Selecting a Patient-Reported Outcome Measure:
Minimum Standards from the International Society for Quality of Life Research

Bryce B. Reeve, PhD

Funded by the Patient-Centered Outcomes Research Institute and Genentech

PROTEUS
Patient-Reported Outcomes Tools: Engaging Users and Stakeholders

The PROTEUS Consortium

Patient-Reported Outcome Tools: Engaging Users & Stakeholders

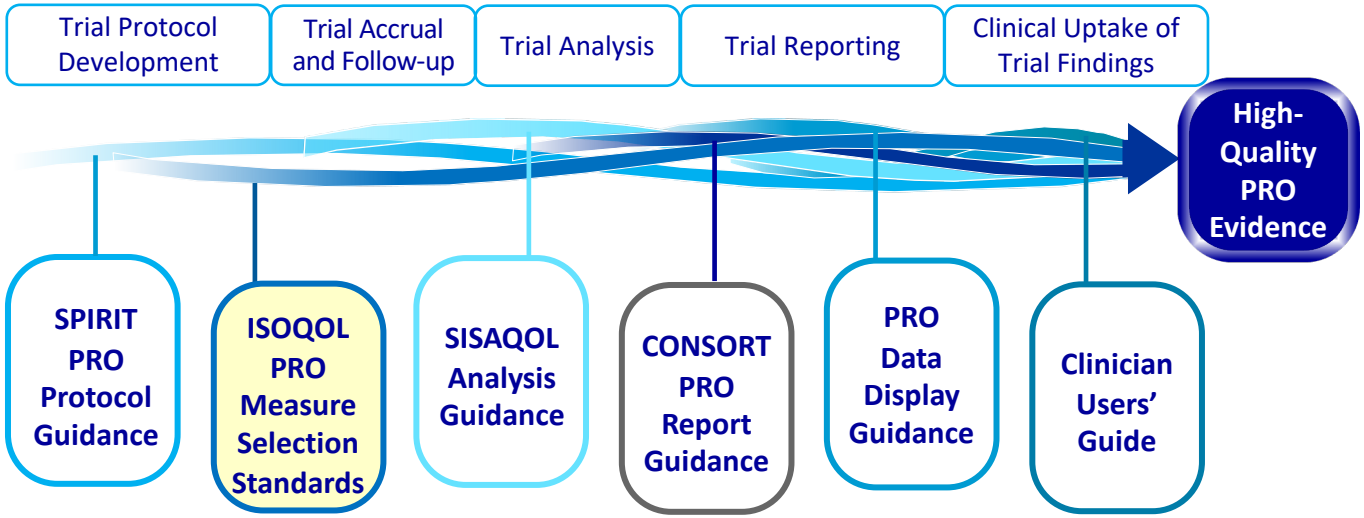
PROTEUS

TheProteusConsortium.org

Overview of Presentations

Introduction to PROs and PROTEUS

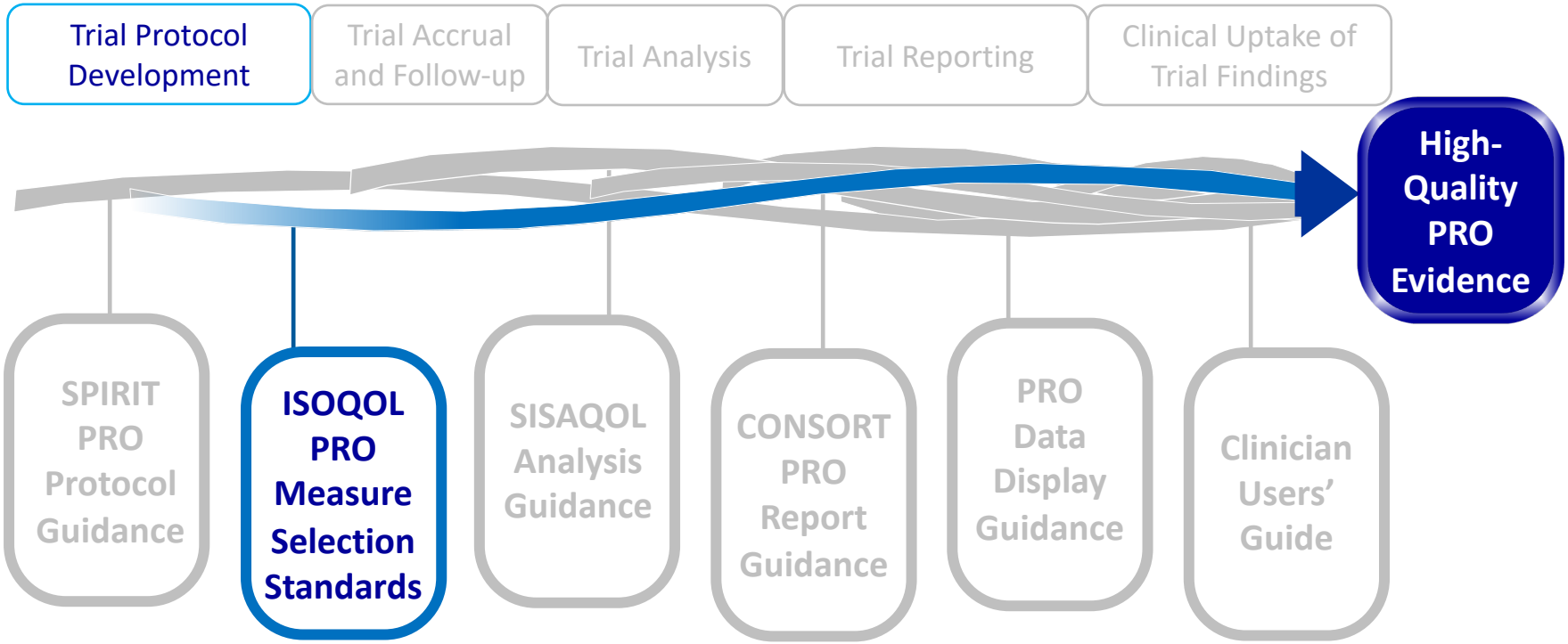
Introduction to the PROTEUS Tools



Overview of Tool Recommendations

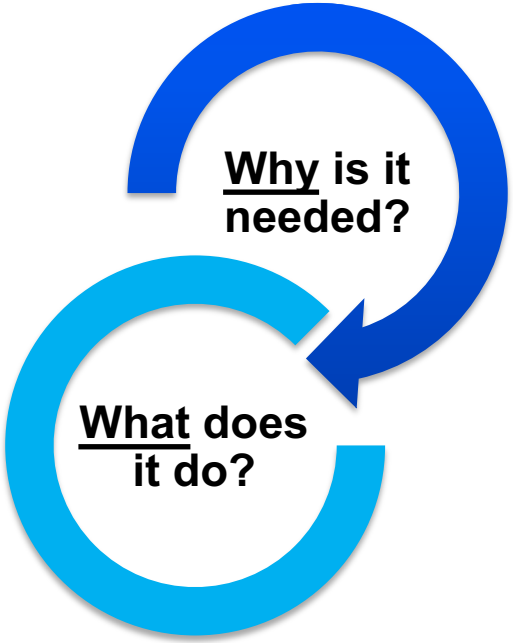
How to Apply the Tools

Minimum Standards for PRO Measures



PROTEUS

Measuring PROs Effectively



PROs must be measured in a valid, standardized way using appropriate tools and methods to ensure valid conclusions

Provides guidance for selecting PRO measures for use in patient-centered and comparative effectiveness research

PROTEUS

Minimum Standards for PRO Measures

Qual Life Res (2013) 22:1889–1905
DOI 10.1007/s11136-012-0344-y

ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research

Bryce B. Reeve · Kathleen W. Wyrwich · Albert W. Wu · Galina Velikova · Caroline B. Terwee · Claire F. Snyder · Carolyn Schwartz · Dennis A. Revicki · Carol M. Moinpour · Lori D. McLeod · Jessica C. Lyons · William R. Lenderking · Pamela S. Hinds · Ron D. Hays · Joanne Greenhalgh · Richard Gershon · David Feeny · Peter M. Fayers · David Cella · Michael Brundage · Sara Ahmed · Neil K. Aaronson · Zeeshan Butt

Accepted: 17 December 2012 / Published online: 4 January 2013
© Springer Science+Business Media Dordrecht 2013

Reeve et al, Qual Life Res 2013, 22, 1889-1905

PROTEUS

Acknowledgements

ISOQOL Scientific Advisory Task Force (SATF)

Neil Aaronson, PhD
University of Amsterdam
Sara Ahmed, PhD
McGill University
Michael Brundage, MD
Queens University
Zeeshan Butt, PhD***
Northwestern University
David Cella, PhD
Northwestern University
Peter Fayers, PhD
University of Aberdeen
David Feeny, PhD
University of Alberta
Richard Gershon, PhD
Northwestern University
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
Jessica Lyons, MS
University of North Carolina
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

ISOQOL
International Society for Quality of Life Research

PROTEUS

Why We Need PRO Measure Selection Standards

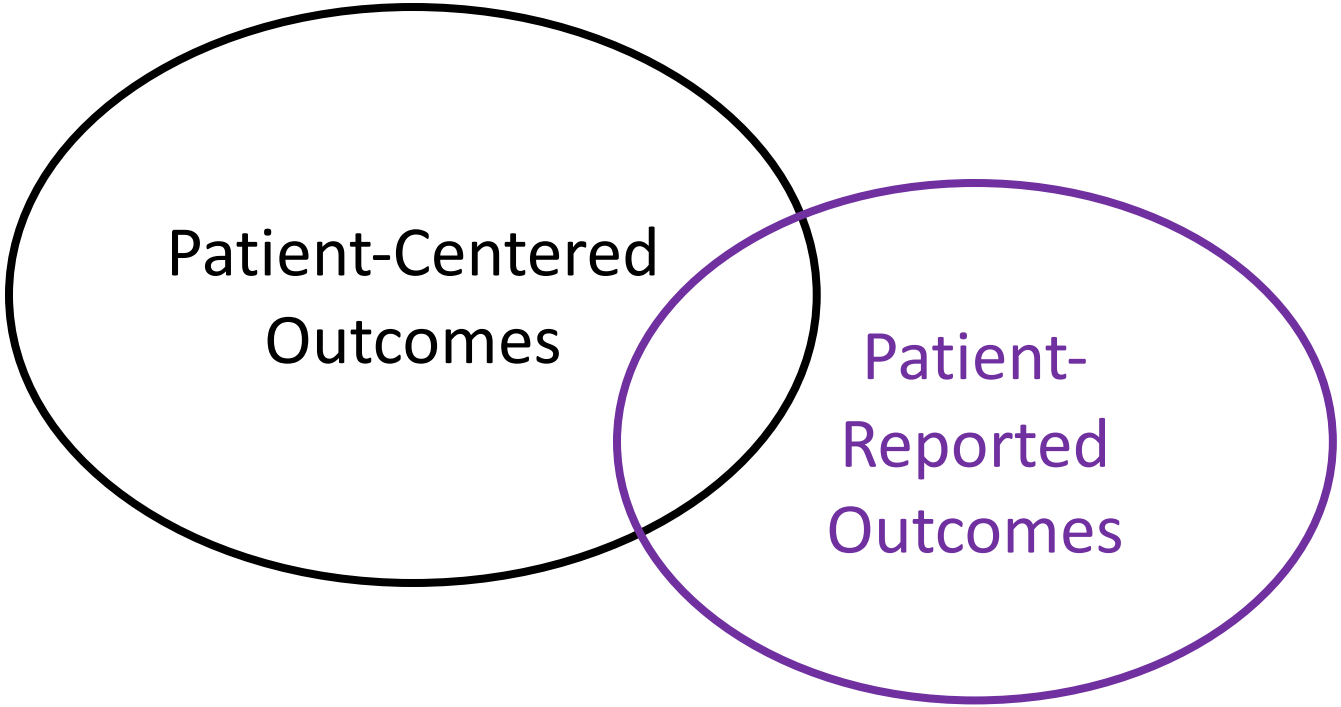
- An essential aspect of patient-centered outcomes research (PCOR) and comparative effectiveness research (CER) is integration of patient perspectives and experiences about their health with clinical and biological data to evaluate the safety and effectiveness of interventions
- Widely accepted that patient's report is the best source of information about what they are experiencing
- Challenge for PCOR and CER is how to best capture patient-reported data to inform decision making in healthcare delivery, research, and policy settings

Why We Need PRO Measure Selection Standards

- To draw valid research conclusions regarding patient-centered outcomes, PROs must be measured in a standardized way using appropriate methods
- A PRO is the measurement of any aspect of a patient's health that comes directly from them without interpretation by another[1]
- PROs can be symptoms (e.g. pain, anxiety, nausea, fatigue), aspects of functioning (e.g. role, physical, emotional, social) and multidimensional constructs (e.g. HRQOL)
- A PRO measure is the questionnaire, index, checklist, instrument, or tool[2], along with the algorithm used to score patient responses into summary scores for analysis and reporting

[1] *https://www.fda.gov/media/77832/download*
[2] *Mayo, 2015 ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement*

Why We Need PRO Measure Selection Standards



PROTEUS

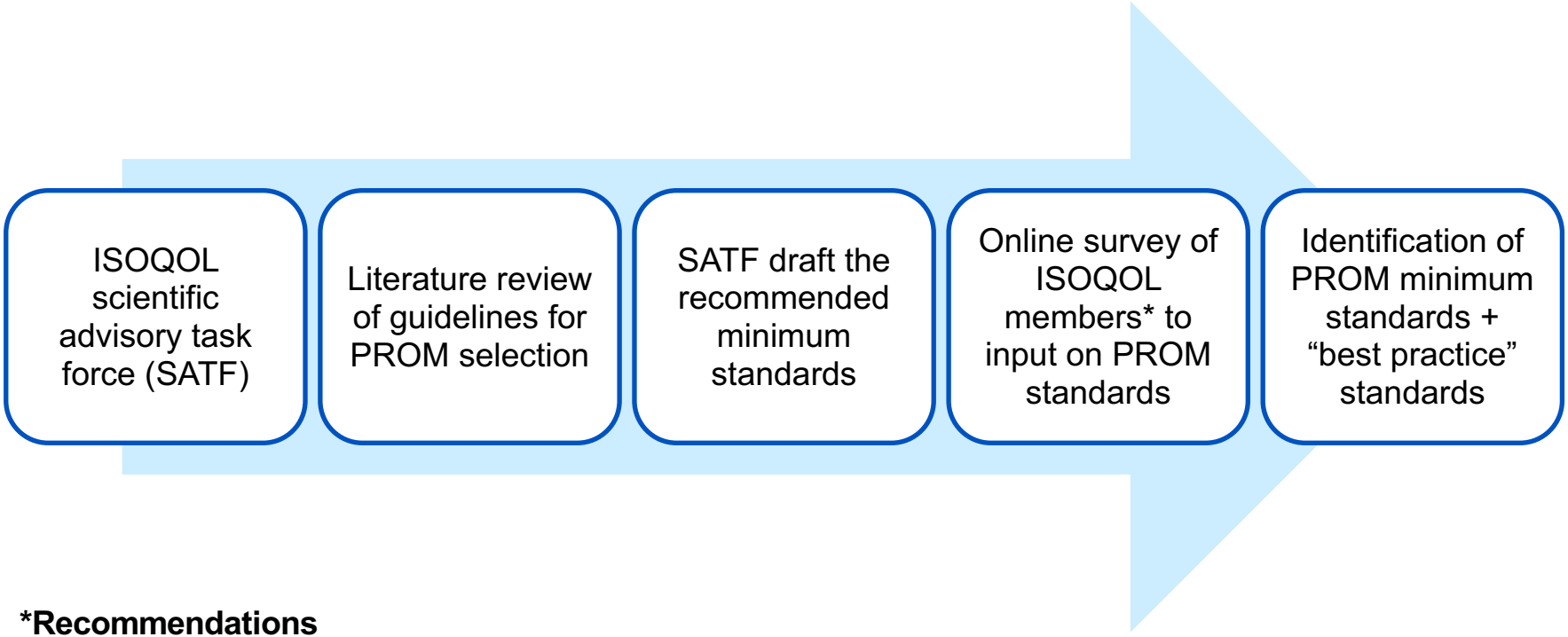
Objective of the Resource

- To develop PRO measure minimum standards for the design and selection of a PRO measure for use in PCOR and CER

- The standards represent the minimum criteria required for a PRO measure to be judged suitable for a PCOR or CER study

- These standards are intended to promote the appropriate use of PRO measures in PCOR and CER, which in turn can improve the effectiveness and efficiency of healthcare delivery

Methods for Resource Development



***Recommendations**

>50% had to endorse as "required as a minimum standard"

PROTEUS

Overview of PRO Measure Selection Standards

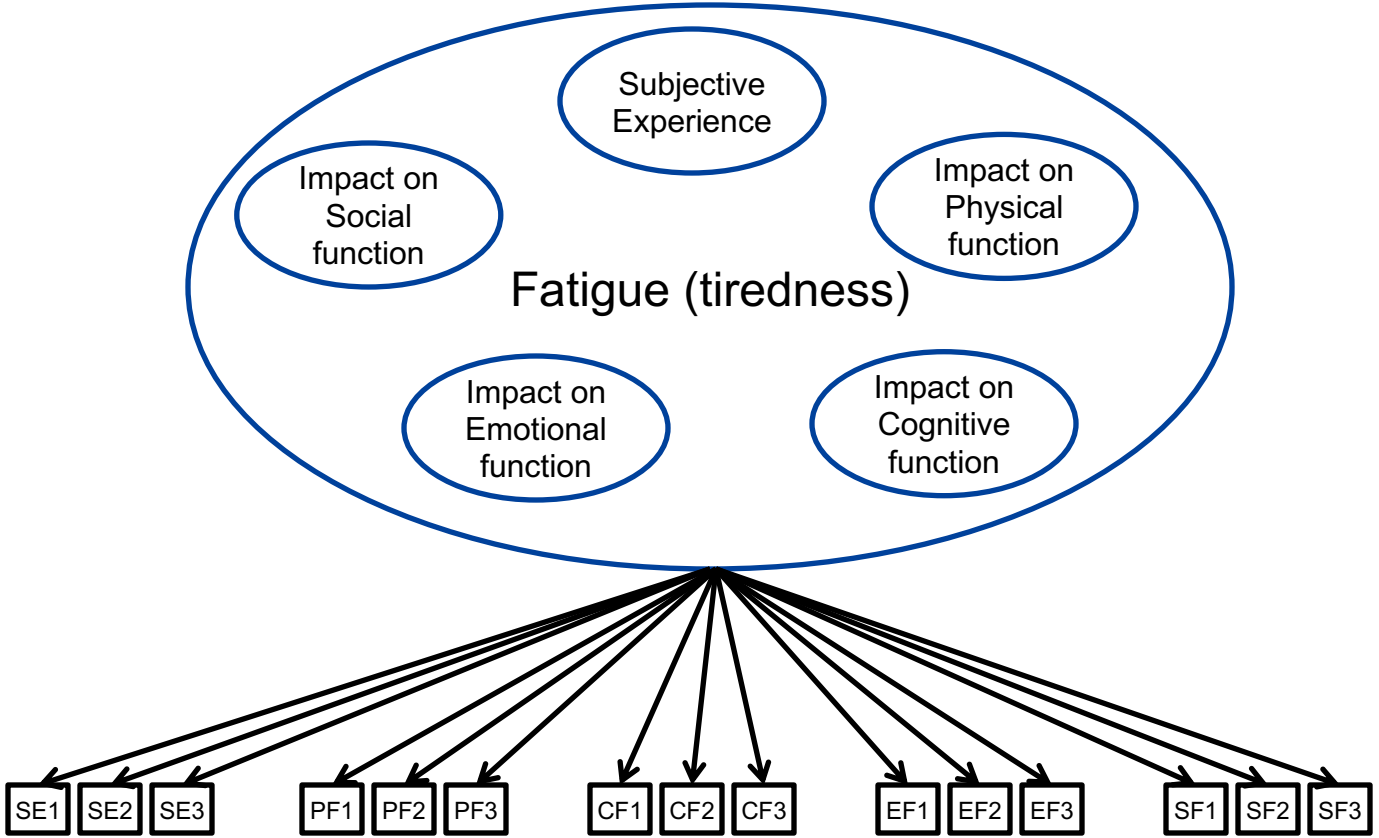
1. Conceptual and measurement model
2. Reliability
   a. Internal consistency
   b. Test-retest
3. Validity
   a. Content
   b. Construct
   c. Responsiveness
4. Interpretability of scores
5. Translations
6. Patient and investigator burden

Conceptual and Measurement Model

A PRO measure should have documentation:

1. defining and describing the concept(s) included and the intended population(s) for use

2. 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



Barsevick et al, J Pain Symptom Manage 2010, 39(6), 1086-1099

Reliability

Degree to which an instrument is free from measurement error.

"…expresses how well patients with true systematic differences can be distinguished from each other in spite of, or after accounting for, the presence of measurement error." (*p. 53)

Cappelleri et al, Patient-Reported Outcomes Measurement, Implementation and Interpretation. 2014, Boca Raton, FL.

Reliability

Two types of reliability for PRO measures:

1. Internal Consistency (for multi-item scales)
2. Test-Retest

The reliability of a PRO measure should preferably be at or above 0.70 for group level comparisons, but may be lower if appropriately justified.

Validity

Extent a scale measures what it purports to measure.

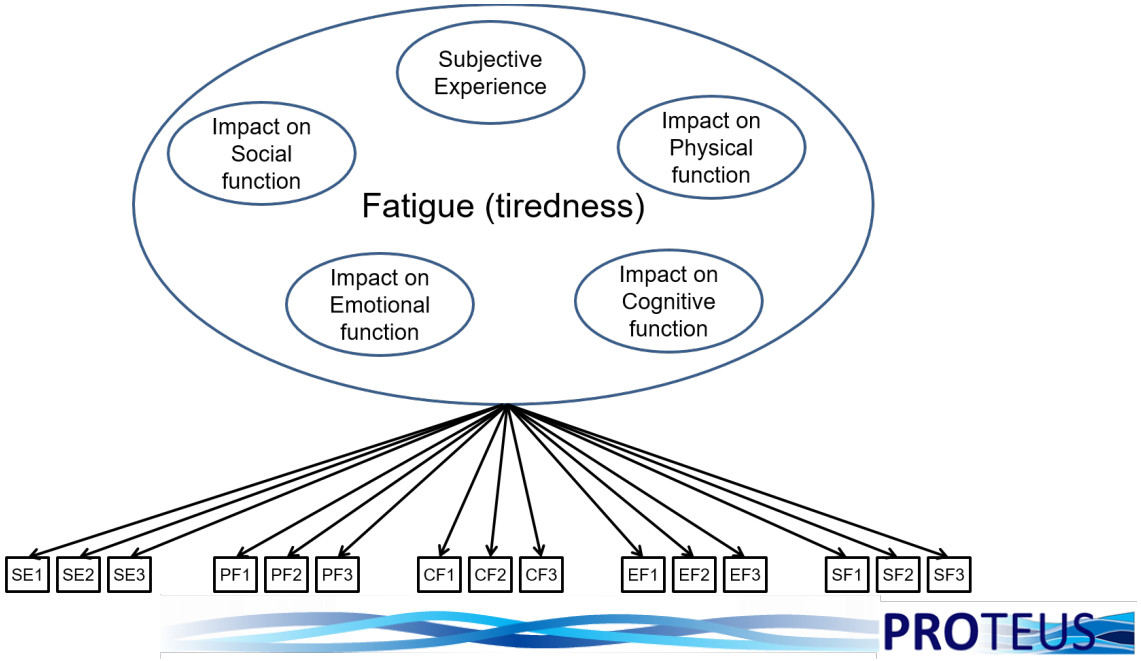
There are multiple types of validity; the more frequently assessed types for PRO measures are:

1. Content Validity
2. Construct Validity
3. Responsiveness

Content Validity

Extent to which the PRO instrument measures the appropriate content and represents the variety of attributes that make up the measured construct.

Bryce's new Fatigue Measure



PROTEUS

Content Validity

A PRO measure should have evidence supporting its content validity, including evidence that patients and 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:

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,
3) justification for the recall period for the measurement application.

Construct Validity

Extent to which the PRO measure “behaves” in a way consistent with theoretical hypotheses and represents how well scores on the instrument are indicative of the theoretical construct.

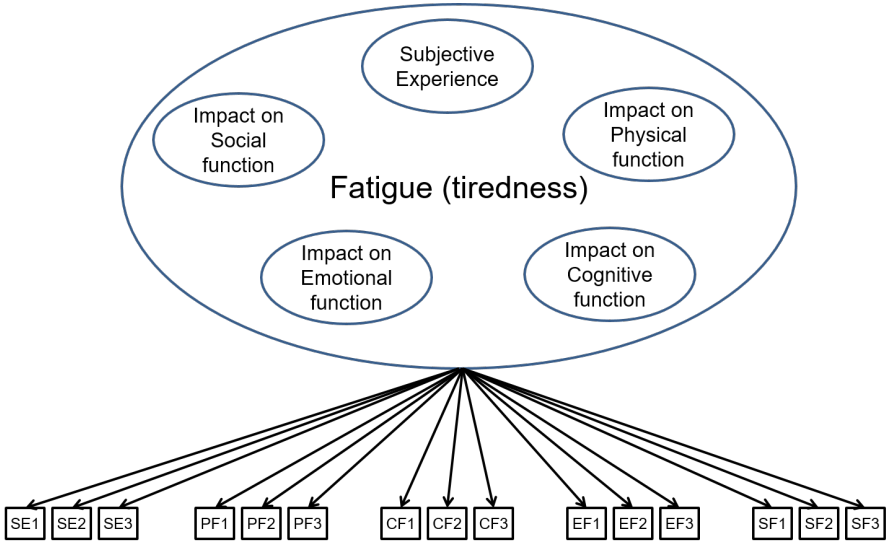
4 types of construct validity:

1. Structural Validity
2. Convergent Validity
3. Discriminant Validity
4. Known Groups Validity

Construct Validity

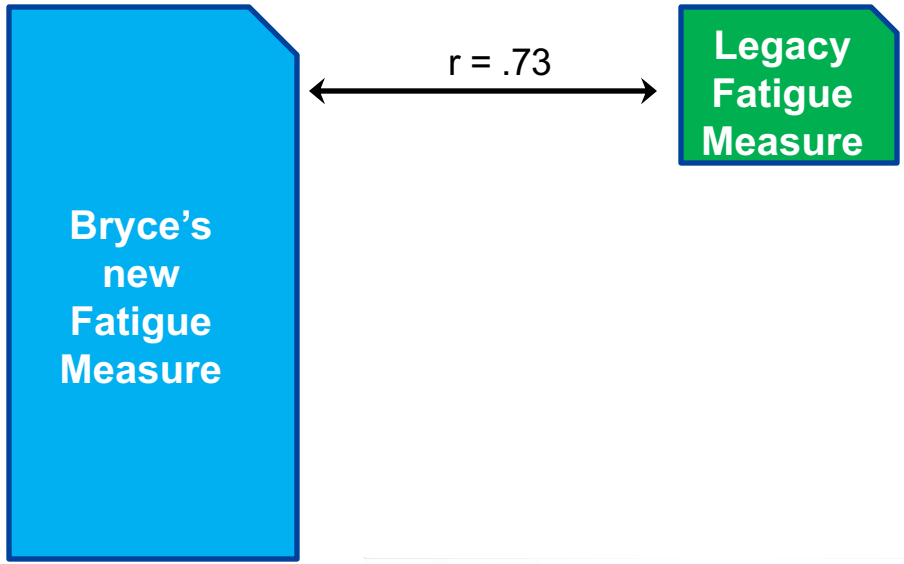
Structural Validity – extent the empirical data support the conceptual model.

Bryce's new Fatigue Measure



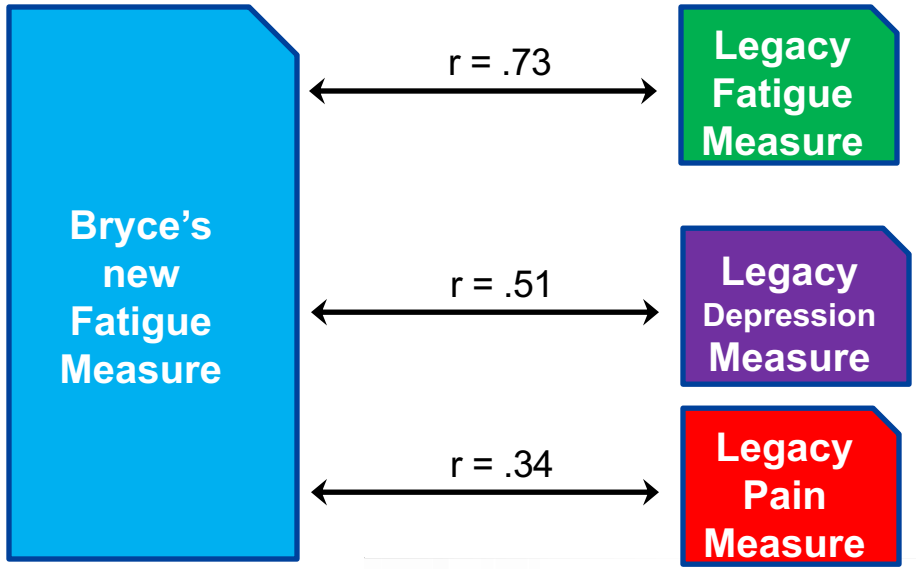
Construct Validity

Convergent Validity – extent the PRO measure is similar to other established measures measuring the same concept.



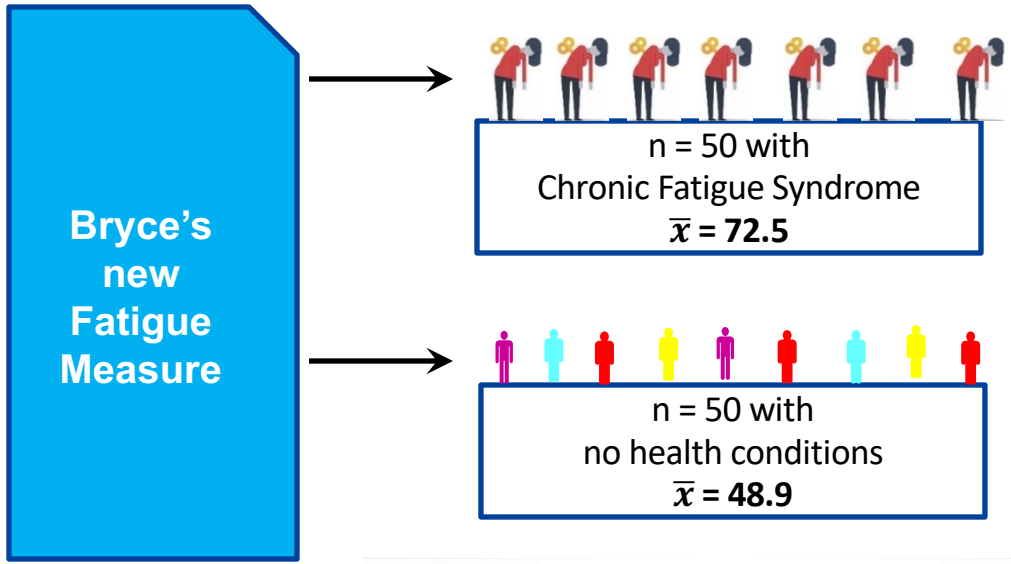
Construct Validity

Discriminant Validity – extent the PRO measure is dissimilar to other established measures measuring different concepts.



Construct Validity

Known Groups Validity – extent the PRO measure can differentiate between groups known to differ on the measured concept.



Construct Validity (standard)

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.

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 measured PRO in the target population for the research application.

Baseline

n = 50 with
Chronic Fatigue Syndrome
$\bar{x}$ = 72.5

Bryce's new Fatigue Measure

Proven Interval Training Therapy

6-week Follow-up

n = 48 with
Chronic Fatigue Syndrome
$\bar{x}$ = 61.3

Bryce's new Fatigue Measure

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.

Your fatigue score is:

63

Translation of the PRO Measure

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.

Patient and Investigator Burden

A PRO measure must not be overly burdensome for patients or investigators:

- length of the PRO measure should be considered in the context of other PRO measures included in the assessment,
- frequency of PRO data collection,
- literacy demand of the items in the PRO measure should usually be at a 6th grade education level or lower (i.e., 12 year old or lower); however, it should be appropriately justified for the context of the proposed application.

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 Examples
1	Conceptual and Measurement Model	Conduct concept elicitation interviews among stakeholders to define the PRO construct and its attributes. Review literature.
2	Reliability	Longitudinal pilot study*: Internal Consistency (Cronbach's alpha) at baseline; Test-retest between Baseline & 1 day later.
3	Validity	
3a	- Content Validity	Conduct multiple rounds of cognitive testing interviews to evaluate interpretability of PRO measure items and their relevance.
3b	- Construct Validity	In pilot study*, include other existing/established measures of similar and related constructs as the PRO measure. Include sub-groups whose scores should be different on the PRO measure. Conduct factor analysis of items to evaluate fit to conceptual model.
3c	- Responsiveness	In pilot study*, collect longitudinal data on the PRO measure and other key health indicators at times when the participants are expected to change on the PRO being evaluated.
4	Interpretability of Scores	The conceptual and measurement model should define scoring direction. Collect data in representative samples to create reference or norm scores.
5	Translations	Conduct language translations following best practice guidelines and linguistic validation using cognitive testing and psychometric testing.
6	Patient and Administrator Burden	Consider appropriate length of measure for its intended use balanced with need for a valid and reliable scale. Consider appropriateness of administration mode (web-based, phone)

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

#	Attribute	Current 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 study target 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	Cognitive testing in older men with prostate cancer (Thomas et al. 2012)	Need to conduct cognitive testing with young prostate cancer patients to confirm relevant attributes.
3b	- Construct Validity	Evidence for convergent validity	Need to test for known-groups validity.
3c	- Responsiveness	No data	Need to conduct pilot study to examine changes over time.
4	Interpretability of Scores	High scores represent more of the construct being measured	Conduct study to estimate minimally important differences (MID)
5	Translations	Available in English and Spanish	Translate to French and validate translation.
6	Patient and Administrator Burden	Tested in web-based platform	Evaluate validity for administering over phone by live interviewers.

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	Evidence for convergent, discriminant, and known groups validity	Evidence for convergent validity.
3c	- Responsiveness	No evidence	No evidence
4	Interpretability of Scores	Low and High scores defined	Low and high scores defined, MID = 5 points
5	Translations	English, Spanish, French, Simplified Chinese	English only
6	Patient and Administrator Burden	20 items	16 items

Recap

Minimum standards for the selection of PRO measures for use in PCOR and CER include:

- Conceptual and measurement model
- Reliability (internal consistency, test-retest)
- Validity (Content and Construct)
- Responsiveness
- Interpretability of scores
- Translations
- Patient and Investigator Burden

If a PRO measure does not meet these criteria, it is considered NOT suitable for a PCOR or CER study.

Further Reading

Reeve BB, Wyrwich KW, Wu AW, et al. (2013) ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*; 22:1889–1905.

*Crossnohere NL, Brundage M, Calvert MJ, King M, Reeve BB, Thorner E, Wu AW, Snyder C. (2020) International guidance on the selection of patient-reported outcome measures in clinical trials: A review. *Qual Life Res*. 14 Sept [Epub ahead of print].

** This paper reviews alternative guidance documents regarding selecting a PRO measure and compares the recommendations to the ISOQOL Minimum Standards.*