For example, both qualitative and quantitative studies may require quota sampling based on race/ethnicity that reflects the prevalence of the condition in the study target population.

Literacy demand is also an important consideration for use of PROs. Data collected from PRO measures is only valid if the participants in a study can understand what is asked of them and can provide a response that accurately reflects their experiences or perspectives. It is critical that developers of PRO measures be attentive to make sure the questions and response options are clear and easy to understand. Pre-testing of the instrument (e.g., cognitive testing) should include individuals with low literacy to evaluate the questions.[281]

Response burden must be considered when selecting a PRO measure and using it in a PCOR study. The instrument must not be overly burdensome for patients as they are often sick and cannot be subjected to long questionnaires or be asked repeatedly to provide repeated, longitudinal data that may significantly disrupt their lives.

Finally, researchers much carefully consider the strength of evidence for the measurement properties. There is no threshold for which an instrument is valid or not valid for any or all populations or applications. In addition, there can be no single study that confirms all the measurement properties for all contexts. Like any scientific discipline, measurement science relies on an iterative, accumulating body of evidence examining key properties in different contexts. Thus, it is the weight of the evidence that informs the evaluation of the appropriateness of a PRO. Older PROs will have the benefit of having more evidence than more

recent PROs; yet the newer PROs tend to have improved basic measurement properties that warrant attention.

PRO Characteristics for Consideration

Global versus Condition-specific Measures

One of the primary factors to consider when selecting a patient-level PRO measure is whether to use a global versus a condition-specific PRO. Several elements inform the selection of global versus condition-specific measures.[282] The specific population of interest may guide whether one opts to use a global or condition specific PRO. For example, if the target population is largely comprised of healthy individuals, a global measure may be the preferred choice; conversely, if the goal is to examine a specific subset of patients with a particular health concern, then a condition-specific measure may be more appropriate. Similarly, the outcomes of interest may guide the selection process given that global measures may capture a different category of outcomes when compared to a condition-specific PRO. Additionally, the assessment purpose will likely influence the selection of global versus specific measures. An excellent example of this stems from guidance from the Food and Drug Administration guidance stating that pharmaceutical company claims of improved QOL must be specific to the QOL domain that was measured, with the recommended that the assessment of specific symptoms is an appropriate starting point for improved measurement of QOL domains.[283]

Global PRO measures have several important advantages. They allow for comparability across patients/populations,[282] although they are more suitable for comparison across groups than for individual clinical use.[284] Global PROs also allow for normative data which can be used to interpret scores.[282] This enables comparison to population norms or directly with other disease conditions. They can also be applied to individuals without specific health conditions, and can differentiate groups on indexes of overall health/well-being.[282] In spite of these advantages, global PROs also have several disadvantages. They have tended to be less sensitive to change and therefore may underestimate health changes in specific patient populations.[285] Additionally, they may fail to capture important condition-specific concerns[285] when applied in specific disease populations.

Condition-specific PROs serve as an alternative to global PROs. One advantage of condition-specific PROs is greater sensitivity to change, because they focus on the concerns pertinent to the given condition.[282] They also enable differentiation of groups at level of specific symptoms/concerns.[282] However, given their condition-specific focus, one notable limitation is the difficulty in making comparisons across patient/disease populations.[282]

Given their respective unique benefits and limitations, the use of a combination of global and condition-specific measures is recommended. Global and condition-specific PRO measures may measure different aspects of QOL when administered in combination,[286] resulting in more comprehensive assessment. Consequently, hybrid measurement systems have emerged to facilitate the combination of global and condition-specific PROs. For example, the FACIT system consists of a generic HRQL measure plus condition-specific subscales. The PROMIS measurement system, which was developed to create item banks that are appropriate for use across common chronic disease conditions,[287] represents another example of a hybrid system of PROs that combine both global and targeted approaches.

Measurement Precision

Another factor to consider when selecting a patient-level PRO measure is measurement precision. Measurement precision refers to the level of variation in multiple measurements of the same factor, such that measures with greater precision have less variation across measurement time points. PROs with greater measurement precision also demonstrate greater sensitivity to change.[288] Given that most PROs were originally developed as research outcome measures, they may lack the level of precision necessary for assessment of individuals.[289] Although performance measures will aggregate to provider or organization, adequate measurement precision at the patient-level is still needed.

When considering measurement precision in the selection of PROs, measures based on IRT tend to have greater precision than measures based on classical test theory.[289] Specifically, computerized adaptive tests (CATs) offer greater precision than static short-forms derived from item banks; however, short forms are acceptable alternative when CAT is not feasible.[290,291] Although CATs include a greater number of items in an item bank, they allow tailored measurement, resulting in shorter instrument length and improved precision. Consequently, the use of PROs derived from IRT methodology is recommended in order to achieve the greatest measurement precision.

Sensitivity to Change/Responsiveness

Sensitivity to change constitutes another important factor to consider when selecting a PRO measure because the ability to detect a small, but important change is necessary when monitoring patients and implementing clinical interventions.[30] Sensitivity to change is a component of construct validity characterized by within subject changes over time on the PRO following an intervention.[292,293] There is great variation in how responsiveness is conceptualized, resulting in different findings and interpretations.[294] Definitions of sensitivity to change range from the ability to detect any kind of change, regardless of meaningfulness (e.g., a statistically significant change post-treatment), to the ability to detect a clinically important change. In order to be clinically useful, PROs must demonstrate sensitivity to change when individuals improve as well as when they deteriorate.[293]

Just as great variation exists in how responsiveness is defined, there is also great variation in the methods for assessing responsiveness. These methods primarily differ in terms of whether they are intended to demonstrate statistically significant changes versus quantify the magnitude of change.[294] The lack of equivalence across methods for detecting change can be problematic for interpretation, given that the different methods for detecting responsiveness produce different classifications of who is improved or not.[295] However, solely relying on statistical tests of responsiveness is not recommended, given that it may not accurately reflect what is meaningful to patient or clinician.[296]

There are a number of factors which may limit a PRO measure's sensitivity to change. First, the use of multi-trait scales containing items that are not relevant to the population being assessed may fail to capture change over time.[297] The responsiveness of a PRO measure may also be constrained by the use of scales that offer categorical or a limited range of response options.[297] PRO measures that utilize an extensive timeframe for reporting also will not be likely

to demonstrate change if administered regularly over a brief period of time.[297] The responsiveness of a PRO measure is also limited by the inclusion of items that reflect stable characteristics which are unlikely to change as well as scales that contain items with floor or ceiling effects.[297] It is also important to note that a PRO measure's sensitivity to change may depend upon the direction of the change. For example, Eurich and colleagues found that PROs were more responsive to change when there was clinical improvement, relative to deterioration.[30]

In addition to those factors, a growing body of research suggests that condition-specific PROs are more sensitive to change than generic PROs.[30,32,298-300] This reflects the fact that responsiveness to change is likely impacted by the purpose for which the measure was originally developed.[300] For example, measures developed to emphasize specific content areas would be expected to show greater change secondary to treatment in those content areas.[293] Thus, the greater sensitivity to change in condition-specific PROs is likely due to the strong content validity inherent in disease-specific measures. [30] As a result, the use of a combination of disease-specific and generic PRO measures may yield the most meaningful data.[30,32]

Minimally Important Differences and Changes

The difference between clinical versus statistical significance also merits consideration when selecting a PRO measure. Historically, research has relied upon tests of statistical significance to examine differences in PRO scores between patients or within patients over time. However, concerns arise regarding whether statistically significant differences truly reflect differences that would be perceived as important to the patient or the clinician. Consequently, attention has shifted to the concept of clinically significant differences in PRO scores. A variety of approaches to determining clinical significance have been proposed. For example, clinically significant change has been defined as "changes in patient functioning that are meaningful for individuals who undergo psychosocial or medical interventions."[301] Similarly, meaningful change is defined as ""one that results in a meaningful reduction in symptoms or improvement in function…" [from the patient perspective].[302] Minimally important differences (MIDs) represent a specific approach to clinical significance, and are defined as "…the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important."[303] Finally, minimum clinically important differences (MCIDs) comprise an even more specific category of MID and are defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management."[304]

The examination of clinically significant differences carries a number of important implications.[303] First, investigating clinically significant (versus statistically significant) differences in scores aids in the interpretation of PROs. Second, the focus on clinically significant differences also emphasizes the importance of the patient perspective, which may not be adequately captured when strictly looking at statistically significant differences. Third, the ability to look at clinically significant differences in PRO scores informs the evaluation of the success of a clinical intervention. Finally, in the context of clinical research, clinically significant differences can assist with sample size estimation.

Currently, no methodological "gold standard" exists for estimating MIDs;[302,305] however, two primary methods are currently in-use: the anchor-based method and the distribution-based

method. The anchor-based method of establishing MIDs assesses the relationship between scores on the PRO and some independent measure which is interpretable.[303] Several options exist for the type of anchor selected when using the anchor-based method. First, clinical anchors which are correlated with the PRO measure at the r ≥ 0.30 level may serve as appropriate anchors. [276,306] Clinical trial experience can be used to inform the selection of these clinical anchors,[307] which also enables the use of multiple clinical anchors.[308] Transition ratings represent another potential source of anchors when establishing MIDs. Transition ratings are within-person global ratings of change made by a patient.[306,309] However, due to concerns about validity, it is recommended that researchers examine the correlation between pre-and post-test PRO scores and the transition rating.[310] Between-person differences made by patients can also be used as anchors when establishing MIDs for PRO measures.[314,317] Additional sources for anchors when establishing MIDs include HRQL-related functional measures used by clinicians [306,309] and objective standards (e.g., hospital admissions, time away from work.[310] Although the anchor-based method offers promise for establishing MIDs in PRO measures, several limitations should be considered. First, the transition rating approach to anchor selection is subject to recall bias on the part of the patient. [302] Second, global ratings may only account for some variance in PRO scores.[302] Third, the anchor based method does not take into consideration measurement precision of instrument.[302]

The distribution-based method represents the second method of establishing MIDs in PRO measures. The distribution-based method uses the statistical characteristics of the PRO scores when establishing MIDs.[303] Specifically, the distribution-based approach evaluates change in scores in relation to the probability that the change occurred at random.[302] As in the case of the anchor-based method, there are several methods available when applying a distribution-based approach to MID establishment. First, the t-test statistic has been used to establish MID when examining change over time.[302] However, given that this relies solely on statistical significance, it may not reflect change that is clinically meaningful and it is also subject to variation due to sample size.[302] Distribution-based methods may also be grounded in measurement precision and the standard error of mean (SEM). [302] Specifically, it has been suggested that the 1 SEM criterion can be used as an alternative to MID when assessing the magnitude of PRO score changes.[311] Sample variation, such as effect size and standardized response mean, constitutes another method for establishing MIDs using the distribution-based method.[302] When using this method, it is recommended that the effect size by specific to the population being studied.[309] Evidence suggests that MID estimates using sample variation are approximately half of a standard deviation.[312] Finally, reliable change constitutes another method of using the distribution-based approach to establishing MIDs.[302] Reliable change is based on the standard error of measurement difference (SEMD) and indicates how much the observed change exceeds fluctuations in an imprecise measure that are random in nature.[302] While the distribution-based approach serves as a possible alternative to the anchor-based methods, there is little consensus on the benchmarks for establishing changes that are clinically significant.[302]

Given limitations of the anchor- and distribution-based approaches, it is recommended that multiple methods and triangulation should be used to determine the MID.[276,302,312] Moreover, the final selection of MID values should be based on systematic review and an evaluation process such as the Delphi method.[276] When considering MIDs for PRO measures, a single MID should not be applied to situation involving that particular PRO, given that MID varies by population/context.[276] Consequently, it is recommended that the distribution around the MID be

provided rather than just a single MID value.[308] Finally, because the criteria for assessing clinically important change in individuals do not directly translate to evaluating clinically important group differences,[306] a useful strategy is to calculate the proportion of patients who experience a clinically significant change.[252,306]

Essential Conditions to Integrate PROs into the Electronic Health Record

Health information technology (HIT) has the potential to enable dramatic transformation in health care delivery, but the empirical research evidence base supporting its benefits is limited.[313]

E-health refers to health-related Internet applications that deliver a range of content, connectivity and clinical care.[1] This includes health information, online formularies, prescription refills, appointment scheduling, test results, advance care planning and health care proxy designation, and physician-patient communication.[314] Patient-Centered E-Health (PCEH) is an emerging discipline that is defined as the combination of three themes:[4]

- Patient-focus: PCEH applications are developed primarily based on needs and perspectives of patients.
- Patient-activity: PCEH application designs assume that patients can participate meaningfully in providing and consuming information about, and of interest to, them,
- Patient-empowerment: PCEH applications assume that patients want to, and are able to, control far-ranging aspects of their health care via a PCEH application.

Although e-health applications have become common, they tend to focus on the needs of health care providers and organizations. Patients desire a range of services to be brought online by their own health care provider.[315] However, there is little evidence about whether the services offered by providers are services that patients desire.[2] It is important that providers attend to patient acceptability factors.[2,316]

Measurement of PROs will constitute an important aspect of future stages of "meaningful use" of electronic health records (EHRs).[48,49] There is the potential for enhanced access by allowing entry directly from commonly used devices such as smart phones. Enabling clinical decision support by providing structured data directly into EHRs will permit PROs to (a) be used for tracking patient progress over time, or (b) use individual question responses to drive change in care plans or care processes concurrently thus improving outcomes over time. The use of a standardized instrument registered in an established code system (e.g., LOINC) enables EHRs to incorporate the instrument as an observation with a known set of responses using standard terminology (SNOMED-CT) or numerical responses. Each question in the standardized instrument can also be coded (structured) to drive changes based on those responses.

Unfortunately, in an updated systematic review of health information technology studies published during 2004-2007, PROs were not mentioned at all.[314]

The passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act creates a mix of incentives and penalties that will induce a large proportion of physicians and hospitals to move toward EHR systems by the end of this decade.[317] The discussion should now focus on whether HIT will support the models of care delivery that will help achieve broader policy goals: safer, more effective, and more efficient care.

Three features of EHRs are critical to enable accountable care organizations to succeed: interoperability and widespread health information exchange; automated, real-time quality and cost measurement; and smarter analytic capacities. Having a complete picture of the patient's care is a critical start, yet most EHRs are not interoperable and have limited data-sharing capabilities.[318] In summary, important issues include: a) the patient perspective (patients want to be involved "as a participant and partner in the flow of information" relating to their own health care[319]); b) clinical buy-in; c) compatibility with clinical flow; and d) meaningful use.

Examples. Health care centers are beginning to implement ways to use patient-reported information ("the voice of the patient") to provide higher quality care.[320] Three recent case studies (two in the U.S. and one in Sweden) are particularly informative to illustrate "lessons learned" about such initiatives.[320] The Dartmouth Spine Center collects health survey data from patients before each visit, either at home or in the clinic. The data are summarized in a report and are available for use by the patients and clinicians to develop or modify the care plan, and to monitor results over time to guide treatment decisions. Longitudinal changes are incorporated into the report with each new assessment. The Karolinska University hospital (Stockholm, Sweden) developed a Swedish Rheumatology Quality registry in 1995 to improve the quality and value of care for people suffering from arthritis and other rheumatic diseases. Paper forms have now been replaced with a web-based system that makes use of real-time data provided by patients, clinicians and diagnostic tests. Longitudinal summaries of PRO measures and other health information are incorporated into graphical reports that are available to patients and providers. An electronic Health Risk Assessment has been integrated with an electronic health record at Group Health Cooperative in the State of Washington. Patients can complete PRO measures, make appointments, fill prescriptions, review health benefits, communicate with their providers, and get vetted health information. Customized reports are available to patients and providers.

Both patients and clinicians have generally favorable reactions to the patient-reported measurement systems implemented in these three very different health care settings. The information gathered helps to support patient-centered care by focusing attention on the health issues and outcomes that are important to patients. Although both patients and clinicians acknowledge that using PROs takes extra time for data collection, both groups report that it makes the care more effective and efficient. Key design principles to successful use of patient-reported measurement systems include fitting PRO measures into the flow of care, designing the systems with stakeholder engagement, merging PRO data with other types of data (clinician reports, medical records, claims), and engaging in continuous improvement of the systems based on users' experiences and new technology.

Other examples can be found in the use of PROs in the management of advanced cancer where the primary goals of care are to maximize symptom management and minimize treatment toxicity. Clinicians and patients often base treatment decisions on informal assessments of health-related quality of life (HRQL). Integrating formal HRQL assessment into treatment decision-making has the potential to improve patient-centered care for advanced cancer patients. Computer-based PRO assessment can reduce patient and administrative burden while enabling real-time scoring and presentation of HRQL data. Two pilot studies conducted with advanced lung cancer patients reported that the computer technology was acceptable and feasible for patients and physicians.[151,321] Patients felt that the HRQL questionnaire helped them

focus on issues to discuss with their physicians, and physicians indicated that the HRQL report helped them to evaluate patient responses over time.

A new initiative in the Robert H. Lurie Comprehensive Cancer Center involves the development and implementation of patient-reported symptom assessment in Gynecologic Oncology clinics. Prior to clinic visits, outpatients complete instruments measuring fatigue, pain, physical function, depression and anxiety through the electronic health record (EHR) patient communication portal at home or in-clinic using an iPad. Results immediately populate the EHR. Severe symptoms trigger EHR notifications to providers. The EHR provides automated triage for psychosocial and nutritional care when indicated.

Selection of PROs that Meet the Recommended Characteristics for use in Performance Measures

A number of characteristics have been recommended when evaluating the appropriateness of a PRO for use in performance measures, as indicated in Table 3. Given that PROs are not yet in widespread use in clinical practice, little is known about how best to aggregate these patient-level outcomes for the purpose of measuring performance of the health care entity. In spite of this, in order to accommodate the needs of patients with diverse linguistic, cultural, educational and functional skills, evidence is needed regarding the equivalence of multiple methods and modes of questionnaire administration. Additionally, scoring, analysis and reporting of PRO response data needs to be user-friendly and understandable to clinicians for use in real-time in clinical settings. Moreover, the timing of measurement must include pre-intervention in order to allow for measurement of responsiveness to change, to allow for risk adjustment, and to facilitate candidate screening for clinical intervention. In order to illustrate the application of these recommended characteristics when evaluating the appropriateness of a PRO for use as a performance measure, we provide the following example related to the evaluation of a PRO for use as a performance measure when evaluating the success of total hip arthroplasty.

Example of Applying Recommended Characteristics to Evaluate a Hip Osteoarthritis PRO for use in Performance Measurement

Total hip arthroplasty has emerged as an acceptable surgical treatment for individuals experiencing intractable pain and remarkable functional impairments for whom conservative treatment has yielded minimal improvement.[322,323 324,325] The most common indication for total hip arthroplasty is joint deterioration secondary to osteoarthritis.[326] Consequently, the aging of the population is likely to result in an increased demand for both primary, as well as revision total hip arthroplasty procedures.[327-329] Patient-reported outcomes have increasingly been included alongside more traditional indices of surgical outcome such as morbidity and mortality when evaluating the success of total hip arthroplasty as an intervention. With the increasing focus on patient-reported outcomes, such as functioning and quality of life, a widespread array of PROs have been developed and applied to the measurement of total hip arthroplasty outcomes.[326] Consequently, total hip arthroplasty provides a relevant context in which to review the use of recommended characteristics in the selection of PRO measures. Table 3 illustrates the application of important characteristics and best practices to evaluate and select a PRO for use as a performance measure for hip replacement outcomes. In this example, we illustrate the process of examining the characteristics of the Western Ontario and McMaster Universities

Osteoarthritis Index (WOMAC), a PRO measure developed to examine pain, stiffness, and physical function in individuals with osteoarthritis.[330]

Conclusion

Patient Reported Outcome (PRO) measures have reached a level of sophistication to enable their use in performance measures in the clinical setting. Attention to the many methodological considerations discussed in this paper will help produce meaningful, actionable results. Judicious use of a mixture of generic and disease-specific assessment, along with modern measurement methods such as item response theory, and the application of technology to enable standardized, equitable assessment across a range of patients, such as that applied in the development and validation of the PROMIS instruments, can effectively shorten assessment time without compromising accuracy, meeting the demands of clinical application of PROs for performance measurement.

Table 3[3]. Important characteristics and best practices to evaluate and select PROs for use in performance measures[260,265]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
1.	Conceptual and Measurement Model		
	A PRO measure should have documentation defining and describing the concept(s) included and the intended population(s) for use.	Target PRO concept should be a high priority for the health care system.	Factorial validity of the physical function and pain subscales has been inadequate.[331]
	There should be documentation of how the concept(s) are organized into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts.		
2.	Reliability		
	The degree to which an instrument is free from random error.		
2a.	*Internal consistency (multi-item scales)*	▪ reliability estimate ≥ 0.70 for group-level purposes ▪ reliability estimate ≥ 0.90 for individual-level purposes	Cronbach alphas for the three subscales range from 0.86 to 0.98.[332-334]
2b.	*Reproducibility (stability over time)* ▪ type of test-retest estimate depends on the response scale (dichotomous, nominal ordinal, interval, ratio)		Test-retest reliability has been adequate for the pain and physical function subscales, but less adequate for the stiffness subscale.[334]
3.	Validity		
	The degree to which the instrument reflects what it is supposed to measure.	There are a limited number of PRO instruments that have been validated for performance measurement.	
3a.	*Content Validity*		
	The extent to which a measure samples a representative range of the content.		
	A PRO measure should have evidence supporting its content validity, including evidence that patients and/or experts consider the content of the PRO measure relevant and comprehensive for the concept, population, and aim of the measurement		Development involved expert clinician input, and survey input from patients,[335] as well as a review of existing

[3] This table is adapted from recommendations contained within a report from the Scientific Advisory Committee of the Medical Outcomes Trust and a report submitted to the PCORI Methodology Committee. The recommendations from these sources have been adapted to enhance relevance to PRO selection for performance measurement.

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
	application.		measures.
	Documentation of qualitative and/or quantitative methods used to solicit and confirm attributes (i.e., concepts measured by the items) of the PRO relevant to the measurement application.		
	Documentation of the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, socio-economic status, literacy).		
	Documentation of sources from which items were derived, modified, and prioritized during the PRO measure development process.		
	Justification for the recall period for the measurement application.		
3b.	*Construct and Criterion-related Validity*		
	A PRO measure should have evidence supporting its construct validity, including: • documentation of empirical findings that support predefined hypotheses on the expected associations among measures similar or dissimilar to the measured PRO • documentation of empirical findings that support predefined hypotheses of the expected differences in scores between "known" groups		Patient ratings of satisfaction with arthroplasty were correlated with WOMAC scores in the expected direction.[22,336,337]
	A PRO measure should have evidence that shows the extent to which scores of the instrument are related to a criterion measure.		
3c.	*Responsiveness*		
	A PRO measure for use in longitudinal initiatives should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the target population.	If a PRO measure has cross-sectional data that provides sufficient evidence in regard to the reliability (internal consistency), content validity, and construct validity but has no data yet on responsiveness over time (i.e., ability of a PRO measure to detect changes in the construct being measured over time), would you accept use of the PRO measure to provide valid data over time in a longitudinal study if no other PRO measure was available?	Demonstrates adequate responsiveness and ability to detect change in response to clinical intervention.[338]
		Important to emphasize responsiveness because there is an expectation of consequences. Need to be able to demonstrate responsiveness if action is to be taken.	

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[330] for use in hip arthroplasty
4.	Interpretability of Scores		
	A PRO measure should have documentation to support interpretation of scores, including: • what low and high scores represent for the measured concept • representative mean(s) and standard deviation(s) in the reference population • guidance on the minimally important difference in scores between groups and/or over time that can be considered meaningful from the patient and/or clinical perspective	▪ If different PROs are used, it is important to establish a link or cross-walk between them. ▪ Because the criteria for assessing clinically important change in individuals does not directly translate to evaluating clinically important group differences, [306] a useful strategy is to calculate the proportion of patients who experience a clinically significant change[252,306]	Availability of population-based, age- and gender-normative values[339] Availability of minimal clinically important improvement values[340] Can be translated into a utility score for use in economic and accountability evaluations[341]
5.	Burden		
	The time, effort, and other demands on the respondent and the administrator.	In a busy clinic setting, PRO assessment should be as brief as possible, and reporting should be done in real-time.	Short form available[342] Average time to complete mobile phone WOMAC = 4.8 minutes[343]
6.	Alternatives modes and methods of administration	The use of multiple modes and methods can be useful for diverse populations. However, there should be evidence regarding their equivalence.	Validated mobile phone and touchscreen based platforms[344,345]
7.	Cultural and language adaptations	The mode, method and question wording must yield equivalent estimates of PRO measures.	Available in over 65 languages[346]
8.	Electronic health records (EHR)	Critical features: ▪ interoperability ▪ automated, real-time measurement and reporting ▪ sophisticated analytic capacities	Electronic data capture may allow for integration within EHR[343]

References

1. Maheu MM, Whitten P, Allen A. *E-Health, telehealth, and telemedicine: a guide to start-up and success.* San Francisco: Jossey-Bass; 2001.

2. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. *J. Am. Med. Inform. Assoc.* 2004;11(4):241-248.

3. National Quality Forum (NQF). *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care.* Washington, DC: National Quality Forum; 2009.

4. Wilson EV. *Patient-centered e-health.* Hershey, PA: Medical Information Science Reference; 2009.

5. US Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009; http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071975.pdf Accessed November 26, 2011.

6. Stephens RJ, Hopwood P, Girling DJ, Machin D. Randomized trials with quality of life endpoints: Are doctors' ratings of patients' physical symptoms interchangeable with patients' self- ratings? *Qual. Life Res.* 1997;6(3):225-236.

7. Justice AC, Rabeneck L, Hays RD, Wu AW, Bozzette SA, for the Outcomes Committee of the ACTG. Sensitivity, Specificity, Reliability, and Clinical Validity of Provider-Reported Symptoms: A Comparison With Self-Reported Symptoms. *J. Acquir. Immune Defic. Syndr.* 1999;21(2):126-133.

8. Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol.* 2006;7(11):903-909.

9. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J. Natl. Cancer Inst.* 2009;101(23):1624-1632.

10. Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med.* 2010;362(10):865-869.

11. Bech P. Quality of life measurements in chronic disorders. *Psychother. Psychosom.* 1993;59(1):1-10.

12. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. *J. Clin. Oncol.* 1993;11(3):570-579.

13. Guyatt GH. A taxonomy of health status instruments. *J. Rheumatol.* 1995;22(6):1188-1190.

14. Rothrock NE, Kaiser KA, Cella D. Developing a Valid Patient-Reported Outcome Measure. *Clin. Pharmacol. Ther.* 2011;90(5):737-742.

15. Osoba D. A taxonomy of the uses of health-related quality-of-life instruments in cancer care and the clinical meaningfulness of the results. *Med. Care.* 2002;40(6 Suppl):III31-38.

16. Benson T, Sizmur S, Whatling J, Arikan S, McDonald D, Ingram D. Evaluation of a new short generic measure of health status: howRu. *Inform. Prim. Care.* 2010;18(2):89-101.

17. Barry MJ, Fowler FJ, Jr., O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK. Measuring disease-specific health status in men with benign prostatic hyperplasia. Measurement Committee of The American Urological Association. *Med. Care.* 1995;33(4 Suppl):AS145-155.

18. Ware JE, Jr., Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I.Conceptual Framework and Item Selection. *Med. Care.* 1992;30(6):473-483.

19. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and final revision of a health status measure. *Med. Care.* 1981;19(8):787-805.

20. Caplan D, Hildebrandt N. *Disorders of Syntactic Comprehension.* Cambridge, Mass.: MIT Press; 1988.

21. Andresen EM, Rothenberg BM, Panzer R, Katz P, McDermott MP. Selecting a generic measure of health-related quality of life for use among older adults. A comparison of candidate instruments. *Eval. Health Prof.* 1998;21(2):244-264.

22. Bombardier C, Melfi CA, Paul J, et al. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med. Care.* 1995;33(4 Suppl):AS131-144.

23. Lundgren-Nilsson A, Tennant A, Grimby G, Sunnerhagen KS. Cross-diagnostic validity in a generic instrument: an example from the Functional Independence Measure in Scandinavia. *Health Qual. Life Outcomes.* 2006;4:55.

24. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. *Med. Care.* 2007;45(5 Suppl 1):S3-S11.

25. Cella D, Riley W, Stone A, et al. Initial item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) network: 2005-2008. *J. Clin. Epidemiol.* 2011;63(11):1179-1194.

26. Cella D, Lai JS, Nowinski C, et al. Neuro-QOL: Brief Measures of Health-related Quality of Life for Clinical Research in Neurology. *Neurology.* 2012;Epub ahead of print.

27. Tulsky DS, Kisala PA, Victorson D, et al. Developing a Contemporary Patient-Reported Outcomes Measure for Spinal Cord Injury. *Arch. Phys. Med. Rehabil.* 2011;92(10, Supplement):S44-S51.

28. Cella D. *Manual of the Functional Assessment of Chronic Illness Therapy (FACIT Scales). Version 4* Elmhurst, IL: FACIT.org; 1997.

29. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *Can. Med. Assoc. J.* 1986;134(8):889-895.

30. Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health Qual. Life Outcomes.* 2006;4:89.

31. Huang IC, Hwang CC, Wu MY, Lin W, Leite W, Wu AW. Diabetes-specific or generic measures for health-related quality of life? Evidence from psychometric validation of the D-39 and SF-36. *Value Health.* 2008;11(3):450-461.

32. Krahn M, Bremner KE, Tomlinson G, Ritvo P, Irvine J, Naglie G. Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Qual. Life Res.* 2007;16(3):509-522.

33. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch. Phys. Med. Rehabil.* 2000;81(12 Suppl 2):S21-29.

34. Bombardier C, Tugwell P. Methodological considerations in functional assessment. *J. Rheumatol.* 1987;14 Suppl 15:6-10.

35. Gabel CP, Michener LA, Burkett B, Neller A. The Upper Limb Functional Index: development and determination of reliability, validity, and responsiveness. *J. Hand Ther.* 2006;19(3):328-348; quiz 349.

36. Hobart J, Kalkers N, Barkhof F, Uitdehaag B, Polman C, Thompson A. Outcome measures for multiple sclerosis clinical trials: relative measurement precision of the Expanded Disability Status Scale and Multiple Sclerosis Functional Composite. *Mult. Scler.* 2004;10(1):41-46.

37. Kaasa T, Loomis J, Gillis K, Bruera E, Hanson J. The Edmonton Functional Assessment Tool: preliminary development and evaluation for use in palliative care. *J. Pain Symptom Manage.* 1997;13(1):10-19.

38. Mausbach BT, Moore R, Bowie C, Cardenas V, Patterson TL. A review of instruments for measuring functional recovery in those diagnosed with psychosis. *Schizophr. Bull.* 2009;35(2):307-318.

39. Olarsch S. *Validity and responsiveness of the late-life function and disability instrument in a facility-dwelling population*, Boston University; 2008.

40. Litwin MS, Hays R, Fink A, Ganz PA, Leake B, Brook RH. The UCLA Prostate Cancer Index: Development, reliability, and validity of health-related quality of life measure. *Med. Care.* 1998;26(7):1002-1012.

41. Rosen R, Brown C, Heiman J, et al. The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *J. Sex Marital Ther.* 2000;26(2):191-208.

42. Cleeland CS. Symptom burden: multiple symptoms and their impact as patient-reported outcomes. *J. Natl. Cancer Inst. Monogr.* 2007(37):16-21.

43. Smith E, Lai JS, Cella D. Building a measure of fatigue: the functional assessment of chronic illness therapy fatigue scale. *PM R.* 2010;2(5):359-363.

44. Yount SE, Choi SW, Victorson D, et al. Brief, Valid Measures of Dyspnea and Related Functional Limitations in Chronic Obstructive Pulmonary Disease (COPD). *Value Health.* 2011;14(2):307-315.

45. Cella D, Rosenbloom SK, Beaumont JL, et al. Development and Validation of 11 Symptom Indexes to Evaluate Response to Chemotherapy for Advanced Cancer. *J. Natl. Compr. Canc. Netw.* 2011;9(3):268-278.

46. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain.* 2010;150(1):173-182.

47. Deering DE, Sellman JD, Adamson SJ, Horn J, Frampton CMA. Development of a brief treatment instrument for routine clinical use with methadone maintenance treatment clients: the methadone treatment index. *Subst. Use Misuse.* 2008;43(11):1666-1680.

48. United States Department of Health and Human Services, Office of the National Coordinator for Health Information Technology. Meaningful use. 2011; http://healthit.hhs.gov/portal/server.pt?open=512&objID=2996&mode=2. Accessed July, 2011.

49. Estabrooks PA, Boyle M, Emmons KM, et al. Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial factors. *J. Am. Med. Inform. Assoc.* 2012;Epub ahead of print.

50. National Prevention Health Promotion and Public Health Council. National Prevention Strategy. 2011; http://www.healthcare.gov/prevention/nphpphc/strategy/report.html. Accessed July, 2011.

51. Bonevski B, Campbell E, Sanson-Fisher RW. The validity and reliability of an interactive computer tobacco and alcohol use survey in general practice. *Addict. Behav.* 2010;35(5):492-498.

52. Couwenbergh C, van der Gaag RJ, Koeter M, de Ruiter C, van den Brink W. Screening for substance abuse among adolescents validity of the CAGE-AID in youth mental health care. *Subst. Use Misuse.* 2009;44(6):823-834.

53. Smith PC, Schmidt SM, Allensworth-Davies D, Saitz R. Primary Care Validation of a Single-Question Alcohol Screening Test. *J. Gen. Intern. Med.* 2009;24(7):783-788.

54. Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System Survey Questionnaire.* Atlanta, Georgia: US Department of Health and Human Services, Center for Disease Control and Prevention; 2009.

55. Khazaal Y, Chatton A, Atwi K, Zullino D, Khan R, Billieux J. Arabic validation of the Compulsive Internet Use Scale (CIUS). *Subst. Abuse Treat. Prev. Policy.* 2011;6:32.

56. Storholm ED, Fisher DG, Napper LE, Reynolds GL, Halkitis PN. Proposing a tentative cut point for the Compulsive Sexual Behavior Inventory. *Arch. Sex. Behav.* 2011;40(6):1301-1308.

57. Paxton AE, Strycker LA, Toobert DJ, Ammerman AS, Glasgow RE. Starting the conversation performance of a brief dietary assessment and intervention tool for health professionals. *Am. J. Prev. Med.* 2011;40(1):67-71.

58. Sallis R. Developing healthcare systems to support exercise: exercise as the fifth vital sign. *Br. J. Sports Med.* 2011;45(6):473-474.

59. Wong SL, Leatherdale ST, Manske SR. Reliability and validity of a school-based physical activity questionnaire. *Med. Sci. Sports Exerc.* 2006;38(9):1593-1600.

60. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. *J. Clin. Hypertens.* 2008;10(5):348-354.

61. Buysse DJ, Moul DE, Germain A, Yu L, Stover A, Dodds NE. Development of a patient-reported outcome measure for sleep-wake function: scale development and initial validation. In preparation.

62. Agency for Healthcare Research and Quality. Notice Number: NOT-HS-05-005. Special Emphasis Notice: Research Priorities for the Agency for Healthcare Research and Quality. 2005 http://grants.nih.gov/grants/guide/notice-files/NOT-HS-05-005.html. Accessed June 25, 2012.

63. Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century.* Washington, D.C.: National Academy Press; 2001.

64. Hall JA, Dornan MC. Meta-analysis of satisfaction with medical care: Description of research domain and analysis of overall satisfaction levels. *Soc. Sci. Med.* 1988;27(6):637-644.

65. Lewis JR. Patient views on quality care in general practice: Literature review. *Soc. Sci. Med.* 1994;39(5):655-670.

66. Locker D, Dunt D. Theoretical and methodological issues in sociological studies of consumer satisfaction with medical care. *Soc. Sci. Med.* 1978;12:283-292.

67. Pascoe GC. Patient satisfaction in primary health care: A literature review and analysis. *Eval.Program.Plann.* 1983;6(3-4):185-210.

68. Williams B. Patient satisfaction: A valid concept? *Soc. Sci. Med.* 1994;38(4):509-516.

69. Shikiar R, Rentz AM. Satisfaction with medication: an overview of conceptual, methodologic, and regulatory issues. *Value Health.* 2004;7(2):204-215.

70. Linder-Pelz SU. Toward a theory of patient satisfaction. *Soc. Sci. Med.* 1982;16(5):577-582.

71. Oberst MT. Patients' perceptions of care. Measurement of quality and satisfaction. *Cancer.* 1984;53(10):2366-2375.

72. National Quality Forum (NQF). *Safe Practices for Better Healthcare–2010 Update.* Washington, D.C.: National Quality Forum; 2010.

73. Ware JE, Jr., Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. *Eval. Program Plann.* 1983;6(3-4):247-263.

74. Cella D, Bonomi A, Leslie WT, VonRoenn J, Tchekmedyian NS. Quality of life and nutritional well-being: measurement and relationship. *Oncology.* 1993;7(11, Suppl):S105-S111.

75. Rubin HR, Gandek B, Rogers WH, Kosinski M, McHorney CA, Ware JE, Jr. Patients' Ratings of Outpatient Visits in Different Practice Settings. Results from the Medical Outcomes Study. *JAMA.* 1993;270(7):835-840.

76. Graham J. Foundation for accountability(FACCT): a major new voice in the quality debate. In: Boyle J, ed. *1997 Medical Outcomes & Guidelines Sourcebook : a progress report and resource guide on medical outcomes research and practice guidelines : developments, data, and documentation*. New York: Faulkner & Gray; 1996.

77. Hays RD, Davies AR, Ware JE. Scoring the Medical Outcomes Study Patient Satisfaction Questionnaire: PSQ-III. Unpublished work 1987.

78. Moinpour CM. Assessment of quality of life in clinical trials. *Quality of life assesment in cancer clinical trials. Report of the Workshop on Quality of Life Research in Cancer Clinical Trials, July 16-17, 1990*. Bethesda, MD: U.S. Department of Health and Human Services; 1991.

79. Williams S. Consumer satisfaction surveys: health plan report cards to guide consumers in selecting benefit programs. In: Boyle J, ed. *1997 Medical Outcomes & Guidelines Sourcebook : a progress report and resource guide on medical outcomes research and practice guidelines : developments, data, and documentation*. New York: Faulkner & Gray; 1996.

80. Speight J. Assessing patient satisfaction: concepts, applications, and measurement. *Value Health.* 2005;8 Suppl 1:S6-8.

81. Epstein LH, Cluss PA. A behavioral medicine perspective on adherence to long-term medical regimens. *J. Consult. Clin. Psychol.* 1982;50(6):950-971.

82. Sherbourne CD, Hays RD, Ordway L, DiMatteo MR, Kravitz RL. Antecedents of adherence to medical recommendations: Results from the Medical Outcomes Study. *J. Behav. Med.* 1992;15(5):447-468.

83. Hays RD, Kravitz RL, Mazel RM, et al. The impact of patient adherence on health outcomes for patients with chronic disease in the Medical Outcomes Study. *J. Behav. Med.* 1994;17(4):347-360.

84. Hirsh AT, Atchison JW, Berger JJ, et al. Patient Satisfaction With Treatment for Chronic Pain: Predictors and Relationship to Compliance. *Clin. J. Pain.* 2005;21(4):302-310.

85. Ickovics JR, Meisler AW. Adherence in AIDS clinical trials: a framework for clinical research and clinical care. *J. Clin. Epidemiol.* 1997;50(4):385-391.

86. Kincey J, Bradshaw P, Ley P. Patients' satisfaction and reported acceptance of advice in general practice. *J. R. Coll. Gen. Pract.* 1975;25(157):558-566.

87. Augustin M, Reich C, Schaefer I, Zschocke I, Rustenbach SJ. Development and validation of a new instrument for the assessment of patient-defined benefit in the treatment of acne. *Journal der Deutschen Dermatologischen Gesellschaft.* 2008;6(2):113-120.

88. Blais MA. Development of an inpatient treatment alliance scale. *J. Nerv. Ment. Dis.* 2004;192(7):487-493.

89. Brod M, Christensen T, Bushnell D. Maximizing the value of validation findings to better understand treatment satisfaction issues for diabetes. *Qual. Life Res.* 2007;16(6):1053-1063.

90. Flood EM, Beusterien KM, Green H, et al. Psychometric evaluation of the Osteoporosis Patient Treatment Satisfaction Questionnaire (OPSAT-Q), a novel measure to assess satisfaction with bisphosphonate treatment in postmenopausal women. *Health Qual. Life Outcomes.* 2006;4:42.

91. Hudak PL, Hogg-Johnson S, Bombardier C, McKeever PD, Wright JG. Testing a new theory of patient satisfaction with treatment outcome. *Med. Care.* 2004;42(8):726-739.

92. Kumar RN, Kirking DM, Hass SL, et al. The association of consumer expectations, experiences and satisfaction with newly prescribed medications. *Qual. Life Res.* 2007;16(7):1127-1136.

93. Pouchot J, Trudeau E, Hellot SC, Meric G, Waeckel A, Goguel J. Development and psychometric validation of a new patient satisfaction instrument: the osteoARthritis Treatment Satisfaction (ARTS) questionnaire. *Qual. Life Res.* 2005;14(5):1387-1399.

94. Taback NA, Bradley C. Validation of the genital herpes treatment satisfaction questionnaire (GHerpTSQ) in status and change versions. *Qual. Life Res.* 2006;15(6):1043-1052.

95. Cella DF. Quality of life: The concept. *J. Palliat. Care.* 1992;8(3):8-13.

96. Lake T, Kvan C, Gold M. Literature Review: Using Quality Information for Health Care Decisions and Quality Improvement. *Mathematica Policy Research.* 2005;Reference No. 6110-230.

97. Schneider EC, Zaslavsky AM, Landon BE, Lied TR, Sheingold S, Cleary PD. National quality monitoring of Medicare health plans: the relationship between enrollees' reports and the quality of clinical care. *Med. Care.* 2001;39(12):1313-1325.

98. Browne K, Roseman D, Shaller D, Edgman-Levitan S. Analysis & commentary. Measuring patient experience as a strategy for improving primary care. *Health Aff. (Millwood).* 2010;29(5):921-925.

99. Cella DF, Lloyd SR. Data collection strategies for patient-reported information. *Qual. Manag. Health Care.* 1994;2(4):28-35.

100. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *J. Clin. Epidemiol.* 2002;55(11):1130-1143.

101. Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. *Arch. Dis. Child.* 2001;84(3):205-211.

102. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol. Assess.* 2001;5(4):1-157.

103. Weinfurt KP, Trucco SM, Willke RJ, Schulman KA. Measuring agreement between patient and proxy responses to multidimensional health-related quality-of-life measures in clinical trials. An application of psychometric profile analysis. *J. Clin. Epidemiol.* 2002;55(6):608-618.

104. Andresen EM, Vahle VJ, Lollar D. Proxy reliability: Health-related quality of life (HRQoL) measures for people with disability. *Qual. Life Res.* 2001;10(7):609-619.

105. Hart T, Whyte J, Polansky M, et al. Concordance of patient and family report of neurobehavioral symptoms at 1 year after traumatic brain injury. *Arch. Phys. Med. Rehabil.* 2003;84(2):204-213.

106. Matziou V, Perdikaris P, Feloni D, Moshovi M, Tsoumakas K, Merkouris A. Cancer in childhood: Children's and parents' aspects for quality of life. *Eur J Oncol Nurs.* 2008;12(3):209-216.

107. Matziou V, Tsoumakas K, Perdikaris P, Feloni D, Moschovi M, Merkouris A. Corrigendum to: "Cancer in childhood: Children's and parents' aspects for quality of life" [Eur J Oncol Nurs 12 (2008) 209-216] (DOI:10.1016/j.ejon.2007.10.005). *Eur J Oncol Nurs.* 2009;13(5).

108. Oczkowski C, O'Donnell M. Reliability of Proxy Respondents for Patients With Stroke: A Systematic Review. *J. Stroke Cerebrovasc. Dis.* 2010;19(5):410-416.

109. Brown-Jacobsen AM, Wallace DP, Whiteside SPH. Multimethod, multi-informant agreement, and positive predictive value in the identification of child anxiety disorders using the SCAS and ADIS-C. *Assessment.* 2011;18(3):382-392.

110. Agnihotri K, Awasthi S, Singh U, Chandra H, Thakur S. A study of concordance between adolescent self-report and parent-proxy report of health-related quality of life in school-going adolescents. *J. Psychosom. Res.* 2010;69(6):525-532.

111. Dorman PJ, Waddell F, Slattery J, Dennis M, Sandercock P. Are proxy assessments of health status after stroke with the EuroQol questionnaire feasible, accurate, and unbiased? *Stroke.* 1997;28(10):1883-1887.

112. Duncan PW, Lai SM, Tyler D, Perera S, Reker DM, Studenski S. Evaluation of proxy responses to the Stroke Impact Scale. *Stroke.* 2002;33(11):2593-2599.

113. Ostbye T, Tyas S, McDowell I, Koval J. Reported activities of daily living: agreement between elderly subjects with and without dementia and their caregivers. *Age Ageing.* 1997;26(2):99-106.

114. Sneeuw KC, Aaronson NK, de Haan RJ, Loeb JM. Assessing quality of life after stroke. The value and limitations of proxy ratings. *Stroke.* 1997;28(8):1541-1549.

115. Morrow AM, Hayen A, Quine S, Scheinberg A, Craig JC. A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. *Child Care Health Dev.* 2012;38(2):186-195.

116. White-Koning M, Arnaud C, Dickinson HO, et al. Determinants of child-parent agreement in quality-of-life reports: a European study of children with cerebral palsy. *Pediatrics.* 2007;120(4):804-814.

117. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: a review of the literature. *Qual. Life Res.* 2008;17(6):895-913.

118. Hilari K, Owen S, Farrelly SJ. Proxy and self-report agreement on the Stroke and Aphasia Quality of Life Scale-39. *J. Neurol. Neurosurg. Psychiatry.* 2007;78(10):1072-1075.

119. Lynn Snow A, Cook KF, Lin P-S, Morgan RO, Magaziner J. Proxies and Other External Raters: Methodological Considerations. *Health Serv. Res.* 2005;40(5p2):1676-1693.

120. Fowler FJ, Jr., Spilker B. Data Collection Methods. *Quality of Life and Pharmacoeconomics in Clinical Trials.* Vol 2nd. Philadelphia: Lippincott-Raven Publishers; 1996.

121. Naughton MJ, Shumaker SA, Anderson RT, Czajkowski SM, Spilker B. Psychological aspects of health-related quality of life measurement: tests and scales. *Quality of Life and Pharmacoeconomics in Clinical Trials.* Vol 2nd. Philadelphia: Lippincott-Raven; 1996.

122. Groves RM. *Survey methodology.* 2nd ed. Hoboken, NJ: J. Wiley; 2009.

123. Selltiz C, Wrightsman LS, Cook SW. *Research Methods in Social Relations.* New York: Holt, Rinehart and Winston; 1976.

124. Edwards AL. *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts; 1957.

125. Crowne DP, Marlowe D. *The approval motive: Studies in evaluative dependence.* New York: Wiley; 1964.

126. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J. Public Health.* 2005;27(3):281-291.

127. Anderson JP, Bush JW, Berry CC. Classifying function for health outcome and quality-of-life evaluation. Self- versus interviewer modes. *Med. Care.* 1986;24(5):454-469.

128. Cook DJ, Guyatt GH, Juniper E, et al. Interviewer versus self-administered questionnaires in developing a disease-specific, health-related quality of life instrument for asthma. *J. Clin. Epidemiol.* 1993;46(6):529-534.

129. McHorney CA, Kosinski M, Ware JE, Jr. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: Results from a national survey. *Med. Care.* 1994;32(6):551-567.

130. Chan KS, Orlando M, Ghosh-Dastidar B, Duan N, Sherbourne CD. The interview mode effect on the Center for Epidemiological Studies Depression (CES-D) scale: an item response theory analysis. *Med. Care.* 2004;42(3):281-289.

131. Weinberger M, Oddone EZ, Samsa GP, Landsman PB. Are health-related quality-of-life measures affected by the mode of administration? *J. Clin. Epidemiol.* 1996;49(2):135-140.

132. Chambers LW, Haight M, Norman G, MacDonald L. Sensitivity to change and the effect of mode of administration on health status measurement. *Med. Care.* 1987;25(6):470-480.

133. Wu AW, Jacobson DL, Berzon RA, et al. The effect of mode of administration on medical outcomes study health ratings and EuroQol scores in AIDS. *Qual. Life Res.* 1997;6(1):3-10.

134. Teresi JA. Overview of quantitative measurement methods: equivalence, invariance, and differential item functioning in health applications. *Med. Care.* 2006;44(11 Suppl 3):S39-S49.

135. Teresi JA. Different approaches to differential item functioning in health applications: advantages, disadvantages and some neglected topics. *Med. Care.* 2006;44(11 Suppl 3):S152-S170.

136. Borsboom D. When does measurement invariance matter? *Med. Care.* 2006;44(11 Suppl 3):S176-S181.

137. Hambleton RK. Good practices for identifying differential item functioning. *Med. Care.* 2006;44(11 Suppl 3):S182-S188.

138. McHorney CA, Fleishman JA. Assessing and Understanding Measurement Equivalence in Health Outcome Measures: Issues for Further Quantitative and Qualitative Inquiry. *Med. Care.* 2006;44(11 Suppl 3):S205-S210.

139. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health.* 2009;12(4):419-429.

140. Hahn E, Cella D, Dobrez D, et al. The Talking Touchscreen: a new approach to outcomes assessment in low literacy. *Psychooncology.* 2004;13(2):86-95.

141. Hahn EA, Cella D, Dobrez DG, et al. Quality of life assessment for low literacy Latinos: A new multimedia program for self-administration. *J. Oncol. Manag.* 2003;12(5):9-12.

142. Greist JH, Van Cura LJ, Erdman HP. Computer interview questionnaires for drug use/abuse. In: Lettieri DJ, National Institute on Drug Abuse, eds. *Predicting adolescent drug abuse : a review of issues, methods and correlates*. Rockville, Md.; Washington: U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute on Drug Abuse; 1975:164-174.

143. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value Health.* 2008;11(2):322-333.

144. Dalal AA, Nelson L, Gilligan T, McLeod L, Lewis S, DeMuro-Mercon C. Evaluating patient-reported outcome measurement comparability between paper and alternate versions, using the lung function questionnaire as an example. *Value Health.* 2011;14(5):712-720.

145. Abernethy AP, Herndon JE, Wheeler JL, et al. Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. *Health Serv. Res.* 2008;43(6):1975-1991.

146. Snyder CF, Jensen R, Courtin SO, Wu AW. PatientViewpoint: a website for patient-reported outcomes assessment. *Qual. Life Res.* 2009;18(7):793-800.

147. Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *J. Clin. Oncol.* 2004;22(4):714-724.

148. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: A randomized controlled trial. *JAMA.* 2002;288(23):3027-3034.

149. Velikova G, Brown JM, Smith AB, Selby PJ. Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. *Br. J. Cancer.* 2002;86(1):51-59.

150. Suh SY, Leblanc TW, Shelby RA, Samsa GP, Abernethy AP. Longitudinal patient-reported performance status assessment in the cancer clinic is feasible and prognostic. *J Oncol Pract.* 2011;7(6):374-381.

151. Chang CH, Cella D, Masters GA, et al. Real-time clinical application of quality-of-life assessment in advanced lung cancer. *Clin Lung Cancer.* 2002;4(2):104-109.

152. Wright EP, Selby PJ, Crawford M, et al. Feasibility and compliance of automated measurement of quality of life in oncology practice. *J. Clin. Oncol.* 2003;21(2):374-382.

153. Valderas J, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual. Life Res.* 2008;17(2):179-193.

154. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J. Eval. Clin. Pract.* 2006;12(5):559-568.

155. Mullen KH, Berry DL, Zierler BK. Computerized symptom and quality-of-life assessment for patients with cancer part II: acceptability and usability. *Oncol. Nurs. Forum.* 2004;31(5):E84-E89.

156. Jones JB, Snyder CF, Wu AW. Issues in the design of Internet-based systems for collecting patient-reported outcomes. *Qual. Life Res.* 2007;16(8):1407-1417.

157. Cleeland CS, Wang XS, Shi Q, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. *J. Clin. Oncol.* 2011;29(8):994-1000.

158. Basch E, Artz D, Iasonos A, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *J. Am. Med. Inform. Assoc.* 2007;14(3):264-268.

159. Hardwick ME, Pulido PA, Adelson WS. The use of handheld technology in nursing research and practice. *Orthop. Nurs.* 2007;26(4):251-255.

160. Sebille V, Hardouin J-B, Le Neel T, et al. Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients--a simulation study. *BMC Med. Res. Methodol.* 2010;10:24.

161. Bjorner JB, Chang C-H, Thissen D, Reeve BB. Developing tailored instruments: Item banking and computerized adaptive assessment. *Qual. Life Res.* 2007;16(Suppl1):95-108.

162. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: time to let the CAT out of the bag? *Health Serv. Res.* 2005;40(5 Pt 2):1694-1711.

163. Cook KF, Teal CR, Bjorner JB, et al. IRT health outcomes data analysis project: an overview and summary. *Qual. Life Res.* 2007;16 Suppl 1:121-132.

164. Coster W, Ludlow L, Mancini M. Using IRT variable maps to enrich understanding of rehabilitation data. *J. Outcome Meas.* 1999;3(2):123-133.

165. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual. Life Res.* 2007;16 Suppl 1:5-18.

166. Fayers PM. Applying item response theory and computer adaptive testing: the challenges for health outcomes assessment. *Qual. Life Res.* 2007;16 Suppl 1:187-194.

167. Fries JF, Bruce B, Cella D. The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. *Clin. Exp. Rheumatol.* 2005;23(S38):S33-S37.

168. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br. J. Clin. Psychol.* 2007;46(Pt 1):1-18.

169. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care.* 2007;45(5 Suppl 1):S22-S31.

170. Nunnally JC, Bernstein IH. *Psychometric Theory.* New York: McGraw-Hill, Inc.; 1994.

171. Fleiss JL. *The Design and Analysis of Clinical Experiments.* New York: John Wiley & Sons; 1986.

172. Lord FM, Novick MR. *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley 1968.

173. Allen MJ, Yen WM. *Introduction to measurement theory.* Monterey, CA: Brooks/Cole Publishing; 1979.

174. DeVellis RF. Classical test theory. *Med. Care.* 2006;44(11 Suppl 3):S50-S59.

175. DeVellis RF. *Scale development theory and applications.* Thousand Oaks, CA: Sage; 2003.

176. Fayers P, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes.* 2nd ed. Chichester: John Wiley & Sons; 2007.

177. Martinez-Martin P. Composite rating scales. *J. Neurol. Sci.* 2010;289(1-2):7-11.

178. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use.* New York: Oxford University Press; 2003.

179. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory.* Newbury Park, CA: SAGE Publications, Inc.; 1991.

180. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. *Med. Care.* 2000;38(9 Suppl):II60-II65.

181. van der Linden WJ, Hambleton RK. *Handbook of Modern Item Response Theory.* New York Springer-Verlag; 1997.

182. Wright BD, Masters GN. *Rating scale analysis: Rasch measurement.* Chicago: MESA Press; 1985.

183. Cook KF, Monahan PO, McHorney CA. Delicate balance between theory and practice: Health status assessment and Item Response Theory. *Med. Care.* 2003;41(5):571-574.

184. McHorney CA, Cohen AS. Equating health status measures with Item Response Theory: Illustrations with functional status items. *Med. Care.* 2000;38(9 Suppl):1143-1159.

185. Dorans N. Comparing or combining scores from multiple instruments: instrument linking (equating). Paper presented at: Advances in Health Outcomes Measurement2004; Bethesda, MD.

186. *Quality First: Better Health Care for All Americans. Final Report of the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry.* Washington, DC: US Government Printing Office;1998.

187. Hahn EA, Cella D. Health outcomes assessment in vulnerable populations: Measurement challenges and recommendations. *Arch. Phys. Med. Rehabil.* 2003;84(Suppl 2):S35-S42.

188. United States Agency for Healthcare Research Quality. *National healthcare disparities report 2010* Rockville, Md.: Agency for Healthcare Research and Quality 2011.

189. Hahn E, Cella D, Dobrez D, et al. The impact of literacy on health-related quality of life measurement and outcomes in cancer outpatients. *Qual. Life Res.* 2007;16(3):495-507.

190. Kirsch I, Jungeblut A, Jenkins L, Kolstad A. *Adult literacy in America: A first look at the results of the National Adult Literacy Survey.* Washington, DC: National Center for Education Statistics, U.S. Department of Education; 1993.

191. Kutner M, National Center for Education Statistics. *Literacy in everyday life: results from the 2003 National Assessment of Adult Literacy (NCES 2007-480). U.S. Department of Education.* Washington, DC: National Center for Education Statistics; 2007.

192. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. Health literacy: Report of the Council on Scientific Affairs. *JAMA.* 1999;281(6):552-557.

193. Parikh NS, Parker RM, Nurss JR, Baker DW, Williams MV. Shame and health literacy: The unspoken connection. *Patient Educ. Couns.* 1996;27(1):33-39.

194. Baker DW, Parker RM, Williams MV, Coates WC, Pitkin K. Use and effectiveness of interpreters in an emergency department. *JAMA.* 1996;274(10):783-788.

195. Lennon C, Burdick H. The Lexile Framework As An Approach For Reading Measurement And Success. 2004; http://www.lexile.com/m/uploads/whitepapers/Lexile-Reading-Measurement-and-Success-0504_MetaMetricsWhitepaper.pdf. Accessed January 25, 2011.

196. Klare GR. *The measurement of readability.* Ames: Iowa State University Press; 1963.

197. Liberman IY, Mann VA, Shankweiler D, Werfelman M. Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex.* 1982;18(3):367-375.

198. Shankweiler D, Crain S. Language mechanisms and reading disorder: a modular approach. *Cognition.* 1986;24(1-2):139-168.

199. Crain S, Shankweiler D. Syntactic Complexity and Reading Acquisition. In: Davidson A, Green GM, eds. *Linguistic complexity and text comprehension: readability issues reconsidered.* Vol Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1988:167-192.

200. Brach C, Keller D, Hernandez LM, et al. *Ten Attributes of Health Literate Health Care Organizations.* Washington, D.C.: National Academies Press; 2012.

201. U.S. Department of Health and Human Services OoMH, ,. *National Standards for Culturally and Linguistically Appropriate Services in Health Care.* Washington, DC: US Department of Health and Human Services; 2001.

202. Drasgow F, Kanfer R. Equivalence of psychological measurement in heterogeneous populations. *J. Appl. Psychol.* 1985;70:662-680.

203. Hui CH, Triandis HC. Measurement in cross-cultural psychology. *J Cross Cult Psychol.* 1985;16(2):131-152.

204. Angel R, Thoits P. The impact of culture on the cognitive structure of illness. *Cult. Med. Psychiatry.* 1987;11 23-52.

205. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual. Life Res.* 1993;2(6):451-459.

206. Hayes RP, Baker DW. Methodological problems in comparing English-speaking and Spanish-speaking patients' satisfaction with interpersonal aspects of care. *Med. Care.* 1998;36(2):230-236.

207. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: Translation and preliminary validity studies. *J. Clin. Epidemiol.* 1998;51(11):991-999.

208. Hunt SM. Cross-Cultural Comparability of Quality of Life Measures. *Drug Inf. J.* 1993;27(2):395-400.

209. Atkinson MJ, Lennox RD. Extending basic principles of measurement models to the design and validation of Patient Reported Outcomes. *Health Qual. Life Outcomes.* 2006;4(1):65.

210. da Mota Falcao D, Ciconelli RM, Ferraz MB. Translation and cultural adaptation of quality of life questionnaires: an evaluation of methodology. *J. Rheumatol.* 2003;30(2):379-385.

211. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. *Qual. Life Res.* 1998;7(4):323-335.

212. Herdman M, Fox-Rushby J, Badia X. Equivalence and the translation and adaptation of health-related quality of life questionnaires. *Qual. Life Res.* 1997;6(3):237-247.

213. Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value Health.* 2009;12(4):430-440.

214. Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient reported outcomes(PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health.* 2005;8(2):94-104.

215. Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health.* 2008;11(3):509-521.

216. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186-3191.

217. Dewolf L, Koller M, Velikova G, Johnson C, Scott N, Bottomley A. EORTC Quality of Life Group: Translation Procedure. 2009; 3rd:http://groups.eortc.be/qol/downloads/translation_manual_2009.pdf. Accessed November 26, 2011.

218. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval. Health Prof.* 2005;28(2):212-232.

219. Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology.* 2004;126(1 Suppl 1):S124-128.

220. Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. *Int. J. Technol. Assess. Health Care.* 1995;11(3):525-551.

221. Centers for Disease Control and Prevention. Prevalence and most common causes of disability among adults - United States, 2005. *MMWR Morb. Mortal. Wkly. Rep.* 2009;58(16):421-426.

222. National Council on Disability. The current state of health care for people with disabilities. 2009; http://purl.fdlp.gov/GPO/gpo3755.

223. Agency for Healthcare Research and Quality. Developing Quality of Care Measures for People with Disabilities: Summary of Expert Meeting. AHRQ Publication No. 10-0103. 2010; http://www.ahrq.gov/populations/devqmdis/.

224. North Carolina State University College of Design. Center for Universal Design. 2008; http://www.ncsu.edu/project/design-projects/udi/.

225. Story MF. Maximizing Usability: The Principles of Universal Design. *Assist. Technol.* 1998;10(1):4-12.

226. Section 508 of the Rehabilitation Act, as amended by the Workforce Investment Act of 1998 (P.L. 105-220). 1998; http://www.section508.gov/. Accessed February 20, 2010.

227. Harniss M, Amtmann D, Cook D, Johnson K. Considerations for Developing Interfaces for Collecting Patient-Reported Outcomes That Allow the Inclusion of Individuals With Disabilities. *Med. Care.* 2007;45(5 Suppl 1):S48-S54.

228. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc. Sci. Med.* 1999;48(11):1531-1548.

229. Nolte S, Elsworth GR, Sinclair AJ, Osborne RH. Tests of measurement invariance failed to support the application of the "then-test". *J. Clin. Epidemiol.* 2009;62(11):1173-1180.

230. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual. Life Res.* 2002;11(3):207-221.

231. Brossart DF, Clay DL, Willson VL. Methodological and statistical considerations for threats to internal validity in pediatric outcome data: response shift in self-report outcomes. *J. Pediatr. Psychol.* 2002;27(1):97-107.

232. Schwartz CE. Applications of response shift theory and methods to participation measurement: a brief history of a young field. *Arch. Phys. Med. Rehabil.* 2010;91(9 Suppl):S38-43.

233. Ring L, Hofer S, Heuston F, Harris D, O'Boyle CA. Response shift masks the treatment impact on patient reported outcomes (PROs): the example of individual quality of life in edentulous patients. *Health Qual. Life Outcomes.* 2005;3:55.

234. Ahmed S, Bourbeau J, Maltais F, Mansour A. The Oort structural equation modeling approach detected a response shift after a COPD self-management program not detected by the Schmitt technique. *J. Clin. Epidemiol.* 2009;62(11):1165-1172.

235. Mayo NE, Scott SC, Ahmed S. Case management poststroke did not induce response shift: the value of residuals. *J. Clin. Epidemiol.* 2009;62(11):1148-1156.

236. Ramachandran S, Lundy JJ, Coons SJ. Testing the measurement equivalence of paper and touch-screen versions of the EQ-5D visual analog scale (EQ VAS). *Qual. Life Res.* 2008;17(8):1117-1120.

237. Dillman DA, Smyth JD, Christian LM. *Internet, mail, and mixed-mode surveys: the tailored design method.* Hoboken, N.J.: Wiley & Sons; 2009.

238. Troxel AB, Fairclough DL, Curran D, Hahn EA. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat. Med.* 1998;17(5-7):653-666.

239. Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* Hoboken, NJ: John Wiley & Sons, Inc.; 2002.

240. Keeter S, Kennedy C, Dimock M, Best J, Craighill P. Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey. *Public Opin. Q.* 2006;70(5):759-779.

241. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. *JAMA.* 2012;307(17):1805-1806.

242. Johnson TP, Holbrook AL, Ik Cho Y, Bossarte RM. Nonresponse Error in Injury-Risk Surveys. *Am. J. Prev. Med.* 2006;31(5):427-436.

243. Cull WL, O'Connor KG, Sharp S, Tang S-fS. Response Rates and Response Bias for 50 Surveys of Pediatricians. *Health Serv. Res.* 2005;40(1):213-226.

244. Purdie DM, Dunne MP, Boyle FM, Cook MD, Najman JM. Health and demographic characteristics of respondents in an Australian national sexuality survey: comparison with population norms. *J. Epidemiol. Community Health.* 2002;56(10):748-753.

245. Voigt LF, Koepsell TD, Daling JR. Characteristics of telephone survey respondents according to willingness to participate. *Am. J. Epidemiol.* 2003;157(1):66-73.

246. Fairclough DL. Design and analysis of quality of life studies in clinical trials. Boca Raton: Chapman & Hall/CRC Press; 2002.

247. Little RA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc.* 1995;90(431):1112-1121.

248. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for MIXED Models.* Cary, NC: SAS Institute, Inc.; 1996.

249. Hahn EA, Glendenning GA, Sorensen MV, et al. Quality of life in patients with newly diagnosed chronic phase chronic myeloid leukemia on imatinib versus interferon alfa plus low-dose cytarabine: Results from the IRIS Study. *J. Clin. Oncol.* 2003;21(11):2138-2146.

250. Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Stat. Med.* 1998;17(5-7):781-796.

251. Basch EM, Reeve BB, Mitchell SA, et al. Electronic toxicity monitoring and patient-reported outcomes. *Cancer J.* 2011;17(4):231-234.

252. Guyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? *Qual. Life Res.* 2007;16(7):1097-1105.

253. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual. Life Res.* 2000;9(8):887-900.

254. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med. Care.* 1989;27(3 Suppl):S254-268.

255. Lipscomb J, Donaldson MS, Arora NK, et al. Cancer outcomes research. *J Natl Cancer Inst Monogr.* 2004(33):178-197.

256. 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.* 2011:S76–S85.

257. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J. Clin. Epidemiol.* 2010;63(7):737-745.

258. Revicki DA, Gnanasakthy A, Weinfurt K. Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: the PRO Evidence Dossier. *Qual. Life Res.* 2007;16(4):717-723.

259. Schunemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual. Life Outcomes.* 2006;4:62.

260. Butt Z, Reeve B. Enhancing the Patient's Voice: Standards in the Design and Selection of Patient-Reported Outcomes Measures (PROMs) for Use in Patient-Centered Outcomes Research. 2012; http://www.pcori.org/assets/Enhancing-the-Patients-Voice-Standards-in-the-Design-and-Selection-of-Patient-Reported-Outcomes-Measures-for-Use-in-Patient-Centered-Outcomes-Research.pdf. Accessed June 15, 2012.

261. U. S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. 2009; http://purl.access.gpo.gov/GPO/LPS113413.

262. US Food and Drug Administration. Draft Guidance for industry. Qualification process for drug development tools. 2010; http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf. Accessed November 26, 2011.

263. Erickson P, Willke R, Burke L. A concept taxonomy and an instrument hierarchy: tools for establishing and evaluating the conceptual framework of a patient-reported outcome (PRO) instrument as applied to product labeling claims. *Value Health.* 2009;12(8):1158-1167.

264. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health.* 2007;10 Suppl 2:S125-137.

265. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality of life instruments: attributes and review criteria. *Qual. Life Res.* 2002(11):193-205.

266. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Med. Res. Methodol.* 2010;10:82.

267. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Med. Res. Methodol.* 2011;11(1):152.

268. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med. Res. Methodol.* 2010;10:22.

269. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual. Life Res.* 2010;19(4):539-549.

270. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual. Life Res.* 2011;21(4):651-657.

271. Johnson C, Aaronson N, Blazeby JM, et al. EORTC Quality of Life Group: Guidelines for Developing Questionnaire Modules. 2011; 4th:http://groups.eortc.be/qol/Pdf%20presentations/Guidelines%20for%20Developing%20questionnaire-%20FINAL.pdf. Accessed November 26, 2011.

272. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health.* 2009;12(8):1075-1083.

273. Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health.* 2005;8(2):94-104.

274. Magasi S, Ryan G, Revicki D, et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual. Life Res.* 2011.

275. Valderas JM, Ferrer M, Mendivil J, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health.* 2008;11(4):700-708.

276. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J. Clin. Epidemiol.* 2008;61(2):102-109.

277. Kazis LE, Miller DR, Skinner KM, et al. Applications of methodologies of the Veterans Health Study in the VA healthcare system: conclusions and summary. *J. Ambulatory Care Manage.* 2006;29(2):182-188.

278. Kazis LE, Selim A, Rogers W, Ren XS, Lee A, Miller DR. Dissemination of methods and results from the veterans health study: final comments and implications for future monitoring strategies within and outside the veterans healthcare system. *J. Ambulatory Care Manage.* 2006;29(4):310-319.

279. Haffer SC, Bowen SE. Measuring and improving health outcomes in Medicare: the Medicare HOS program. *Health Care Financ. Rev.* 2004;25(4):1-3.

280. National Committee for Quality Assurance, Committee on Performance Measurement. *HEDIS 2006: health plan employer data & information set.* Washington, DC: National Committee for Quality Assurance; 2006.

281. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J. Clin. Epidemiol.* 2011;64(4):366-379.

282. Cella D, Nowinski C. Measuring quality of life in chronic illness: The Functional Assessment of Chronic Illness Therapy Measurement System. *Arch. Phys. Med. Rehabil.* 2002;83(Suppl. 2):S10-S17.

283. FDA Center for Drug Evaluation and Research Quality of Life Subcommittee Oncologic Drugs Advisory Committee. Meeting transcript. February 10, 2000. 2000; http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3591b1a.pdf.

284. Shearer D, Morshed S. Common generic measures of health related quality of life in injured patients. *Injury.* 2011;42(3):241-247.

285. Owolabi MO. Which Is More Valid for Stroke Patients: Generic or Stroke-Specific Quality of Life Measures? *Neuroepidemiology.* 2010;34(1):8-12.

286. Bergland A, Thorsen H, Kåresen R. Association between generic and disease-specific quality of life questionnaires and mobility and balance among women with osteoporosis and vertebral fractures. *Aging Clin. Exp. Res.* 2011;23(4):296-303.

287. Rothrock N, Hays R, Spritzer K, Yount SE, Riley W, Cella D. 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(11):1195-1204.

288. Chakravarty EF, Bjorner JB, Fries JF. Improving patient reported outcomes using item response theory and computerized adaptive testing. *J. Rheumatol.* 2007;34(6):1426-1431.

289. Donaldson G. Patient-reported outcomes and the mandate of measurement. *Qual. Life Res.* 2008;17(10):1303-1313.

290. Lai JS, Cella D, Choi SW, et al. How Item Banks and Their Application Can Influence Measurement Practice in Rehabilitation Medicine: A PROMIS Fatigue Item Bank Example. *Arch. Phys. Med. Rehabil.* 2011;92(10 Supplement):S20-S27.

291. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J. Clin. Epidemiol.* 2008;61(1):17-33.

292. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J. Chronic Dis.* 1985;38(1):27-36.

293. McClendon DT, Warren JS, Green KM, Burlingame GM, Eggett DL, McClendon RJ. Sensitivity to change of youth treatment outcome measures: a comparison of the CBCL, BASC-2, and Y-OQ. *J. Clin. Psychol.* 2011;67(1):111-125.

294. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual. Life Res.* 2003;12(4):349-362.

295. Beaton DE, van Eerd D, Smith P, et al. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. *J. Clin. Epidemiol.* 2011;64(5):487-496.

296. Andresen EM, Meyers AR. Health-related quality of life outcomes measures. *Arch. Phys. Med. Rehabil.* 2000;81(12 Suppl 2):S30-45.

297. Vermeersch DA, Lambert MJ, Burlingame GM. Outcome Questionnaire: item sensitivity to change. *J. Pers. Assess.* 2000;74(2):242-261.

298. Shikiar R, Willian MK, Okun MM, Thompson CS, Revicki DA. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. *Health Qual. Life Outcomes.* 2006;4:71.

299. Schroter S, Lamping DL. Responsiveness of the coronary revascularisation outcome questionnaire compared with the SF-36 and Seattle Angina Questionnaire. *Qual. Life Res.* 2006;15(6):1069-1078.

300. Kaplan RM, Tally S, Hays RD, et al. Five preference-based indexes in cataract and heart failure patients were not equally responsive to change. *J. Clin. Epidemiol.* 2011;64(5):497-506.

301. Bauer S, Lambert MJ, Nielsen SL. Clinical significance methods: a comparison of statistical techniques. *J. Pers. Assess.* 2004;82(1):60-70.

302. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *Jouornal of Clinical Epidemiology.* 2003;56(5):395-407.

303. Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual. Life Outcomes.* 2006;4(69):1-7

304. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control. Clin. Trials.* 1989;10(4):407-415.

305. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual. Life Res.* 1993;2(3):221-226.

306. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J. Pain.* 2008;9(2):105-121.

307. Revicki D, Hays R, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J. Clin. Epidemiol.* 2008;61(2):102-109.

308. Farivar SS, Liu H, Hays RD. Half standard deviation estimate of the minimally important difference in HRQOL scores? *Expert Rev. Pharmacoecon. Outcomes Res.* 2004;4(5):515-523.

309. Guyatt GH. Making sense of quality-of-life data. *Med. Care.* 2000;38(9 Suppl):II175-179.

310. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J. Clin. Epidemiol.* 2002;55(9):900-908.

311. Rejas J, Pardo A, Ruiz MA. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. *J. Clin. Epidemiol.* 2008;61(4):350-356.

312. Norman G, Sloan J, Wyrwich K. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med. Care.* 2003;41(5):582-592.

313. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann. Intern. Med.* 2006;144(10):742-752.

314. Goldzweig CL, Maglione M, Shekelle PG, Towfigh A. Costs and benefits of health information technology: New trends from the literature. *Health Aff. (Millwood).* 2009;28(2):w282-w293.

315. Harris Interactive, ARiA Marketing. *Healthcare Satisfaction Study.* Rochester, NY: Harris Interactive; 2000.

316. Davis F. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int J Man Mach Stud.* 1993;38(3):475-487.

317. Bitton A, Flier LA, Jha AK. Health information technology in the era of care delivery reform: to what end? *JAMA.* 2012;307(24):2593-2594.

318. Adler-Milstein J, Jha AK. Sharing clinical data electronically: A critical challenge for fixing the health care system. *JAMA.* 2012;307(16):1695-1696.

319. Masys D, Baker D, Butros A, Cowles KE. Giving patients access to their medical records via the internet: the PCASSO experience. *J. Am. Med. Inform. Assoc.* 2002;9(2):181-191.

320. Nelson EC, Hvitfeldt H, Reid RM, et al. *Using Patient-Reported Information to Improve Health Outcomes and Health Care Value: Case Studies from Dartmouth, Karolinska and Group Health.* Lebanon, NH: Dartmouth Institute for Health Policy and Clinical Practice;2012.

321. Davis K, Yount S, Del Ciello K, et al. An innovative symptom monitoring tool for people with advanced lung cancer: A pilot demonstration. *J. Support. Oncol.* 2007;5(8):381-387.

322. Harris WH, Sledge CB. Total hip and total knee replacement (1). *N Engl J Med.* 1990;323(11):725-731.

323. Harris WH, Sledge CB. Total hip and total knee replacement (2). *N Engl J Med.* 1990;323(12):801-807.

324. Liang MH, Cullen KE, Poss R. Primary total hip or knee replacement: evaluation of patients. *Ann. Intern. Med.* 1982;97(5):735-739.

325. Kroll MA, Otis JC, Sculco TP, et al. The relationship of stride characteristics to pain before and after total knee arthroplasty. *Clin. Orthop.* 1989(239):191-195.

326. Ethgen O, Bruyère O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J. Bone Joint Surg. Am.* 2004;86-A(5):86.

327. Birrell F, Johnell O, Silman A. Projecting the need for hip replacement over the next three decades: influence of changing demography and threshold for surgery. *Ann. Rheum. Dis.* 1999;58(9):569-572.

328. Rissanen P, Aro S, Sintonen H, Asikainen K, Slätis P, Paavolainen P. Costs and cost-effectiveness in hip and knee replacements. A prospective study. *Int. J. Technol. Assess. Health Care.* 1997;13(4):575-588.

329. Williams MH, Newton JN, Frankel SJ, Braddon F, Barclay E, Gray JAM. Prevalence of Total Hip Replacement: How Much Demand Has Been Met? *J. Epidemiol. Community Health.* 1994;48(2):188-191.

330. Bellamy N. *WOMAC Osteoarthritis Index: user guide IX.* Brisbane: Nicholas Bellamy; 2008.

331. Pua YH, Cowan SM, Wrigley TV, Bennell KL. Discriminant Validity of the Western Ontario and McMaster Universities Osteoarthritis Index Physical Functioning Subscale in Community Samples With Hip Osteoarthritis. *Arch. Phys. Med. Rehabil.* 2009;90(10):1772-1777.

332. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J. Rheumatol.* 1988;15(12):1833-1840.

333. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J. Bone Joint Surg. Br.* 2001;83(3):339-344.

334. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum.* 2001;45(5):453-461.

335. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin. Rheumatol.* 1986;5(2):231-241.

336. Bullens P, van Loon C, de Waal Malefijt M, Laan R, Veth R. Patient satisfaction after total knee arthroplasty. *J. Arthroplasty.* 2001;16(6):740-747.

337. Robertsson O, Dunbar MJ. Patient satisfaction compared with general health and disease-specific questionnaires in knee arthroplasty patients. *J. Arthroplasty.* 2001;16(4):476-482.

338. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res.* 1999;12(3):172-179.

339. Bellamy N, Wilson C, Hendrikz J. Population-based normative values for the Western Ontario and McMaster (WOMAC) Osteoarthritis Index: part I. *Semin. Arthritis Rheum.* 2011;41(2):139-148.

340. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann. Rheum. Dis.* 2005;64(1):29-33.

341. Marshall D, Pericak D, Grootendorst P, et al. Validation of a Prediction Model to Estimate Health Utilities Index Mark 3 Utility Scores from WOMAC Index Scores in Patients with Osteoarthritis of the Hip. *Value Health.* 2008;11(3):470-477.

342. Tubach F, Baron G, Falissard B, et al. Using patients' and rheumatologists' opinions to specify a short form of the WOMAC function subscale. *Ann. Rheum. Dis.* 2005;64(1):75-79.

343. Bellamy N, Patel B, Davis T, Dennison S. Electronic data capture using the Womac NRS 3.1 Index (m-Womac): A pilot study of repeated independent remote data capture in OA. *Inflammopharmacology.* 2010;18(3):107-111.

344. Bellamy N, Wilson C, Hendrikz J, et al. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. *J. Clin. Epidemiol.* 2011;64(2):182-190.

345. Theiler R, Bischoff-Ferrari HA, Good M, Bellamy N. Responsiveness of the electronic touch screen WOMAC 3.1 OA Index in a short term clinical trial with rofecoxib. *Osteoarthritis Cartilage.* 2004;12(11):912-916.

346. American College of Rheumatology. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). 2011; http://www.rheumatology.org/practice/clinical/clinicianresearchers/outcomes-instrumentation/WOMAC.asp. Accessed July 6, 2012.

Appendix D – Workshop Participant List

Patient-Reported Outcomes Workshop
July 30-31, 2012
Workshop Participants (On-Site)

Baranowski, Rebecca
American Board of Internal Medicine

Bershadsky, Julie
Human Services Research Institute

Berzon, Rick
National Institutes of Health

Blum, Steven
Forest Research Institute

Chang, Victor
VA New Jersey Health Care System/ UMDNJ

Cotter, Frances
Substance Abuse and Mental Health Administration

Dailey, Maureen
American Nurses Association

DeSoto, Mia
AHRQ

Deutsch, Anne
RTI International

Diamond, Louis
QHC

Doermann Byrd, Katherine
American College of Cardiology

Faerberg, Jennifer
AAMC

Gage, Barbara
The Brookings Institution

Garfinkel, Danielle
RTI International

Giovannetti, Erin
National Committee for Quality Assurance

Hinds, Pamela
Children's National Medical Center

Ireland, Andrea
NCQA

James, Tom
Humana

Kelleher, Cindy
RTI

Keller, San
American Institutes for Research

Kennedy, Cille
DHHS/ASPE

Lentz, Lisa
CMS

Mabry-Hernandez, Iris
AHRQ

Makadia, Preyanka
Agency for Healthcare Research and Quality

Mastanduno, Melanie
The Dartmouth Institute for Health Policy & Clinical Practice

McGuinn, Kristyne
American College of Cardiology

McNiff, Kristen
American Society of Clinical Oncology

Mitchell, Sandra
National Cancer Institute

Moon, JeanHee
Children's Hospital of Philadelphia

Moore, Jennifer
Agency for Healthcare Research & Quality

Patawaran, Wally
The John A. Hartford Foundation

Petrillo, Jennifer
Novartis

Riley, William
National Cancer Institute

Ross, Clarke
American Association on Health and Disability

Rubin, Koryn
American Association of Neurological Surgeons

Servies, Tammy
Uniformed Services University of the Health Sciences

Suarez, Monica
George Washington Internal Medicine Residency

Teschendorf, Bonnie
Adelphi Values

Tonkins, Phil
NIH/NIAMS

Wright, Lacey
Agency for Healthcare Research and Quality

Yang, DerShung
BrightOutcome Inc.

Patient-Reported Outcomes Workshop
July 30-31, 2012
Workshop Participants (Off-Site)

Aravamudhan, Krishna
Dental Quality Alliance

Asher, Anthony
American Association of Neurological Surgeons

Bilimoria, Karl
Northwestern

Brill, Joel
American Gastroenterological Association

Chang, Chih-Hung
Northwestern University Feinberg School of Medicine

Charlifue, Susan
Craig Hospital

Chauhan, Cynthia
Mayo Clinic Breast SPORE

Chen, Christine
Pacific Business Group on Health

Chisolm, Deena
The Research Institute at Nationwide Children's Hospital

Chiu, Jensen
American College of Cardiology

DeMark Neumann, Holly
Rehabilitation Institute of Chicago

Destefano, Alicia
Merck

Edwards, Todd
University of Washington

Gruber-Baldini, Ann
University of Maryland School of Medicine

Haas, Niina
BrightOutcome, Inc

Hagan, Eileen
American College of Cardiology

Han, Jane
The Society of Thoracic Surgeons

Harder, Joel
SCAI

Heinemann, Allen
Rehabilitation Institute of Chicago

Jalundhwala, Yash
UIC

Jensen, Sally
Northwestern University

Jewell, Kay
Tara Center LLC

Kidin, Lisa
UT MD Anderson

Ko, Clifford
American College of Surgeons

Lai, Jin-Shei
Northwestern University

Lepore, Michael
Planetree

Lewis, Barbara
Regeneron Pharmaceuticals, Inc

Maddux, Suzanne
ASCO

Matheson, James
EIP Consulting

McGonigal, Lisa
KCP

Miller, Lesley-Ann
GlaxoSmithKline

Myslinski, Rachel
American College of Rheumatology

Otte, Diane
Mayo Clinic Health System - Franciscan Healthcare

Shahriary, Melanie
American College of Cardiology

Spinks, Tracy
MD Anderson Cancer Center

Swain-Eng, Rebecca
American Academy of Neurology

Tavernier, Susan
University of Utah

Tobin, Judith
CMS

Whiteneck, Gale
Craig Hospital