Enhancing the Patient's Voice: Draft Guidance on Minimum Standards for the Design and Selection of Patient-Reported Outcomes (PRO) Measures for use in Patient Centered Outcomes Research

NORTHWESTERN UNIVERSITY

Zeeshan Butt, PhD
David Cella, PhD
Richard Gershon, PhD
Elizabeth Rauch
Kathleen Swantek, MLS

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Bryce Reeve, PhD
Jessica Lyons, MS

ISOQOL Scientific Advisory Task Force (SATF)

- **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**
  RTI – Health Solutions
- **Carol Moinpour, PhD**
  Fred Hutchinson Cancer Research Center
- **Dennis Revicki, PhD**
  United BioSource Corporation
- **Carolyn Schwartz, ScD**
  DeltaQuest Found. / Tufts Univ. Med. Sch.
- **Claire Snyder, PhD**
  Johns Hopkins University
- **Caroline Terwee, PhD**
  VU University Medical Center
- **Galina Velikova, MD, PhD**
  University of Leeds
- **Albert Wu, MD**
  Johns Hopkins University
- **Kathleen Wyrwich, PhD**
  United BioSource Corporation

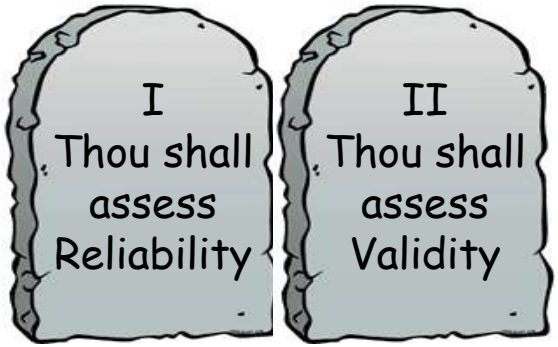
NORTHWESTERN UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

- Comprehensive literature review
- Interact with ISOQOL SATF
- Field ISOQOL member survey.

Reports

How we spent our winter vacation.



Literature Review

- **Sources**
  - ISOQOL membership
  - Pre-contract activity
  - Pubmed
  - PsycINFO
  - CINAHL

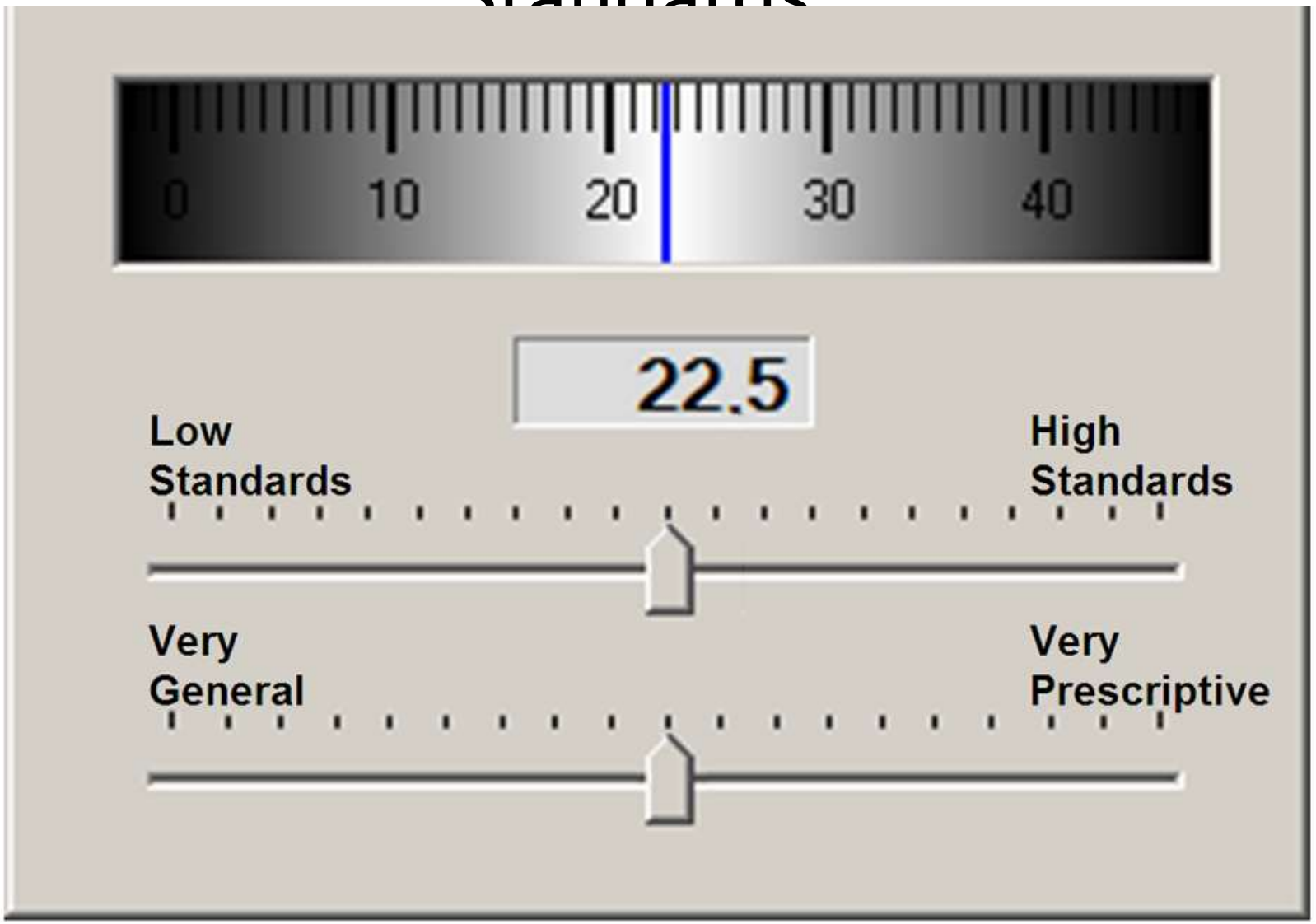
- **301**
- *24 reviewed* → **12 unique**
- *544 reviewed* → **14 unique**
- *172 reviewed* → **22 unique**
- *126 reviewed* → **4 unique**

- **Search Parameters**
  - Adapted the methodological search filter developed by Terwee et al (2009) *Qual Life Res*
  - Focus was on consensus statements, guidelines, and evidence-based papers
  - Included population-, but not instrument-specific papers, if concepts were generalizable
  - Included unpublished and published sources

ISOQOL Membership Survey

- Designed with ISOQOL SATF
- Sought consensus on draft recommendations presented today and explored more in depth on specific issues of debate.
- Study was approved by UNC IRB.
- Sent out on Feb. 20 to approximately 500 ISOQOL members with deadline of Feb. 29.
- ISOQOL SATF members reviewed responses to survey.

Challenges with writing recommendations for Minimum Standards



Instructions on ISOQOL Survey

Please remember as you answer the questions in this survey that we are developing the **minimum** standards for the selection and design of a PRO measure for use in patient-centered outcomes research (PCOR).

That is, we are saying a PRO measure that does not meet the "minimum standard" should not be considered appropriate for the research study.

Quality of Life Research **11**: 193–205, 2002.

Assessing health status and quality-
criteria

Scientific Advisory Committee of the Medical O

Accepted in revised form 8 January 2002

Abstract

The field of health status and quality of life (Qo
theoretical framework, accepted methods, and di
of 30 years. To identify health status and QoL i
precursor to creating an instrument library for
created an independently functioning Scientific
defined a set of attributes and criteria to carry
revised these materials to take account of the
instruments were being developed. This paper
attributes of health status and QoL instrumen
validity; responsiveness; interpretability; respond
tural and language adaptations) and the criteria
attributes. These are suggested guidelines for the
become both more familiar and more sophisticat
these criteria accordingly.

Key words: Health status, Item response theory,
Validity

Introduction

The field of health assessment

The field of health status and quality of life (QoL
measurement – as a formal discipline with a c

[1] Neil Aaronson, PhD, The Netherlands Cancer Institut
Amsterdam; Jordi Alonso, MD, Institut Municipal d'Invest
gacio Medica (IMIS-IMAS), Barcelona, Spain; Audrey Bu
nam, PhD, The RAND Corporation, Santa Monica, C
Kathleen N. Lohr, PhD, RTI International, Research Triang
Park, NC, and Program on Health Outcomes, University
North Carolina at Chapel Hill; Donald L. Patrick, Phl
MSPH, Department of Health Services, University of Was
ington, Seattle; Edward Perrin, PhD, Department of Healt
Services, University of Washington, Seattle, Ruth E.K. Stei
MD Albert Einstein College of Medicine/Children's Hospital
Montefiore, Bronx, NY.

Qual Life Res (2010) 19:539–549
DOI 10.1007/s11136-010-9606-8

The COSMIN checklist for
of studies on measurement
measurement instruments: a

**Lidwine B. Mokkink · Caroline B. Terwee ·
Donald L. Patrick · Jordi Alonso · Paul W. S
Dirk L. Knol · Lex M. Bouter · Henrica C. W**

Accepted: 2 February 2010 / Published online: 19 Febru

Abstract

Background Aim of the COSMIN study (C
based Standards for the selection of health st
surement INstruments) was to develop a consc
checklist to evaluate the methodological quality
on measurement properties. We present the
checklist and the agreement of the panel on the i
checklist.

Methods A four-round Delphi study was perfo
international experts (psychologists, epidemiolo
isticians and clinicians). Of the 91 invited e
agreed to participate (63%). Panel members we
rate their (dis)agreement with each proposal on a
scale. Consensus was considered to be reache

L. B. Mokkink (✉) · C. B. Terwee · D. L. Knol ·
L. M. Bouter · H. C. W. de Vet
Department of Epidemiology and Biostatistics and the E
Institute for Health and Care Research, VU University
Center, Van der Boechorststraat 7, 1081 BT Amsterdam
The Netherlands
e-mail: w.mokkink@vumc.nl
URL: www.emgo.nl; www.cosmin.nl

C. B. Terwee
e-mail: cb.terwee@vumc.nl

D. L. Knol
e-mail: d.knol@vumc.nl

L. M. Bouter
e-mail: lm.bouter@dienst.vu.nl

H. C. W. de Vet
e-mail: hcw.devet@vumc.nl

D. L. Patrick
Department of Health Services, University of Washingt
Thur Canal St Research Office, 146N Canal Suite 310,
WA 98103, USA
e-mail: donald@u.washington.edu

Guidance for Industry

Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2009
Clinical/Medical

Attributes of a PRO Measure

#	Attribute
1	Conceptual and Measurement Model
2	Reliability
3	Validity
3a	- Content Validity
3b	- Construct Validity
3c	- Responsiveness
4	Interpretability of Scores
5	Translations
6	Patient and Administrator Burden

ISOQOL Sample Characteristics (n = 98)

- **Degrees**: 62% PhD; 17% MD; 45% Masters
- **Role**: 71% Academic; 19% Clinicians; 19% Industry Consultant; 8% Industry; 7% Government
- **Geographic**: 48% N. A. (85% US); 33% Europe; 9% Asia; 6% S.A.; 3% Australia; 1% Africa; 0% Antarctica
- **Psychometric training**: 81% mod – extensive; 16% little
- **Qualitative training**: 53% mod – extensive; 40% little
- **Competency**: 50% very competent; 39% competent; 8% somewhat competent
- **Average # of years HRQOL/PRO research**: 15 years (range 1 – 40 years)

Survey Question: Please provide your opinion on the following items regarding the **minimum standard** for a PRO measure:

RECOMMENDATION

- Required as a minimum standard
- Desirable but not required as a minimum standard
- Not required at all (not needed for a PRO measure)
- Not sure
- No opinion

Survey Question: Please provide your opinion on the following items regarding the **minimum standard** for a PRO measure's conceptual and measurement model:

RECOMMENDATION

General rule:
> 50% accepted;
< 50% consider as "best practice"

- **Required as a minimum standard**
- Desirable but not required as a minimum standard
- Not required at all (not needed for a PRO measure)
- Not sure
- No opinion

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. *(91% required)*

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. *(62% required)*

2. Reliability

The reliability of a PRO measure should ideally be at or above 0.70 for group level comparisons.

- *55% agree;*
- *35% → "no minimum level…it should be appropriately justified for the context of the proposed application"*

Reliability for multi-item scales should include an assessment of internal consistency *(81% required)* **and test-retest reliability;** *(44% required)*

Differences of opinion

...and reliability for a single item measure should be assessed by test-retest reliability. *(63% required)*

3. Validity: How critical is each type of validity?

	Content	Construct	Responsiveness	Criterion
Cross-sectional studies				
Must have, or I wouldn't use PROM	60%	49%	26%	8%
Would expect to have in most cases	35%	49%	52%	35%
Nice to have, but not critical	4%	1%	14%	44%
Not critical at all.			7%	11%
Longitudinal studies				
Must have, or I wouldn't use PROM	63%	45%	60%	11%
Would expect to have in most cases	33%	52%	38%	34%
Nice to have, but not critical	4%	3%	2%	41%
Not critical at all.				12%

Note: If a PRO Measure had cross-sectional data to support reliability and validity (content, construct); but no data on responsiveness: 64% would use PROM; 33% would require evidence of responsiveness before using it.

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. This includes 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;**

- **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;**

- **ways from which items were developed and/or sources from which items were derived, modified, and prioritized during the PRO measure development process; and**

- **justification for the recall period for the measurement application.**

3a. Content Validity (part i)

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. *(78% required)*

3a. Content Validity (part ii)

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;** *(53% required)*
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;** *(53% required)*
3. **ways from which items were developed and/or sources from which items were derived, modified, and prioritized during the PRO measure development process;** *(47% required)*
4. **justification for the recall period for the measurement application.** *(42% required)*

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,** *(55% required)*
- **or expected differences in scores on that PRO measure between "known" groups.** *(41% required)*

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. *(57% required)*

Note: However, there may be circumstances in which a PRO measure with cross-sectional data to support reliability and validity (content, construct), but no data on responsiveness, could be used (64% supporting such use).

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;** *(65% required)*
- **representative mean(s) and standard deviation(s) in the reference population;** *(40% required)*
- **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.** *(23% required)*

5. Translations

A PRO measure translated to one or more languages should have evidence of the equivalence of measurement properties for translated versions, allowing comparison or combination of data across language forms. *(48% required)*

	Must have	Expect in most cases	Nice to have, but not critical
Qualitative evidence	42%	39%	18%
Quantitative evidence	23%	43%	30%

This includes documentation of:

1) **background and experience of the persons involved in the translation;** *(42% required)*
2) **methods used to translate and evaluate the PRO measure in each language;** *(81% required)*
3) **extent of harmonization across different language versions.** *(38% required)*

5. Translations

A PRO measure translated to one or more languages should have documentation of the methods used to translate and evaluate the PRO measure in each language. Studies should at least include evidence from qualitative methods (e.g., cognitive testing) to evaluate the translations.

Proposed new wording!!!!!

6. Patient and Administrator Burden

A PRO measure must not be overly burdensome for patients or administrators. The length of the PRO measure should be considered in the context of other PRO measures included in the assessment, the frequency of PRO data collection, and the characteristics of the study population.

The literacy demand of the items in the PRO measure should usually be at a 6th grade education level or lower.

- 4th grade: 6%
- 6th grade: 23%
- 8th grade: 6%
- 44% endorsed "There should be no minimum requirement for the literacy level of the PRO measure; however, it should be appropriately justified for the context of its proposed application."

6. Patient and Administrator Burden

A PRO measure must not be overly burdensome for patients or administrators. The length of the PRO measure should be considered in the context of other PRO measures included in the assessment, the frequency of PRO data collection, and the characteristics of the study population.

The literacy demand of the items in the PRO measure should usually be at a 6th grade education level or lower; however, it should be appropriately justified for the context of the proposed application.

Proposed new wording!!!!!

How can these standards be used by a PRO Measure Developer or Investigator?

#	Attribute
1	Conceptual and Measurement Model
2	Reliability
3	Validity
3a	- Content Validity
3b	- Construct Validity
3c	- Responsiveness
4	Interpretability of Scores
5	Translations
6	Patient and Administrator Burden

How can these standards be used by a PRO Measure Developer?

#	Attribute	Method	Results
1	Conceptual and Measurement Model	PRO defined and items generated through expert panel.	This domain includes 4 attributes measured by 20 items.
2	Reliability	Internal Consistency (Cronbach's alpha) at baseline; Test-retest between Baseline & 1 day later.	Internal consistency (n = 258) α = .84; Test-retest (n = 50) r = .72
3	Validity		
3a	- Content Validity	Literature review and focus groups to identify sub-domains of concept. Cognitive testing to evaluate relevance of items and modify problematic items.	2 focus groups (n = 17) identified 4 attributes. Consistent with theory of Awesomeness. Cog. Testing (n = 8, round 1; n=6, round 2) 4 items modified.
3b	- Construct Validity		
3c	- Responsiveness		
4	Interpretability of Scores		
5	Translations		
6	Patient and Administrator Burden		

How can these standards be used by an Investigator? (1)

#	Attribute	Evidence?	Relevance & needs for current research application
1	Conceptual and Measurement Model	Definition…. ; 4 Attributes with 20 items. In Smith et al. 2011.	PRO is highly prevalent and bothersome in our population.
2	Reliability	Internal consistency (n = 258) α = .84; Test-retest (n = 50) r = .72. in Smith et al. 2011.	Acceptable for prospective study among 2 arms.
3	Validity		Validation study in Thomas et al was older than our target pop.
3a	- Content Validity	2 focus groups in older men with prostate cancer (Thomas et al. 2012)	Need to conduct 2 focus groups with young prostate cancer patients to confirm relevant attributes.
3b	- Construct Validity		
3c	- Responsiveness		
4	Interpretability of Scores		
5	Translations		
6	Patient and Administrator Burden		

How can these standards be used by an Investigator? (2)

#	Attribute	Evidence for Instrument A	Evidence for Instrument B
1	Conceptual and Measurement Model	Definition…. ; 4 Attributes with 20 items.	Definition….; 2 Attributes with 16 items.
2	Reliability	Internal consistency (n = 258) α = .84; Test-retest (n = 50) r = .72.	Internal consistency (n = 312) α = .91; Test-retest (n = 78) r = .74.
3	Validity		
3a	- Content Validity	2 focus groups in older men with prostate cancer.	3 focus groups in younger and older prostate cancer population.
3b	- Construct Validity		
3c	- Responsiveness		
4	Interpretability of Scores		
5	Translations		
6	Patient and Administrator Burden		

Beyond the PCORI Contract Period

- Develop “Ideal” or “Best Practices” standards...
- PRO Measure Maturity (or Stepped) Model
  - From PROMIS, “… describes the stages of instrument scientific development from conceptualization through evidence of psychometric properties in multiple diverse populations.”
- Standards for PRO Measures to be used in the healthcare setting.