

**Enhancing the Patient's Voice: Standards in the Design and Selection of
Patient-Reported Outcomes Measures (PROMs) for
Use in Patient-Centered Outcomes Research**
Methodology Committee Report

Zeeshan Butt, PhD
Northwestern University

Bryce Reeve, PhD
University of North Carolina – Chapel Hill

submitted to
Patient Centeredness Workgroup
PCORI Methodology Committee
March 30, 2012

DISCLAIMER

All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. PCORI has not peer-reviewed or edited this content, which was developed through a contract to support the Methodology Committee's development of a report to outline existing methodologies for conducting patient-centered outcomes research, propose appropriate methodological standards, and identify important methodological gaps that need to be addressed. The report is being made available free of charge for the information of the scientific community and general public as part of PCORI's ongoing research programs. Questions or comments about this report may be sent to PCORI at info@pcori.org or by mail to 1828 L St., NW, Washington, DC 20036.

1. Introduction

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 survival or tumor shrinkage are still very important, we also need to look at how patients' health-related quality of life (HRQOL) is affected by the disease and treatment. For such HRQOL endpoints, it is well accepted that the patient is the best source for reporting what they are experiencing in most cases. The challenge for PCOR 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.

Increasingly, longitudinal observational and experimental studies have included patient-reported outcome measures (PROMs), defined by the FDA as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."¹ Patients can report on a number of domains that are important for evaluating an intervention including symptom experiences (e.g., pain, fatigue, nausea), functional status (e.g., sexual, bowel, or urinary functioning), wellbeing (e.g., physical, mental, social), quality of life, and satisfaction with care or with a treatment.¹⁻⁴ In order to optimize decision making in PCOR, these PROMs must be measured in a standardized way using questionnaires that demonstrate specific measurement properties.⁵⁻¹⁰

The goal for this study was to identify the minimum standards for the design or selection of a PROM for use in PCOR. Central to this work was to develop an understanding of the critical attributes for which a PROM is judged to be appropriate or inappropriate for a PCOR study. 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 PROM and PCOR to seek consensus on the minimum standards.⁵

The identification of these standards will be a first step towards enabling PCOR to achieve its goals of enhancing healthcare delivery. Access to psychometrically sound and decision-relevant PROM will allow investigators to collect the empirical evidence on the differential benefits of a study intervention.^{7, 10-12} This 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.¹³

2. Method

2.1. Literature Review

We conducted a comprehensive review of the literature to identify existing PROM guidance documents. The review identified the current practices in selecting PROMs in PCOR, relevant questionnaire attributes (reliability, validity, response burden, and interpretability), and use of qualitative and quantitative methods to assess these questionnaire properties.

For the literature review strategy, we adapted a published MEDLINE search strategy to identify vital characteristics of patient-reported outcomes measures.¹⁴ This strategy was supplemented by developing a standardized list of search terms by consulting the MEDLINE thesaurus online, Medical Subject Headings (MeSH), and the American Psychological Association's (APA) online Thesaurus of Psychological terms. We then conducted parallel searches in several relevant electronic databases, including MEDLINE, PsycINFO, and Combined Index to Nursing and Allied Health Literature (CINAHL). Specific search strategies employed across the databases

are attached in Appendix A. (Other databases were initially considered, but not described in detail in this report because of irrelevant, low yield).

The titles and abstracts of identified articles and guidelines were reviewed by Dr. Butt. The full text of relevant articles were obtained and reviewed. The references cited in the included articles were reviewed to identify additional relevant articles. Dr. Butt abstracted the necessary information for the study; Drs. Cella, and Gershon of the Northwestern team independently coded several relevant articles to ensure coding consistency. Our focus was on consensus statements, guidelines, and evidence-based papers. We targeted articles or documents that described broadly generalizable principles, although some papers that were population- or instrument-specific provided this framework.

After synthesizing the existing published and unpublished guidelines, we reviewed standards for designing and selecting PROMs for PCOR. We reviewed standards for reliability (internal consistency and test-retest),^{4, 15-17} validity (content, construct, criterion-related),^{4, 18, 19} responsiveness (sensitivity to change),²⁰⁻²⁸ interpretability of scores and change in scores (i.e., clinically meaningful differences),^{21, 24-29} respondent and administrative burden,^{30, 31} comparability of different assessment modes (paper, computer, interviewer-administered),^{5, 32, 33} and cultural and language translations.³⁴⁻⁴² In addition, we reviewed appropriate means to obtain input from patients throughout the instrument development and evaluation process using both qualitative and quantitative methodologies to yield quality PROMs. In addition, we reviewed how some of these criteria (minimal standards) may vary depending on the population, for example, in pediatric populations, where respondent burden is a concern.^{43, 44}

The bulk of the evidence needed to specify minimum recommended standards came from our synthesis of the available guidelines identified in the literature review and expert survey. We looked for where there was agreement among the guidelines and where there may be variation in recommendations, giving emphasis to guidelines judged to have a high quality development process (e.g., external review process, patient's preferences and views were sought, systematic and thorough review).

2.2. Expert Input for Creating the Minimum Standards

We sought out the expertise of members of the International Society for Quality of Life Research (ISOQOL) to help develop the minimum standards for the design and selection of a PROM for use in PCOR. The ISOQOL is dedicated to advancing the scientific study of HRQOL and other patient-centered outcomes to identify effective interventions, enhance the quality of health care and promote the health of populations. Since 1993, ISOQOL has been an international collaborative network including pediatric and adult researchers, clinicians, patient advocates, government scientists, industry representatives, and policy makers. PROM methodologists are the backbone of ISOQOL. They concentrate on integrating qualitative and quantitative methods to improve the measurement and application of patient-reported data in research, healthcare delivery and population surveillance. Many of the PROMs used in research as well as the guidelines for developing and evaluating a PROM were created by ISOQOL members. Dr. Reeve (co-PI on this study) is the current President of ISOQOL.

This study engaged the members of ISOQOL in two ways to help write, refine, and seek consensus on the minimum standards. The first approach was the creation of an ISOQOL Scientific Advisory Task Force (SATF). The second approach was a standardized survey among ISOQOL members.

2.2.1. The ISOQOL Scientific Advisory Task Force

The 18-member ISOQOL SATF reflected expertise in the design, evaluation, and translation of psychometric-based and preference-based PROMs to capture experiences and perspectives from the general population and diverse patient populations. They also have expertise in using PROMs in a variety of healthcare delivery and research settings for decision support. The ISOQOL SATF was involved throughout the project period. Specifically, they supported our project team by 1) identifying any guidance document they were aware of in the literature or from their organization; 2) helping write the draft guidance on minimum standards; 3) helping design the survey to ISOQOL membership (described below); 4) reviewing the results from the ISOQOL survey and refining recommendations; and 5) identifying key issues to address for the project report.

The ISOQOL SATF included: Neil Aaronson, PhD (University of Amsterdam), Sara Ahmed, PhD (McGill University), Michael Brundage, MD (Queens University), Peter Fayers, PhD (University of Aberdeen), David Feeny, PhD (University of Alberta), Joanne Greenhalgh, PhD (University of Leeds), Ron Hays, PhD (University of California – Los Angeles), Pamela Hinds, PhD (National Children’s Hospital, Wash. DC), William Lenderking, PhD (United BioSource Corporation), Lori McLeod, PhD (Research Triangle Institute), Carol Moinpour, PhD (Fred Hutchinson Cancer Research Center), Dennis Revicki, PhD (United BioSource Corporation), Carolyn Schwartz, ScD (DeltaQuest Foundation/Tufts University Medical School), Claire Snyder, PhD (Johns Hopkins University), Caroline Terwee, PhD (VU University Medical Center), Galina Velikova, MD, PhD (University of Leeds), Albert Wu, MD, MPH (Johns Hopkins University), and Kathleen Wyrwich, PhD (United BioSource Corporation).

To facilitate our selection of minimum standards, we engaged the ISOQOL SATF throughout the study period via conference calls and e-mail correspondence to review findings and to seek their recommendations for minimal standards, especially for areas where there was disagreement among the existing guidelines.

2.2.2. The ISOQOL Member Survey

We sought input into the minimum standards, drafted with the help of the ISOQOL SATF, among the broader membership of ISOQOL through a structured internet-based survey. In the survey, we used multiple questionnaire types to seek input and consensus for minimum standards, paying particular attention to areas where there appeared to be disagreement in the literature or among ISOQOL SATF members. For example, we asked ISOQOL members to rank relative importance for measures of reliability including test-retest or internal consistency for multi-item PROMs. In addition, we sought consensus for recommendations for 4 key attributes of a PROM including: 1) Conceptual and Measurement Model, 2) Reliability, 3) Validity, and 4) Interpretability of Scores.

In the survey, it was deemed critical that respondents have a clear definition of a minimum standard. The second screen of the survey provided this guidance: “Please remember as you answer the questions in this survey that we are developing the **minimum** standards for the selection and design of a PROM for use in patient-centered outcomes research (PCOR). That is, we are saying a PROM that does not meet the **minimum standard** should not be considered appropriate for the research study.”

For each recommendation the participant could answer one of the following responses: required as a minimum standard, desirable but not required as a minimum standard, not required at all (not needed for a PROM), not sure, or no opinion. In analyzing the results we used the general

rule that if 50% or more agreed that the recommendation was required as a minimum standard then the recommendation was accepted. If less than 50% of respondents were in agreement than the recommendation was reviewed by the ISOQOL SATF and investigators to determine if the recommendation may have been unclear or if the recommendation may be better considered as a “best practice” (or “ideal standard) for PROMs than a “minimum standard”. Respondents were also encouraged to provide any comments in a free text box that was provided after each recommendation. This text was abstracted from the survey and helped inform the ISOQOL SATF and investigator decisions.

Prior to disseminating this survey it was reviewed and accepted by the Patient-Centered Outcomes Research Institute Methodology Committee. Additionally, the survey and survey methodology were submitted to the IRB at the University of North Carolina (UNC) for review and were determined exempt from IRB approval by the UNC Office of Human Research and Ethics. The online survey was designed and administered using the Qualtrics Software System under the UNC site license. Qualtrics Software enables the development and deployment of web-based surveys and was chosen because of its user-friendly interface and stringent privacy and security standards.

The Qualtrics survey link was sent out through the ISOQOL member email distribution list (n=506) on February 20, 2012. Survey instructions asked members to complete the survey within nine days (by February 29, 2012). Information about the purpose of the voluntary survey, goals of the project, and funding source were included. All responses were anonymous and no personal identifying information was collected. During the period the survey was available, two reminders were sent (mid-way through and last day).

3. Results

3.1 Guidance Identified Through Literature Review

Our team was aware of a number of existing guidance documents, including guidance documents from the FDA;^{1, 45-47} the 2002 Medical Outcomes Trust guidelines on attributes of a good HRQOL measure;² the extensive, international expert-driven recommendations from COSMIN (Consensus-based Standards for the selection of health Measurement INstruments);^{3, 4, 20, 48-51} the European Organization for Research and Treatment of Cancer (EORTC) guidelines for developing questionnaires;⁵² the Functional Assessment of Chronic Illness Therapy (FACIT) approach;⁵³ the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) task force recommendation documents,^{32, 37, 54, 55} and several others.^{18, 21, 56-58} We also had access to the recent standards documents just completed by the NIH’s Patient-Reported Outcomes Measurement Information System[®] (PROMIS[®]) network, which we considered useful for informing both the minimal standards as well as optimal standards for designing PROMs. In addition, the ISOQOL recently completed two guidance documents on use of PROMs in comparative effectiveness research and on integrating PROMs in healthcare delivery settings that were relevant for this landscape review.

The ISOQOL membership identified a total of 301 additional references relevant for our task. Our formal search of the MEDLINE database yielded 821 references, which were individually reviewed, resulting in 60 additional relevant articles. Review of the 172 potentially relevant PsycINFO results provided 22 additional relevant articles, and an additional 4 unique references were uncovered after review of 126 abstracts marked through CINAHL.

In keeping with the elements described in the PCORI RFP (PCORI-SOL-RMWG-001), **Table 1** describes guidance documents included in our report of the recommended minimum guidelines (in the order referenced in this report), **Table 2** describes exemplar guidance documents

discussed in the background document that informed the minimum guidelines but were less central to the recommendations, and **Table 3** reviews the quality of each of the key guidelines identified. As part of our literature review, we identified many more relevant references than indexed here; however, our focus was on existing guidance documents that had broad relevance. Multiple publications describing the same set of guidelines were not considered separately. Tables 1, 2 and 3 are at the end of this report.

3.2. Characteristics of Participants Responding to the ISOQOL Survey

The email invitation for the survey was sent to 506 members of ISOQOL through its distribution list. For the 9 days the survey was open (February 20-29, 2012), 98 ISOQOL members responded. Table 4 summarizes the characteristics of the survey respondents. Approximately 65% of the sample had a PhD and 17% had a MD. The sample included 71% academic researchers, 19% clinicians, 8% industry representatives, 19% industry consultants, and 7% federal government employees. There was diverse geographic distribution including 48% from North American (85% of those in North America were from the US), 33% from Europe, 9% from Asia, 6% from South America, 3% from Australia, and 1% from Africa.

The respondents were also well skilled in qualitative and quantitative methods and felt very comfortable providing guidance for recommendations for PROM standards. Approximately 81% of the sample reported they had moderate to extensive training in quantitative methods and 53% reported they had moderate to extensive training in qualitative methods. Overall, 89% reported they felt competent or very competent providing guidance. On average, the sample had 15 years of patient-reported outcome measurement and research experience in the field.

3.3. Findings and Recommendations for Minimum Standards for Attributes of a PROM for Use in PCOR

Table 5 provides an overview of the results from the ISOQOL survey on recommendations for minimal standards. A review of the findings from our literature review and survey is provided below.

3.3.1. Conceptual and Measurement Model

We recommend as a minimum standard that: “A PROM should have documentation defining and describing the concept(s) included and the intended population(s) for use. In addition, 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 included in the PROM.”^{2, 59-61} The ISOQOL membership was very supportive of this minimum standard with 91% of the sample endorsing the first statement as a requirement and 62% of the sample endorsing the second statement.

3.3.2. Reliability of a PROM

Reliability is a measure of the extent to which a PROM is free from random error.² In other words, it is the extent to which a PROM can distinguish one group of patients from another, despite measurement error.⁵⁹ For PROMs, the two most common types of reliability that are assessed include internal consistency and test-retest reliability. Internal consistency can be measured on one or more assessment (time) points and applies to multi-item scales (i.e., when two or more items are aggregated together to estimate a single score). Cronbach’s Coefficient Alpha⁶² is the most common measure of internal consistency and is approximate to the average across all split-half correlations among the items in the scale. Test-retest reliability is a measure of the reproducibility of the scale to provide consistent scores over time in a stable population. Common measures of test-retest reliability include intra-class correlation coefficients or weighted kappas depending on the scale.⁵⁹

We recommend as a minimum standard that: “The reliability of a PROM should ideally be at or above 0.70 for group level comparisons. Reliability for multi-item scales should include an assessment of internal consistency and test-retest reliability, and reliability for a single item measure should be assessed by test-retest reliability.”¹⁶

ISOQOL members were in agreement with this standard, except the recommendation that a PROM should be required as a minimum standard to have evidence of test-retest reliability. The concerns regarding test-retest reliability was that populations typically studied in PCOR are not stable and their HRQOL can often fluctuate. This pattern would reduce test-retest reliability making the PROM look unreliable when it may be precise and picking up valid change over time. In addition, memory effects will positively influence the test-retest reliability when the two survey points are scheduled close to each other.

The minimum level of reliability of .70 for group level comparisons is commonly accepted in the field.^{2, 59, 63} It represents approximately a half of standard error of measurement. However, there were concerns that establishing an absolute cut-off would be too strict (i.e. estimated reliability coefficient of 0.69 for a PROM deemed unreliable). Some of the ISOQOL members were more supportive of the statement of “no minimum level of reliability should be stated; however the reliability should be appropriately justified for the context of the proposed PROM measurement application.”

As recommendations shift in focus from “minimum” to “best practices”, item response theory (IRT) models are thought to provide very strong evidence of the precision of a PROM.⁶⁴⁻⁶⁹ Measures of reliability (e.g., internal consistency) give the wrong impression that as long as the reliability is above .70, the scale is reliable for measuring the PRO domain in any population. However, the reality is that a scale can be reliable for one population but not another. For example, one may have a very reliable measure of physical functioning to differentiate among athletes with questions like “Can you run 1 mile?” or “Can you run 5 miles?” However, these same items would be unreliable for differentiating among a very ill population that may have trouble getting out of the bed or walking from one room to another. IRT models have the ability to document how accurate (or reliable) a PROM is dependent on the levels of the latent trait (i.e. symptom) experienced in the population.

3.3.3. Validity of a PROM

The most common types of validity that were considered for minimum standards include content validity, construct validity, and responsiveness. Responsiveness is another aspect of construct validity;²³ however, it is discussed separately given its importance to PROM measurement in prospective studies. Criterion-related validity was not considered, as often in the PROM research field, there lacks a “gold standard” to which to compare a PROM measurement tool. The exception may be for measuring physical functioning when a PROM of physical functioning can be compared to observational study comparing what the patient reported to what he/she can perform in a lab.

Content validity is the extent to which the PROM represents the most relevant and important aspects of a concept in the context of a given measurement application.¹⁶ It is felt to be one of the most critical forms of validity to be assessed for a PROM.¹

We recommend as a minimum standard that “A PROM should have evidence supporting its content validity, including evidence that patients and/or experts consider the content of the PROM relevant and comprehensive for the concept, population, and aim of the measurement

application. This includes documentation of: 1) 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; 2) the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, gender, socio-economic status, literacy level) with an emphasis on similarities or differences with respect to the target population; and 3) justification for the recall period for the measurement application.”¹⁸

All these statements were endorsed by the ISOQOL members; however, there was disagreement for the recall period. Most (52%) felt a justification for the recall period was desirable but not required as a minimum standard for a PROM. We kept the recall statement in the recommendation as the reference period must be carefully considered for research participants to provide valid responses. However, no guidance can be recommended for a single reference period as it varies depending on the PRO domain being measured, the research context, and the population being studied.⁷⁰

One statement that was considered, but not supported by the ISOQOL members as a minimum standard was “documentation of sources from which items were derived, modified, and prioritized during the PROM development process.” We recommend this documentation be considered as a “best practice” but not a minimum standard for PROMs.

Construct validity is the extent to which scores on the PROM relate to other measures (e.g., patient-reported or clinical indicators) in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured.^{59 71} Construct validity also includes expected differences in scores among groups “known” to be different.

We recommend as a minimum standard that “A PROM 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.”^{16, 72, 73}

The ISOQOL members supported this recommendation. Another part of our original recommendation considered document of evidence for “known groups” validity requiring empirical findings that support predefined hypotheses of the expected differences in scores between “known” groups. We felt this was an important part of the evaluation of construct validity as it demonstrates the ability of a PROM to distinguish between one group and another where there is past empirical evidence there should be differences between the groups. However, the majority of ISOQOL members (57%) felt it was a desirable but not required standard. This may be considered as a standard for “best practice.”

Responsiveness (also known as sensitivity) is the extent to which a PROM can detect changes in the construct being measured over time.^{2, 23} Responsiveness is an aspect of construct validity and is also referred to as longitudinal validity.^{22, 23}

We recommend as a minimum standard that “A PROM for use in longitudinal research study should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the target population for the research application.”^{21, 74}

This statement was also endorsed by the ISOQOL membership (57%). However, when probed in the survey, 64% of respondents would agree to use a PROM that had no study to support the

responsiveness of the scale, but did have psychometric evidence in a cross-sectional study of the reliability and validity of the scale.

3.3.4. Interpretability of Scores

For a PROM to be well accepted for use in PCOR, it must provide scores that are easily interpretable to different stakeholders including patients, researchers, clinicians, and policy makers. They must be able to know what a high or low score represents. In addition, knowing what a meaningful difference or change in the score from one group to another (or one time to another) would be very informative to understanding the outcome being measured. Another way to enhance the interpretability of PROM scores would involve comparing one's score from a study to known scores in a population (e.g., the general US population or a specific disease population). This would enhance the ability to know how the study group compared to some norm group.

For minimum standards, we recommend "A PROM should have documentation to support interpretation of scores, including what low and high scores represent for the measured concept." This minimum standard was endorsed by 65% of the ISOQOL membership.

There are certainly better approaches to aid in the interpretation of the scores; however these recommendations would adhere to best practices as opposed to minimum standards. In agreement, 56% of ISOQOL members felt it would be good to have norm or reference scores. 72% agreed that estimation of minimally important differences (MIDs) would be highly desirable.^{21, 75, 76}

4. Summary and Conclusions

We have characterized the methods and results of our literature review highlighting existing guidelines that are informative to developing standards for designing or selecting PROMs for use in PCOR. We have also detailed the results of our survey among ISOQOL members, noting to the extent they are in alignment with the recommended standards we put forward. We have also made special note of standards for which there may be different recommendations depending on the population or context. We have distilled our specific recommendations into two general standards, which are attached here, along with abbreviated documentation for reference (Appendices B and C).

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

There are a number of considerations when applying these standards in PCOR. The populations participating in PCOR will likely be more heterogeneous than who is typically included in a phase III type trial. This population heterogeneity should be reflected in the samples that participate in the evaluation of the measurement properties for the PROM. 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 PROMs in PCOR. Data collected from PROMs 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 PROMs be attentive to make sure the questions and response options

are clear and easy to understand. Pre-testing of the PROM (e.g., cognitive testing) should include individuals with low literacy to evaluate the questions.⁷⁷

Response burden must be considered when selecting a PROM and using it in the PCOR study. A PROM 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 must 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 PROM. Older PROMs will have the benefit of having more evidence than younger PROMs, which should be reflected in the standards.

The extent to which a PROM adheres to the standards described in this report will result in good PROM measurement. Investigators wishing to select a PROM for use in PCOR should carefully consider how a PROM meets these minimal standards. In addition, care should be taken to confirm or establish the measurement properties of a PROM for the study target population in which the investigator wants to use the measure.

5. References

1. 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.
2. 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.
3. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol.* 2006;6:2.
4. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* Jul 2010;63(7):737-745.
5. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, Hess R, Miller DM, Reeve BB, Santana M. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res.* Nov 3 2011.
6. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology.* 2010;63(7):737-745.
7. Basch EM, Reeve BB, Mitchell SA, Clauser SB, Minasian L, Sit L, Chilukuri R, Baumgartner P, Rogak L, Blauel E, Abernethy AP, Bruner D. Electronic toxicity monitoring and patient-reported outcomes. *Cancer J.* Jul-Aug 2011;17(4):231-234.
8. 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. *Quality of Life Research.* 2007;16(4):717-723.
9. 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.
10. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care.* Mar 1989;27(3 Suppl):S254-268.
11. Guyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? *Qual Life Res.* Sep 2007;16(7):1097-1105.
12. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, Rothman M. 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.
13. Lipscomb J, Donaldson MS, Arora NK, Brown ML, Clauser SB, Potosky AL, Reeve BB, Rowland JH, Snyder CF, Taplin SH. Cancer outcomes research. *J Natl Cancer Inst Monogr.* 2004(33):178-197.
14. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* Oct 2009;18(8):1115-1123.
15. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* Jan 2011;64(1):96-106.

16. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. Nov-Dec 2007;10 Suppl 2:S94-S105.
17. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. May 2002;11(3):193-205.
18. Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, Snyder C, Boers M, Cella D. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual Life Res*. Aug 25 2011.
19. Streiner DL. A checklist for evaluating the usefulness of rating scales. *The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie*. 1993;38(2):140-148.
20. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Med Res Methodol*. Nov 18 2011;11(1):152.
21. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. Feb 2008;61(2):102-109.
22. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*. 2006;4:1-5.
23. Hays RD, D. H. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res*. 1992;1:73-75.
24. Wyrwich KW, Norquist JM, Lenderking W, Acaster S, Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. *Not published yet*. 2012.
25. Kemmler G, Zabernigg A, Gattringer K, Rumpold G, Giesinger J, Sperner-Unterweger B, Holzner B. A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient. *J Clin Epidemiol*. Feb 2010;63(2):171-179.
26. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. May 2003;56(5):395-407.
27. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics*. Nov 2000;18(5):419-423.
28. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. Feb 1991;59(1):12-19.
29. Sprangers MA, Moinpour CM, Moynihan TJ, Patrick DL, Revicki DA. Assessing meaningful change in quality of life over time: a users' guide for clinicians. *Mayo Clin Proc*. Jun 2002;77(6):561-571.
30. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS, Mayo FDAP-ROCMG. Patient-reported outcomes: instrument development and selection issues. *Value in Health*. 2007;10 Suppl 2:S86-93.
31. Scientific Working Group Quality of Life and Symptoms, European Hematology Association. Scientific working group quality of life and symptoms. <http://www.gemclinic.ru/konf01eng.php>.
32. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, Lenderking WR, Cella D, Basch E. 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*. Jun 2009;12(4):419-429.

33. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: A randomized study of Internet versus paper-and-pencil versions. *Journal of Medical Internet Research*. 2007;9(3):p39-p48.
34. Gawlicki MC, McKown S, Handa M. Preempting difficulties in linguistic validation: the use of face validation to create more sound translations. Paper presented at: International Society for Pharmacoeconomics Outcomes Research (ISPOR) Nov 6-9, 2011, 2011; Prague, Czech Republic.
35. Hagell P, Hedin P-J, Meads DM, Nyberg L, McKenna SP. Effects of method of translation of patient-reported health outcome questionnaires: A randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) instrument for Sweden. *Value in Health*. 2010;13(4):424-430.
36. 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*. May-Jun 2008;11(3):509-521.
37. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. 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*. Mar-Apr 2005;8(2):94-104.
38. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs*. Oct 2004;48(2):175-186.
39. Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology*. Jan 2004;126(1 Suppl 1):S124-128.
40. Schmidt S, Bullinger M. Current issues in cross-cultural quality of life instrument development. *Arch Phys Med Rehabil*. Apr 2003;84(4 Suppl 2):S29-34.
41. Hutchinson A, Bentzen N, Konig-Zahn C. *Cross cultural health outcome assessment - a user's guide*. Ruinen, The Netherlands: European Research Group on Health Outcomes; 1997.
42. 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*. Summer 1995;11(3):525-551.
43. Bevans KB, Riley AW, Moon J, Forrest CB. Conceptual and methodological advances in child-reported outcomes measurement. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2010;10(4):385-396.
44. Children's Oncology Group. Instrument Rating Tool.
45. 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.
46. 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*. Nov-Dec 2009;12(8):1158-1167.
47. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, O'Neill R, Kennedy DL. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health*. Nov-Dec 2007;10 Suppl 2:S125-137.
48. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, Knol DL, Bouter LM, de Vet HC. 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.
49. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. 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.
50. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. 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*. May 2010;19(4):539-549.
 51. 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*. Jul 6 2011.
 52. Johnson C, Aaronson N, Blazeby JM, Bottomley A, Fayers P, Koller M, Kulis D, Ramage J, Sprangers M, Velikova G, Young T. 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.
 53. Cella D. Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) measurement system. Evanston, IL: Northwestern University; 1997.
 54. 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*. Nov-Dec 2009;12(8):1075-1083.
 55. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, Hareendran A, Wiklund I, Chong LY, von Maltzahn R, Cohen L, Molsen E. 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*. Jun 2009;12(4):430-440.
 56. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, Alonso J. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health*. Jul-Aug 2008;11(4):700-708.
 57. Snyder CF, Aaronson NK, Choucair AK, T.E. E, Greenhalgh J, Halyard MY, Hess R, Miller DM, Reeve BB, Santana M. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. . *Qual Life Res*. 2011;epub ahead of print:1-10. .
 58. 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.
 59. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. Jan 2007;60(1):34-42.
 60. Health USDo, Human Services FDACfDE, Research, Human Services FDACfBE, Human Services FDACfD, Radiological H. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health & Quality of Life Outcomes*. 2006;4:79.
 61. Ware JE, Jr. Conceptualization and measurement of health-related quality of life: comments on an evolving field. *Archives of Physical Medicine & Rehabilitation*. 2003;84(4 Suppl 2):S43-51.
 62. Cronbach LJ. Coefficient alpha and the internal structure of tests. . *Psychometrika*. 1951;16:297-334.
 63. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.

64. Sawatzky R, Ratner PA, Kopec JA, Zumbo BD. Latent variable mixture models: a promising approach for the validation of patient reported outcomes. *Qual Life Res.* Aug 5 2011.
65. Sebille V, Hardouin J-B, Le Neel T, Kubis G, Boyer F, Guillemin F, Falissard B. 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 Medical Research Methodology.* 2010;10:24.
66. Revicki DA, Sloan J. Practical and philosophical issues surrounding a national item bank: if we build it will they come? *Quality of Life Research.* 2007;16 Suppl 1:167-174.
67. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clinical & Experimental Rheumatology.* 2005;23(5 Suppl 39):S53-57.
68. Reeve BB. Item response theory modeling in health outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res.* Apr 2003;3(2):131-145.
69. Cella D, Gershon R, Lai J-S, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research.* 2007;16 Suppl 1:133-141.
70. Norquist JM, Girman C, Fehnel S, Demuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Qual Life Res.* Sep 10 2011.
71. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use.* New York: Oxford University Press; 2003.
72. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* Jul 1955;52(4):281-302.
73. Mayo NE, Moriello C, Asano M, van der Spuy S, Finch L. The extent to which common health-related quality of life indices capture constructs beyond symptoms and function. *Qual Life Res.* Jun 2011;20(5):621-627.
74. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research.* 2010;19(4):539-549.
75. 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.
76. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Med Care.* Oct 2001;39(10):1039-1047.
77. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol.* Apr 2011;64(4):366-379.

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Guidance for Industry: Patient-reported outcome measures: Use in medical product development to support labeling claims.	United States Food and Drug Administration	2009	Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health	USA	No	N/A (proposed guidelines)	“This guidance describes how the Food and Drug Administration (FDA) reviews and evaluates existing, modified, or newly created <i>patient-reported outcome (PRO) instruments</i> used to support <i>claims</i> in approved medical product labeling.” It covers conceptual frameworks, content validity, reliability, validity, ability to detect change, modification of PRO, and use of PRO in special populations.
Medical Outcomes Trust	Scientific Advisory Committee	2002	--	USA primarily, but international	No	N/A (proposed guidelines)	Describes 8 key attributes of PROMs, including conceptual and measurement model, reliability, validity, responsiveness,

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
							interpretability, respondent and administrative burden, alternate forms, and cultural and language adaptations.
Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments	COSMIN group	2006	--	International	No	Guidelines established via systematic literature review and iterative Delphi process.	Consensus was reached on the inclusion and assessment of internal consistency, reliability, measurement error, content validity, construct validity, criterion validity, responsiveness, and interpretability.
Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations	Snyder, Aaronson, Choucair, Elliott, Greenhalgh, Halyard, Hess, Miller, Reeve, Santana; ISOQOL	2011	--	International	No	Literature review	The ISOQOL group developed a series of options and considerations to help guide the use of PROs in clinical practice, along with strengths and weaknesses of alternate approaches.

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed	Kottner, Audige, Brorson, Donnor, Gajewski, Hrobjartsson, Roberts, Shoukri, Streiner	2011	--	International	No	Literature review and expert consensus	Proposes a set of guidelines for reporting inter rater agreement, inter rater reliability in health care and medicine.
What is sufficient evidence for the reliability and validity of patient-reported outcome measures?	Frost, Reeve, Liepa, Stauffer, Hays; Mayo/FDA Patient-reported Outcomes Consensus Meeting Group	2007	--	USA	No	Literature review	Article provides specific guidance on necessary psychometric properties of a PROM, with special reference to the FDA guidance, using the literature as a guide for specific statistical thresholds.
Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes	Revicki, Hays, Cella, Sloan	2008	--	USA	No	Literature review and expert opinion	Makes concrete recommendations regarding estimation of minimally important differences (MID), which should be based on patient-based and clinical anchors and

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
							convergence across multiple approaches and methods.
Defining clinically meaningful change in health-related quality of life	Crosby, Kolotkin, Williams	2003	--	USA	No	Literature review	Reviews current approaches to defining clinically meaningful change in health-related quality of life and provides guidelines for their use.
Assessing meaningful change in quality of life over time: A users' guide for clinicians	Sprangers, Moinpour, Moynihan, Patrick, Revicki; The Clinical Significance Consensus Meeting Group	2002	--	International	No	Literature review and expert opinion	Proposes a set of guidelines/questions to help guide clinicians as to how to use PROM data in the treatment decision process.

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures	Coons, Gwaltney, Hays, Lundy, Sloan, Revicki, Lenderking, Cella, Basch; ISPOR ePRO Good Research Practices Task Force	2009	--	International	No	Expert opinion and literature review	Provides a general framework for decisions regarding evidence needed to support migration of paper PROMs to electronic delivery.
Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials	Acquadro, Conway, Hareendran, Aaronson; European Regulatory Issues and Quality of Life Assessment (ERIQA) Group	2008	--	European Union	No	Formal literature review	Call for more empirical research on translation methodology; reviews several existing guidelines; advocates multistep process for translations.

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures	Wild, Grove, Martin, Eremenco, McElroy, Verjee-Lorenz, Erikson; ISPOR Task Force for Translation and Cultural Adaptation	2005	--	International	No	Literature review and expert opinion/consensus	The ISPOR Task Force produced a critique of the strengths and weaknesses of various methods for translation and cultural adaptation of PROMS.
Guidelines for developing questionnaire modules	Johnson, Aaronson, Blazeby, Bottomley, Fayers, Koller, Kulis, Ramage, Sprangers, Velikova, Young; EORTC Quality of Life Group	2011	--	European Union	No	Expert opinion	Provides detailed description of PROM module development per the EORTC methodology related to generation of issues, construction of item list, pre- and field-testing.

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Manual for the Functional Assessment of Chronic Illness Therapy (FACIT)	Cella	1997	--	USA	No	Description of method	Provides summary of FACIT scale development and translation methodologies; presents basic psychometric info for existing measures.
Use of existing patient-reported outcome (PRO) instruments and their modification	Rothman, Burke, Erickson, Leidy, Patrick, Petrie; ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force	2009	--	USA	No	Expert opinion	Discusses key issues regarding the assessment and documentation of content validity for an existing instrument; discusses potential threats to content validity and methods to ameliorate.

Table 1. Description of Key Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data	Wild, Eremenco, Mear, Martin, Houchin, Gawlicki, Hareendran, Wiklund, Chong, von Maltzahn, Cohen, Molsen; ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force	2009	--	International	No	Expert opinion and literature review	Provides decision tools to decide on translation required for PROM; approach to use when same language is spoken in more than one country; and methods to gather evidence to support pooling of data across different language versions.

Table 2. Description of Other Informative Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: The PRO Evidence Dossier	Revicki, Gnanasakthy, Weinfurt	2007	--	USA	No	Report	Describes the purpose and content of a PROM Evidence Dossier, as well as its potential role with respect to regulatory review.
Interpreting the results of patient reported outcome measures in clinical trials: The clinician's perspective	Schunemann, Akl, Guyatt	2006	--	USA, Canada	No	Report based on examples	The authors provided several examples to describe how to attach meaning to PROM score thresholds and/or score differences.
Recommendations on health-related quality of life research to support labeling and promotional claims in the United States	Revicki, Osoba, Fairclough, Barofsky, Berzon, Leidy, Rothman	2000	--	USA, Canada	No	Review	Outlines the importance of an evidentiary base for making claims with respect to medical labeling or promotional claims.

Table 2. Description of Other Informative Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting	Magasi, Ryan, Revicki, Lenderking, Hays, Brod, Snyder, Boers, Cella	2011	--	USA, Netherlands	N/A	Expert presentation and discussion	The paper describes findings from a PROMIS meeting focused on content validity. Several recommendations were outlined as a result, including the need for consensus driven guidelines (none were proposed).
Methods for interpreting change over time in patient-reported outcome measures	Wyrwich, Norquist, Lenderking, Acaster; International Society of Quality of Life Research	2012	Industry Advisory Committee	USA, UK	N/A	Literature review	This article reviews the evolution of the methods and the terminology used to describe and aid in the communication of meaningful PROM change score thresholds.

Table 2. Description of Other Informative Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient	Kemmler, Zabernigg, Gattringer, Rumpold, Giesinger, Sperner-Unterweger, Holzner	2010	--	Austria	N/A	Longitudinal data from a chemotherapy trial	Data from this trial was used to evaluate change for individual participants (vs groups). Stressed the importance of evaluation on the basis of statistical and clinical significance.
The concept of clinically meaningful change in health-related quality of life research: How meaningful is it?	Hays, Woolley	2000	--	USA	No	Expert opinion	Argues against a single threshold to define the minimally clinical important difference.
Patient-reported outcomes: Instrument development and selection issues	Turner, Quittner, Parasuraman, Kallich, Cleeland, Mayo FDAP-ROCMG	2007	--	USA	No	Literature review	Provides a broad summary of concepts and issues to consider in the development and selection of a PROM.

Table 2. Description of Other Informative Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Current issues in cross-cultural quality of life instrument development	Schmidt, Bullinger	2003	--	Germany	No	Literature review	Provides an overview of cross-cultural adaptation of PROM and provides broad development guidelines, as well as a call for additional focus on international research.
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	Erickson, Willke, Burke	2009	--	USA	No	Expert opinion	Proposes a PROM concept taxonomy and instrument hierarchy that may be useful for demonstration of PROM claim for drug development, although they have not been tested for such purpose.

Table 2. Description of Other Informative Guidance Statements

Guideline	Organization or Authors	Year	Program	Country or Region	Guideline subjected to independent external review?	Research Design	Description
Translation procedure	Dewolf, Koller, Velikova, Johnson, Scott, Bottomley; EORTC Quality of Life Group	2009	--	European Union	No	Expert opinion	Provides guidance on the methodology for translating EORTC Quality of Life Questionnaires (QLQ).
Choice of recall period for patient-reported outcome (PRO) measures: Criteria for consideration	Norquist, Girman, Fehnel, Demuro-Mercon, Santanello	2011	--	USA	No	Literature review	Choice of recall period for a PROM depends on nature of the disease, stability of symptoms, and trajectory of symptoms over time.

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

Guideline	The purpose of the work is to define methodological standards for PCOR	The applications of the standards to PCOR is clear	The standards were developed by a professional group	Patient's views and preferences were sought	Stakeholders were involved in the development of the Standards	A systematic process was used to generate recommendations	Details of the systematic process used to generate recommendations are provided	There is an explicit link between the rationale for and the recommended standards (evidence)	The standards underwent independent external reviews (See note)	The recommendation are specific and unambiguous	Key recommendations are clear	The standards are editorially independent from the funding body	Conflicts of interest have been recorded
Guidance for Industry: Patient-reported outcome measures: Use in medical product development to support labeling claims.	Yes	Yes	Yes	Unclear	Limited to experts	Unclear	No	No	No	No	Yes	No	No
Medical Outcomes Trust	Yes	Yes	Yes	No	Limited to experts	Unclear	No	No	No	Yes	Yes	Yes	Yes

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

Guideline	The purpose of the work is to define methodological standards for PCOR	The applications of the standards to PCOR is clear	The standards were developed by a professional group	Patient's views and preferences were sought	Stakeholders were involved in the development of the Standards	A systematic process was used to generate recommendations	Details of the systematic process used to generate recommendations are provided	There is an explicit link between the rationale for and the recommended standards (evidence)	The standards underwent independent external reviews (See note)	The recommendation are specific and unambiguous	Key recommendations are clear	The standards are editorially independent from the funding body	Conflicts of interest have been recorded
Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments	Yes	Yes	No	No	Limited to experts	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations	Yes	Yes	Yes	No	Limited to experts	Unclear	No	Yes	No	No	Yes	Yes	Yes

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

Guideline	The purpose of the work is to define methodological standards for PCOR	The applications of the standards to PCOR is clear	The standards were developed by a professional group	Patient's views and preferences were sought	Stakeholders were involved in the development of the Standards	A systematic process was used to generate recommendations	Details of the systematic process used to generate recommendations are provided	There is an explicit link between the rationale for and the recommended standards (evidence)	The standards underwent independent external reviews (See note)	The recommendation are specific and unambiguous	Key recommendations are clear	The standards are editorially independent from the funding body	Conflicts of interest have been recorded
Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed	Yes	Yes	No	No	Limited to experts	Yes	Yes	Yes	No	Yes	Yes	Yes	Non e noted.
What is sufficient evidence for the reliability and validity of patient-reported outcome measures?	Yes	Yes	No	No	Limited to experts	Unclear	No	Yes	No	Yes	Yes	Yes	Yes

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

Guideline	Conflicts of interest have been recorded	Yes
	The standards are editorially independent from the funding body	Yes
The purpose of the work is to define methodological standards for PCOR	Key recommendations are clear	Yes
	The recommendation are specific and unambiguous	Yes
The standards underwent independent external reviews (See note)	Yes	Yes
	No	No
There is an explicit link between the rationale for and the recommended standards (evidence)	Yes	Yes
	No	No
Details of the systematic process used to generate recommendations are provided	Yes	Yes
	No	No
A systematic process was used to generate recommendations	Yes	Unclear
	No	No
Stakeholders were involved in the development of the Standards	Limited to experts	Limited to experts
	No	No
Patient's views and preferences were sought	No	No
	No	No
The standards were developed by a professional group	No	No
	No	No
The applications of the standards to PCOR is clear	Yes	Yes
	No	No
Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes	Yes	Yes
	No	No
Defining clinically meaningful change in health-related quality of life	Yes	Yes
	No	No

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

	Conflicts of interest have been recorded	Yes
	The standards are editorially independent from the funding body	Yes
	Key recommendations are clear	Yes
	The recommendation are specific and unambiguous	Yes
	The standards underwent independent external reviews (See note)	No
	There is an explicit link between the rationale for and the recommended standards (evidence)	Yes
	Details of the systematic process used to generate recommendations are provided	No
	A systematic process was used to generate recommendations	Unclear
	Stakeholders were involved in the development of the Standards	Limited to experts
	Patient's views and preferences were sought	No
	The standards were developed by a professional group	No
	The applications of the standards to PCOR is clear	Yes
	The purpose of the work is to define methodological standards for PCOR	Yes
Guideline	Assessing meaningful change in quality of life over time: A users' guide for clinicians	

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

	Conflicts of interest have been recorded	Yes
	The standards are editorially independent from the funding body	Yes
	Key recommendations are clear	Yes
	The recommendation are specific and unambiguous	Yes
	The standards underwent independent external reviews (See note)	No
	There is an explicit link between the rationale for and the recommended standards (evidence)	Yes
	Details of the systematic process used to generate recommendations are provided	No
	A systematic process was used to generate recommendations	Unclear
	Stakeholders were involved in the development of the Standards	Limited to experts
	Patient's views and preferences were sought	No
	The standards were developed by a professional group	Yes
	The applications of the standards to PCOR is clear	Yes
	The purpose of the work is to define methodological standards for PCOR	Yes
Guideline	Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures	

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

Guideline	The purpose of the work is to define methodological standards for PCOR	The applications of the standards to PCOR is clear	The standards were developed by a professional group	Patient's views and preferences were sought	Stakeholders were involved in the development of the Standards	A systematic process was used to generate recommendations	Details of the systematic process used to generate recommendations are provided	There is an explicit link between the rationale for and the recommended standards (evidence)	The standards underwent independent external reviews (See note)	The recommendation are specific and unambiguous	Key recommendations are clear	The standards are editorially independent from the funding body	Conflicts of interest have been recorded
Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials	Yes	Yes	Yes	No	Limited to experts	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures	Yes	Yes	Yes	No	Limited to experts	Unclear	No	No	No	Yes	Yes	Yes	Yes

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

Guideline	The purpose of the work is to define methodological standards for PCOR	The applications of the standards to PCOR is clear	The standards were developed by a professional group	Patient's views and preferences were sought	Stakeholders were involved in the development of the Standards	A systematic process was used to generate recommendations	Details of the systematic process used to generate recommendations are provided	There is an explicit link between the rationale for and the recommended standards (evidence)	The standards underwent independent external reviews (See note)	The recommendation are specific and unambiguous	Key recommendations are clear	The standards are editorially independent from the funding body	Conflicts of interest have been recorded
Guidelines for developing questionnaire modules	Yes	Yes	Yes	No	Limited to experts	Unclear	No	No	No	Yes	Yes	Unclear	No
Manual for the Functional Assessment of Chronic Illness Therapy (FACIT)	Yes	Yes	No	No	Process description is expert-guided	Unclear	No	No	No	No	Yes	Yes	Yes
Use of existing patient-reported outcome (PRO) instruments and their modification	Yes	Yes	Yes	No	Limited to experts	Unclear	No	No	No	No	Yes	Yes	Yes

Table 3. Selected Characteristics of Documents Included in Recommended Guidelines

	Conflicts of interest have been recorded	Yes
	The standards are editorially independent from the funding body	Yes
	Key recommendations are clear	Yes
	The recommendation are specific and unambiguous	Yes
	The standards underwent independent external reviews (See note)	No
	There is an explicit link between the rationale for and the recommended standards (evidence)	Yes
	Details of the systematic process used to generate recommendations are provided	No
	A systematic process was used to generate recommendations	Unclear
	Stakeholders were involved in the development of the Standards	Limited to experts
	Patient's views and preferences were sought	No
	The standards were developed by a professional group	Yes
	The applications of the standards to PCOR is clear	Yes
	The purpose of the work is to define methodological standards for PCOR	Yes
Guideline	Multinational trials- recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data	

Table 4: Sample Characteristics

Sample Characteristic	% (n = 98)
Degrees*	
MD	17%
PhD/Other Doctoral Degree (e.g., ScD)	65%
RN/NP	4%
Physical/Occupational Therapist	8%
MA, MSc, MPH, or other Master's	45%
Role*	
Academic Researcher	71%
Clinician	19%
Industry Representative	8%
Industry Consultant/CRO Employee	19%
Federal Government Employee	7%
Patient Advocate	2%
Other	6%
Geographic Location	
North America	48%
United States	(85%)
Europe	33%
South America	6%
Asia	9%
Africa	1%
Australia	3%
Psychometric Training (Quantitative Methods)	
Extensive training	37%
Moderate amount of training	44%
A little training	16%
Not any training	3%
Qualitative Training	
Extensive training	18%
Moderate amount of training	35%
A little training	40%
Not any training	7%
Competency	
Very competent	50%
Competent	39%
Somewhat competent	8%
A little competent	3%
Average number of years in health-related quality (HRQOL) or patient-reported outcomes (PROs) field.	
Mean years in HRQOL or PRO field	15 years; (range 1-40 years)

Note: *More than one response was allowed for this characteristic.

Table 5: Survey Results

	Draft Recommendation for minimal standards	Survey Results (n=98)
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.	<ul style="list-style-type: none"> • Required as a minimum standard – 91% • Desirable but not required as a minimum standard – 9% • Not required – 0% • Not sure – 0% • No opinion – 0%
	In addition, 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 included in the PRO measure.	<ul style="list-style-type: none"> • Required as a minimum standard – 62% • Desirable but not required – 35% • Not required – 3% • Not sure – 0% • No opinion – 0%
2.	Reliability	
	The reliability of a PRO measure should ideally be at or above 0.70 for group level comparisons.	<ul style="list-style-type: none"> • Yes, it should be at or above 0.70 - 55% • No, it should be at or above _(fill in blank_ - 8% • No minimum level of reliability, should be appropriately justified for the context of the proposed application - 35% • No opinion - 2%
	Reliability for a multi-item unidimensional scale should include an assessment of internal consistency.	<ul style="list-style-type: none"> • Required as a minimum standard – 81% • Desirable but not required – 14% • Not required – 2% • Not sure – 1% • No opinion – 2%
	Reliability for a multi-item unidimensional scale should include an assessment of test-retest reliability.	<ul style="list-style-type: none"> • Required as a minimum standard – 44% • Desirable but not required – 51% • Not required – 3% • Not sure – 2%

		<ul style="list-style-type: none"> • No opinion – 0%
	Reliability for a single item measure should be assessed by test-retest reliability	<ul style="list-style-type: none"> • Required as a minimum standard – 63% • Desirable but not required – 32% • Not required – 2% • Not sure – 2% • No opinion – 1%
3.	Validity	
3a.	- Content Validity	
	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.	<ul style="list-style-type: none"> • Required as a minimum standard – 78% • Desirable but not required – 19% • Not required – 2% • Not sure – 0% • No opinion – 1%
	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.	<ul style="list-style-type: none"> • Required as a minimum standard – 53% • Desirable but not required – 44% • Not required – 2% • Not sure – 1% • No opinion – 0%
	Documentation of the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, socio-economic status, literacy).	<ul style="list-style-type: none"> • Required as a minimum standard – 53% • Desirable but not required – 47% • Not required – 0% • Not sure – 0% • No opinion – 0%
	Documentation of sources from which items were derived, modified, and prioritized during the PRO measure development process.	<ul style="list-style-type: none"> • Required as a minimum standard – 47% • Desirable but not required – 45% • Not required – 7% • Not sure – 0% • No opinion – 0%
	Justification for the recall period for the measurement application.	<ul style="list-style-type: none"> • Required as a minimum standard – 42% • Desirable but not required – 52% • Not required – 5% • Not sure – 1%

		<ul style="list-style-type: none"> • No opinion – 0%
3b.	- Construct 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.	<ul style="list-style-type: none"> • Required as a minimum standard – 55% • Desirable but not required – 44% • Not required – 1% • Not sure – 0% • No opinion – 0%
	A PRO measure should have evidence supporting its construct validity, including documentation of empirical findings that support predefined hypotheses of the expected differences in scores between “known” groups.	<ul style="list-style-type: none"> • Required as a minimum standard – 41% • Desirable but not required – 57% • Not required – 2% • Not sure – 0% • No opinion – 0%
3c.	- Responsiveness	
	A PRO measure for use in longitudinal research study should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the target population for the research application.	<ul style="list-style-type: none"> • Required as a minimum standard – 57% • Desirable but not required – 42% • Not required – 1% • Not sure – 0% • No opinion – 0%
	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?	<ul style="list-style-type: none"> • Yes – 64% • No, I would require evidence of responsiveness before accepting it. – 33% • No opinion – 0% • Comments (fill in blank response) – 20%
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.	<ul style="list-style-type: none"> • Required as a minimum standard – 65% • Desirable but not required – 34% • Not required – 1% • Not sure – 0% • No opinion – 0%

	<p>A PRO measure should have documentation to support interpretation of scores, including representative mean(s) and standard deviation(s) in the reference population.</p>	<ul style="list-style-type: none"> • Required as a minimum standard – 40% • Desirable but not required – 56% • Not required – 4% • Not sure – 0% • No opinion – 0%
	<p>A PRO measure should have documentation to support interpretation of scores, including 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.</p>	<ul style="list-style-type: none"> • Required as a minimum standard – 23% • Desirable but not required – 72% • Not required – 5% • Not sure – 0% • No opinion – 0%

Appendix A: Literature Search Strategies

Note: Our general approach for the literature search was to solicit specific recommendations from the ISOQOL membership, along with a focused search of Medline, PsycINFO, and CINAHL. Because the search functions across these databases does not utilize a common scheme, we adapted our search to fit the database in question. All searches used adapted the general strategy outlined by Terwee et al (2009), which appeared in *Quality of Life Research*. All searches were conducted in early March 2012.

1. MEDLINE

- 1 exp Self Report/ 2232 (exploded, including all subcategories)
 - 2 exp Psychometrics/ 46812 (exploded, including all subcategories)
 - 3 exp *"Outcome Assessment (Health Care)"/ 21429 (exploded/ w focus command)
- 1 or 3 (exp Self Report = 2232 OR *"Outcome Assessment (Health Care)"/ = 21429) = 23637
AND Psychometrics/ (exploded, including all subcategories) 46812 = 909

limit to English language
864

limit to (comparative study or guideline or journal article or meta analysis or validation studies)
835 - removed 14 duplicates after export to EndNote = **821 references reviewed**

Comments – using Guidelines (from Thesaurus) proved too narrow

2. PsycINFO

"self report*" OR "patient report*" OR outcome* OR "quality of life" OR "treatment outcome*" OR "health status" OR "outcome assessment*" = TXT (all text) = 305908

5
AND AB"standards" OR AB guideline* OR AB benchmark* OR "gold standard*" OR "best practice*"
Search modes - Boolean/Phrase (66038) (not limited to ABSTRACT)
AND thesaurus term MM Psychometrics exploded 22,580
Results = **172 references reviewed**

3. CINAHL

MM = MeSH term
(MM "Self Report") = 1103
OR (MM "Checklists") = 460
OR (MM "Health Status Indicators") = (1446)
OR (MM "Treatment Outcomes+") = (12658)
OR (MM "Health Status+") (17599)
OR (MM "Outcomes (Health Care)+") (29680)
OR (MM "Outcome Assessment") (4404)
OR (MM "Quality of Life+") (17460)
AND (MM "Psychometrics")
Results = **126 references reviewed**

Appendix B: Standards for Outcomes and Comparators Selected for Use in Patient Centered Outcomes Research

Name of standard	Outcomes and Comparators used in PCOR
Description of Standard	Outcomes and comparators used in PCOR must be demonstrated as noticeable and meaningful to patients based on evidence directly elicited from people representative of the target population. For those outcomes that are best reported by patients, such as symptoms, functional status, or health-related quality of life, patient-reported and/or caregiver-reported outcome measures must be used unless rationale exists for use of another approach.
Current Practice and Examples	<p>While the traditional endpoints of survival or response to treatment are still critical in clinical research, they do not fully capture the range of outcomes that are important to patients. Indeed, in the past two decades, research has shifted to an increasingly patient-centered focus recognizing the importance of treatment side effects and their short term and long term impact on quality of life.</p> <p>The establishment of the Patient-Centered Outcomes Research Institute (PCORI) in 2010 is the best example of the priority our country has put to ensure health-related research is relevant for patients to make informed treatment decisions. The field has also recognized that there are some classes of outcomes for which patients' own report must be the gold standard. For example, fatigue, depression, and pain are by definition subjective states that cannot be meaningfully reported on by an informant (e.g., clinician, caregiver). The field has also recognized that it is feasible and valuable to collect patient-reported symptomatic adverse events they experience while participating in clinical trials [Basch, 2010]. Current practice for soliciting patient input on key outcomes and comparators for a study typically involve the use of focus groups or individual interviews with patients (and/or caregivers) who are representative of the target population of the study. In addition, recent discussions has also highlighted the value of engaging patients/patient advocates as advisors or collaborators on the research team and including their feedback early in the study design process to maximize the identification of key study research questions and outcomes to measure.</p>
Published Guidance	The FDA's (2009) "Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims" defines a PRO as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else." (p. 6). The US FDA's 2010 draft "Guidance for Industry: Qualification Process for Drug Development Tools" goes on to note that use of PRO

	measures "...that assess important aspects of patient health status and integrating them into clinical trials can make certain trials more informative concerning the benefits and risks of treatment." (p. 3-4)
Contribution to Patient Centeredness	This standard embodies "patient centeredness" as this standard requires patient input on what outcomes and comparators are important. It also supports the patient as the gold standard for self-reporting their experiences and perspectives as it relates to the measured outcomes.
Contribution to Scientific Rigor	Multiple studies have found that clinicians underreport the number and severity of symptoms relative to patients. [Litwin et al 1998; Hockenberry et al 2003; Fromme et al 2004; Weingart et al 2005; Pakhomov et al 2008; Basch et al 2009; Gawert et al 2010]. Failure of clinicians to identify these symptoms results in the occurrence of preventable adverse events [Basch 2010]. Together, this suggests that adverse event reporting by clinicians (the standard in clinical trials) imprecisely captures the negative impact of treatments on patients' lives. Thus, when selecting among treatment options, patients and physicians may not have full understanding of the toxicity associated with each treatment. For measures of efficacy, inclusions of patient-reported outcomes as measures of treatment efficacy will expand our understanding of differential impact of treatments above and beyond the traditional survival endpoint; or the PRO may be a primary endpoint in a symptom management trial.
Contribution to Transparency	Capturing and reporting symptoms, functional impact, and quality of life changes directly from patients will definitely improve our understanding of the impact of the disease and its treatment on patients' lives. This information must be summarized in easy to understand language to allow future patients to make the right choice for treatment with full knowledge of the risks and benefits.
Empirical evidence and theoretical basis	The Litwin (1998) study in prostate cancer found clinicians significantly underreported key symptoms than patients, such as bone pain (5% urologist-reported rate vs 43% patient-reported rate), fatigue (10% urologist-reported rate vs 75% patient-reported rate), erectile dysfunction (52% urologist-reported rate vs 97% patient-reported rate), incontinence (21% urologist-reported rate vs 97% patient-reported rate), and diarrhea (2% urologist-reported rate vs 33% patient-reported rate). Hockenberry et al (2003) documented agreement in fatigue ratings by children with cancer receiving chemotherapy (ages 7-12 years) with their parent, and their nurse. Congruence between child and parent ratings was moderate ($r = .35$) and between child and nurse was poor ($r = .16$). The Fromme et al (2004) prostate cancer study found clinicians missed meaningful significant changes in adverse events reported by patients (i.e., at least a 10 point change in the EORTC-QLQ-C30 score) for key symptoms including pain (65% not reported by clinicians), dyspnea (77% missed), insomnia (65% missed), anorexia (70% missed), constipation (60% missed), fatigue (38% missed), and diarrhea (30%

	missed). In a study with lung cancer patients receiving chemotherapy, Basch et al (2009) found patients reported symptoms earlier and more frequently than clinicians; but suggests there is a complementary role for both clinician-reporting and patient-reporting of symptoms. The Basch et al study (2009) found clinician-reported data was more predictive of unfavorable clinical events (death, emergency room visits), whereas patient-reported data better reflected daily health status. Gawert et al 2010 in a study of patients with rheumatoid arthritis found agreement between patients and providers never exceeded 35% for the ten most frequently reported gastro-intestinal related symptoms.
Degree of Implementation Issues	Collecting patient data requires administrative resources, however many existing PRO measures are available and the internet offers the ability to efficiently and reliably collect the data [Basch 2010]. The key to collecting meaningful data is 1) to have a precise and valid patient-reported questionnaire [Scientific Advisory Committee of the Medical Outcomes Trust 2002] that measures the domains identified by patients/patient advocates as important, and 2) to have patients complete the questionnaire at important points during the course of treatment to accurately capture the trajectory of change in scores, but not overburden the patients.
Other Considerations	There may be times within a study or with a patient population when the participant may be unable to self-report, as is the case with very young children, individuals with cognitive or communication impairments, or those who may be too ill or fatigued. However, their health status and quality of life remains extremely important to understanding the impact of the disease and its treatment. Under these circumstances, we recommended collecting proxy (e.g., caregiver, clinicians) data, especially on more observable aspects of health-related quality of life. [Addington-Hall et al 2001; US FDA 2009]. However, it must be noted that there will likely be bias in the responses, as proxies may be influenced by their own feelings about and experiences of caring for the patient [Addington-Hall et al 2001].

References

- Addington-Hall, J., Kalra, L. (2001). Who should measure quality of life? *BMJ*: 322; 1417-1420.
- Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med* 2010;362(10):865-869.
- Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, Appawu M, Iasonos A, Atkinson T, Goldfarb S, Culkin A, Kris MG, Schrag D. Adverse symptom events reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009;101:1624-1632.
- Fromme EK, Eilers KM, Mori M, Hsieh Y-C, Beer TM. How accurate is clinician reporting of chemotherapy adverse events? A comparison with patient-reported symptoms from the Quality-of-life Questionnaire C30. *Journal of Clinical Oncology* 2004;22(17):3485-3490.

Gawert L, Hierse F, Zink A, Strangfeld A. How well do patient reports reflect adverse drug reactions reported by rheumatologists? Agreement of physician- and patient-reported adverse events in patients with rheumatoid arthritis observed in the German biologics register. *Rheumatology* 2011;50(1):152-160.

Hockenberry MJ, Hinds PS, Barrera P, Bryant R, Adams-McNeil J, Hooke C, Rasco-Baggott C, Patterson-Kelly K, Gattuso JS, Manteuffel B. Three instruments to assess fatigue in children with cancer: the child, parent and staff perspectives. *Journal of Pain and Symptom Management* 2003;25:319-28

Litwin MS, Lubeck DP, Henning JM, et al. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: Results of the CaPSURE database. *J Urol* 1998;159:1988-1992.

Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care* 2008;14:530-9.

Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research* 2002;11:193-205.

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/UCM193282.pdf>
Accessed March 14, 2012.

US Food and Drug Administration. Guidance for Industry. Qualification Process for Drug Development Tools. 2010;
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>
Accessed March 14, 2012.

Weingart SN, Gandhi TK, Seger AC, et al. Patient-reported medication symptoms in primary care. *Arch Intern Med* 2005;165:234-40.

**Appendix C: Standard for the Design or Selection of
Patient-Reported Outcome Measures (PROM) for Use in Patient Centered Outcomes Research**

Name of standard	PRO Measures
Description of Standard	<p>The rationale for selection of a PROM for use in PCOR must include:</p> <ul style="list-style-type: none"> ◦ a description of the concept(s) intended to be assessed by the measure and a description of how the concept(s) relates to the goals of the study. ◦ documentation of measure development steps (for example, use of qualitative input from patients in concept elicitation and item wording, and evidence of cognitive interviewing of the final version of the measure to ensure relevance to the target population), and ◦ documentation of measurement properties including content validity, construct validity, reliability, responsiveness to change over time, score interpretability including meaningfulness of score changes in the target population, with consideration of important subgroups. <p>If key properties are not known, a plan for establishing those properties should be provided along with an explanation of the potential consequences of this lack of information on interpretation and use of results.</p> <p>This standard applies to informant reported outcome measures as well, including clinician-reported outcome measures.</p>
Current Practice and Examples	<p>In the past two decades, there has been an increased focus on the need to have better quality instruments/measures to evaluate the safety and efficacy of an intervention. This trend also reflects the emergence of patient-reported outcomes as the gold standard to measure such domains as symptoms, functioning, health, and well-being (see PCORI standard #4). Now, these standards are well accepted in the field and are used as a framework both for instrument developers, using both qualitative and quantitative methods, to design and to evaluate their survey and for investigators to select the appropriate PRO measure for their study. [e.g. Butt et al, 2005; Buysse et al, 2010; Cella et al, 2007; Gujral et al, 2007]</p>
Published Guidance	<p>The importance of the concepts outlined in the standard have been addressed by a number of previous guideline statements, including – and not limited to -- those published by the Scientific Advisory Committee of the Medical Outcomes Trust, 2002; Terwee, et al., 2007; Johnson, et al., 2011; Mokkink, et al, 2010; Revicki, et al., 2008; and the US FDA, 2009. We note that the various standards differ in their specificity and in how prescriptive their specific guidelines are worded.</p>
Contribution to Patient Centeredness	<p>The proposed standard puts the patient at the center of the outcomes assessment strategy. By ensuring that patients define and refine the concepts being measured within a study,</p>

	<p>the standard is necessarily responsive and respectful of patient preferences, needs, and values.</p> <p>It is critical that patient data be collected in a standardized way and at the right time to accurately capture the impact of the intervention on patients' lives. PRO measures provide a systematic means to collect patient data in and across different research settings. PRO measures can be completed on paper, on a computer, through the phone, or via an interviewer. The ability to collect patient data through different mediums facilitates an investigators ability to reach patient sub-groups that may be hard to reach (e.g., rural populations).</p> <p>To the extent an instrument meets and exceeds the standards described above, the better the ability for the study to evaluate the effectiveness of an intervention. Thus, PRO measures provide a key link to the patient-centeredness of PCOR. While the patient may be the gold standard to report on their condition or perspectives, there may be specific patients or study populations that are unable to provide data due to cognitive impairments, illness, or age. In such case, proxy responses may also provide informative information about the patient's status (preferred over missing data), however proxy data may have bias to the extent the outcome is less observable [Addington-Hall & Kalra. 2001].</p>
<p>Contribution to Scientific Rigor</p>	<p>The quality of a PRO measures rests on these key attributes:</p> <p><u>Conceptual and Measurement Model</u> – Definition of the measured concepts and the intended population for use of the PRO measure. This includes 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 included in the PRO measure [Scientific Advisory Committee, 2002; Terwee et al. 2007].</p> <p><u>Reliability</u> – The extent to which the measure is free from random error [Nunnally and Bernstein, 1994; Scientific Advisory Committee, 2002]. In other words, it is the extent to which a PRO measure can distinguish one group of patients from another, despite measurement error [Terwee et al. 2007].</p> <p><u>Content Validity</u> – The extent to which the PRO measure represents the most relevant and important aspects of a concept in the context of a given measurement application [Frost et al 2007].</p> <p><u>Construct Validity</u> – The extent to which scores on the PRO measure relate to other measures (e.g., patient-reported or clinical indicators) in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured [Terwee et al, 2007; Streiner et al, 2005]. Construct validity also includes expected differences in scores among groups “known” to be different.</p> <p><u>Sensitivity (also known as Responsiveness)</u> – The extent to which a PRO measure can detect changes in the construct being measured over time [Hays et al, 1992; Scientific</p>

	<p>Advisory Committee, 2002]. Responsiveness is an aspect of construct validity [Hays et al, 1992; Revicki et al, 2006].</p> <p><u>Interpretability</u> – The degree to which one (e.g., patient, clinician, researcher, policy maker) can assign meaning to a PRO measure’s scores [Scientific Advisory Committee, 2002].</p> <p><u>Meaningfulness of Score Changes</u> (also can be referred to as Minimally Important Differences, Clinically Meaningful Important Differences, Minimally Important Changes) – The smallest difference (or change) in scores that is deemed meaningful (or important, or noticeable depending on the context) either to a patient, clinician, policymaker, or other stakeholder. Note that this concept of clinical significance may have direct implication for clinical care, unlike statistical significance, which is often a function of effect size and sample size [Sloan et al, 2002].</p>
Contribution to Transparency	<p>Documentation of the evidence of the PRO measure to reflect these measurement properties, in peer reviewed literature and on publically accessible websites, will result in greater acceptance of the PRO measure for use in PCOR. To the extent the evidence was obtained from populations similar to the PCOR studies’ target population, the more confidence the investigator will have in the PRO measure to capture patient’s experiences and perspectives. That said, the standards allow flexibility on the part of the researcher with respect to explicit methods to demonstrate key measurement properties</p>
Empirical evidence and theoretical basis	<p>A survey was conducted among 98 members of the International Society for Quality of Life Research who had moderate to extensive qualitative (53%) and/or quantitative training (81%) and had an average of 15 years conducting patient-reported outcomes research. As a minimum standard for requiring evidence before use of the PRO measure in PCOR, 97% of the respondents required some evidence of reliability. For construct validity, 60% required evidence while 35% said they would expect to see evidence of content validity in most cases. For construct validity, 49% required evidence while 49% reported they would expect to see evidence of construct validity in most cases. For sensitivity/responsiveness, 26% of respondents required evidence while 52% reported it would expect to have in most cases. Only 11% of respondents reported that a minimally important difference was required before using a PRO measure while 55% reported they would expect to have a minimally important difference estimate in most applications.</p>
Degree of Implementation Issues	<p>Most of the standards have been well accepted in the field for decades and have been used as benchmarks for which to design or select PRO measures. The standard of “meaningfulness of score changes” (or MID) is a relatively newer standard and more challenging to define for a PRO measure given what is “meaningful” may vary depending on the stakeholder (e.g., patient, clinician, investigator, policy-maker) and context (e.g., clinical practice, clinical trial). It cannot be assumed that a single MID value can be appropriate for all applications and across all patient</p>

	<p>populations [Revicki et al. 2006]. In fact, there is evidence that a MID will vary if it is a worsening or improving PRO depending on the specific disease [Cella et al 2002; Yost et al 2005]. Revicki et al. [2006] conclude that the optimal approach for estimating an MID will likely be study-specific. This may have a huge impact on PCOR studies if they have to include the identification of a MID in their study.</p>
Other Considerations	<p><u>Confirmation of the Measurement Properties in PCOR</u> – The populations participating in PCOR will likely be more heterogeneous than who is typically included in a phase III type trial. This population heterogeneity should be reflected in the samples that participate in the evaluation of the measurement properties for the PRO measure. 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.</p> <p><u>Literacy Demand</u> – 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 PRO measure (e.g., cognitive testing) should include individuals with low literacy to evaluate the questions.</p> <p><u>Strength of Evidence for the Measurement Properties</u> – 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 (for example, see discussion of MID in prior section). Like any scientific discipline, there is an accumulating body of evidence examining the properties in different contexts. Thus, it is the weight of the evidence that informs the evaluation of the appropriateness of a PRO measure. Older PRO measures will have the benefit of having more evidence than younger PRO measures. This has to be reflected in the standards.</p>

References

- Addington-Hall, J., Kalra, L. (2001). Who should measure quality of life? *BMJ*: 322; 1417-1420.
- Butt, Z., Webster, K., Eisenstein, A., Beaumont, J., Eton, D., Masters, G., & Cella, D. (2005). Quality of life in lung cancer: The validity and cross-cultural applicability of the Functional Assessment of Cancer Therapy – Lung scale. *Hematology/Oncology Clinics of North America*, 19, 389-420.
- Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, Johnston KL, Shablesky-Cade MA, Pilkonis PA. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. 2010;33(6):781-792.
- Cella D, Hahn EA, Dineen K. Meaningful changes in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res* 2002;11:207-221.

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., Ader, D., Fries, J. F., Bruce, B., Rose, M., PROMIS Cooperative Group (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*, 45 (5 Suppl 1):S3-S11.

Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health* 2007;10:S94-S105.

Gujral S, Conroy T, Fleissner C, Sezer O, King PM, Avery KNL, Sylvester P, Koller M, Sprangers MAG, Blazeby JM, European Organisation for R, Treatment of Cancer Quality of Life G. Assessing quality of life in patients with colorectal cancer: an update of the EORTC quality of life questionnaire. *European Journal of Cancer*. 2007;43(10):1564-1573.

Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1992; 1:73–75.

Johnson C, Aaronson N, Blazeby JM, Bottomley A, Fayers P, Koller M, Kulis D, Ramage J, Sprangers M, Velikova G, Young T. 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.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. 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.

Nunnally JC, Bernstein IH. *Psychometric Theory* (3rd edition). New York: McGraw-Hill. 1994.

Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes* 2006;4(70):1-5.

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

Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research* 2002;11:193-205.

Sloan JA, Cella D, Frost MH, Guyatt GH, Sprangers MAG, Symonds T. Assessing clinical significance in measuring oncology patient quality of life: introduction to the symposium, content overview, and definition of terms. *Mayo Clin Proc* 2002;77(4):367-370.

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

Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007;60:34-42.

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.

Yost KJ, Cella D, Chawla A, Holmgren E, Eton T, Ayanian JZ, West DW. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution – and anchor-based approaches. *J Clin Epidemiol* 2005;58:1241-1251.