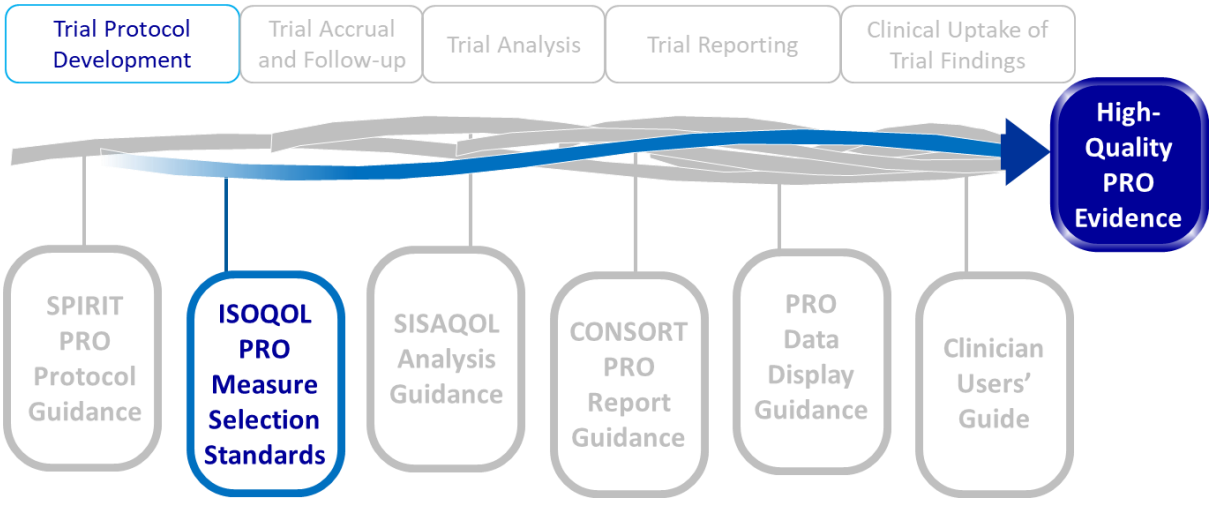
Chapter 3. Selecting PRO Measures



ISOQOL Minimum Standards for PRO Measures in Patient-Centered and Comparative Effectiveness Research

In 2013, the International Society for Quality of Life Research (ISOQOL) led an initiative to inform the selection of PRO measures for use in patient-centered outcomes and comparative effectiveness research by identifying minimum standards. These standards define the critical attributes of a PRO measure for these research studies.

This chapter summarizes the recommendations for selecting PRO measures for research studies.

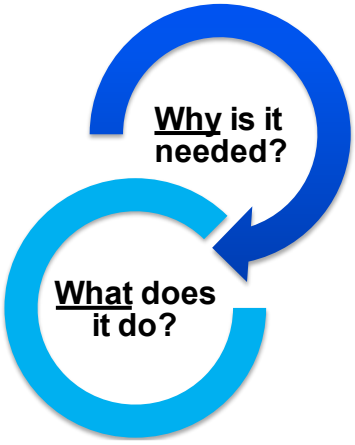
View ISOQOL Minimum Standards article

View the Checklist for the ISOQOL Measure Selection Standards

References

Acknowledgements

Why is This Resource Needed?



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

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

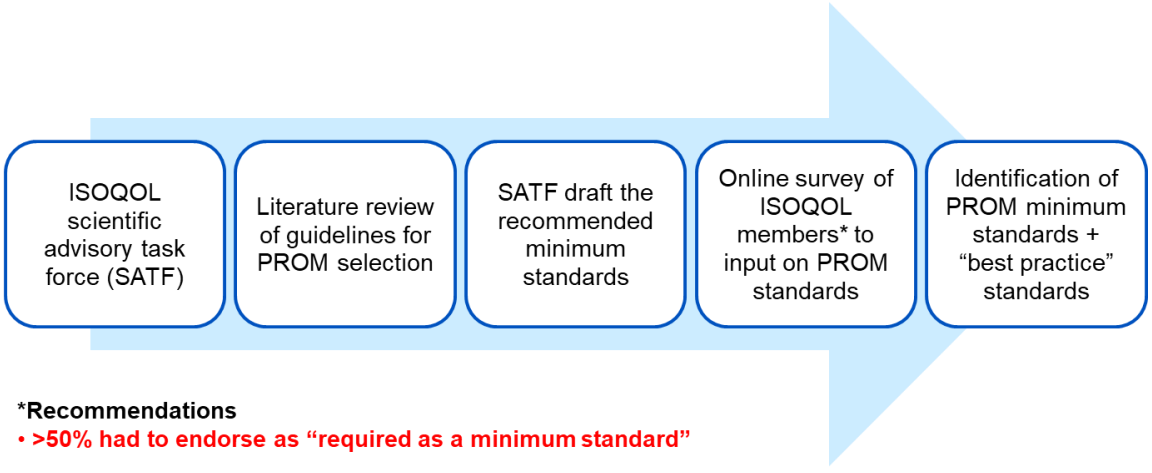
- An essential aspect of patient-centered outcomes research (PCOR) and comparative effectiveness research (CER) is integration of patient perspectives and experiences about their health with clinical and biological data to evaluate the safety and effectiveness of interventions
- Clinical trials are one kind of PCOR/CER; the ISOQOL minimum standards address PCOR/CER more broadly, but we will refer to clinical trials in this handbook
- It is widely accepted that patients' reports are the best source of information about what they are experiencing
- A challenge for PCOR and CER is how to best capture patient-reported data to inform decision making in healthcare delivery, research, and policy settings
- To draw valid research conclusions regarding patient-centered outcomes, PROs should 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
- PROs can be symptoms (e.g., pain, anxiety, nausea, fatigue), aspects of functioning (e.g., role, physical, emotional, social) and multidimensional constructs (e.g., health-related quality of life)
- A PRO measure is the questionnaire, index, checklist, instrument, or tool, along with the algorithm used to score patient responses into summary scores for analysis and reporting

Objective of Resource

The objective of the ISOQOL PRO measure selection guidance was to develop minimum standards for the design and selection of a PRO measure for use in PCOR and CER. These standards represent the minimum criteria required for a PRO measure to be judged suitable for inclusion in a PCOR or CER study. These minimum 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

An ISOQOL Scientific Advisory Task Force (SATF) was established to guide the drafting and final determination of recommended minimum standards. Based on a literature review, the SATF developed draft recommendations, which were subsequently reviewed by ISOQOL members through a formal survey. The literature review and feedback from ISOQOL members informed the final recommendations.



Summary of Recommendations

The ISOQOL PRO measure minimum standards recommends that a PRO measure should include the following attributes:

- Conceptual and measurement model
- Evidence that supports the measure's ability to assess the concepts covered in the measurement model, such as:
  - Reliability
  - Validity
    - Content
    - Construct
    - Responsiveness
- Interpretability of scores
- Translation
- Patient and investigator burden

Conceptual and Measurement Model

The conceptual model provides a description of and framework for the targeted concept(s) to be included in a PRO measure. The measurement model maps the individual items in the PRO measure to the concept(s).

- A PRO measure should have documentation defining and describing the concept(s) included and the intended population(s) for use
- 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

Reliability

Reliability is the degree to which a PRO measure is free from measurement error.

There are two types of reliability relevant for PRO measures:

1. **Internal consistency (for multi-item scales)**

   Internal consistency reliability is the degree of the interrelatedness among the items in a multi-item PRO measure. The internal consistency reliability of a PRO measure should preferably be at or above **0.70** for group-level comparisons, but may be lower if appropriately justified.

2. **Test-retest**

   Test-retest reliability is a measure of the reproducibility of the scale, that is, the ability to provide consistent scores over time in a stable population. However, some populations studied in PCOR are not stable and their health-related quality of life can fluctuate. This phenomenon would reduce estimates of test–retest reliability, making the PRO measure look unreliable when it may be accurately detecting changes over time.

Validity

Validity is the extent to which a PRO 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**

   Content validity is the extent to which the PRO measure includes the most relevant and important aspects of a concept in the context of a given measurement application.

A PRO measure should have evidence supporting its content validity, including evidence that patients and experts considered the content of the PRO measure relevant and comprehensive for the concept, population, and aim of the measurement application.

This includes documentation of:

a. qualitative and/or quantitative methods used to solicit and confirm the attributes (i.e., concepts measured by the items) of the PRO measure relevant to the measurement application
b. the characteristics of the 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
c. justification for the recall period for the measurement application

2. Construct Validity

Construct validity is the degree 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 *a priori* hypotheses concerning the concepts that are being measured.

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 concepts measured by the PRO measure.

Types of construct validity:

a. Structural Validity
   - extent to which the empirical data support the conceptual model
b. Convergent Validity
   - extent to which the PRO measure is similar to other established measures assessing the same concept
c. Discriminant Validity
   - extent to which the PRO measure is dissimilar to other established measures measuring different concepts
d. Known Groups Validity
   - extent to which the PRO measure can differentiate between groups known to differ on the measured concept

3. **Responsiveness**

   Responsiveness is the extent to which a PRO measure can detect changes in the construct being measured over time. A PRO measure for use in longitudinal research studies 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.

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(s). Knowing what comprises a meaningful difference or change in the score from one group to another (or one time to another) improves understanding of the outcome being measured. Another way to enhance the interpretability of PRO measure scores involves comparing scores from a study to known scores in a population (e.g., the general US population or a specific disease population). The availability of such benchmarks improves understanding of how the study group scored as compared to some reference or normative group.

Translation of the PRO Measure

PCOR and CER are often carried out in multi-national or multi-cultural settings that require the PRO measure to be translated into different languages. To be able to compare or combine PRO results across those groups, it is critical that the measured concepts and PRO measure wording is interpreted in the same way across 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. Established international guidance for the linguistic and cross-cultural adaptation of PRO measures should be followed. It is important that not only the words, but also the concepts, are applicable and interpretable across cultural settings. Studies should at least include evidence from forward and backward translations and qualitative methods (e.g., cognitive testing) with the target population to evaluate the translations.

Patient and Investigator Burden

A PRO measure must not be overly burdensome for patients or investigators. The length of the PRO measure should be considered in the context of other PRO measures included in the assessment. How often the PRO measure is administered in the clinical research study should also be considered. Lastly, the literacy demand of the items in the PRO measure should be at a 6th grade education level or lower (i.e., 12 year old or lower) to be acceptable; however, it should be appropriately justified for the context of the proposed application.

Checklist for the ISOQOL Measure Selection Standards

Minimum Standard	Explanation	Notes/comments
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. 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.	
2. Reliability	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. Reliability can be estimated using a variety of methods including internal consistency reliability, test–retest reliability, or item response theory. Each method should be justified.	
3. Validity <u>3a Content validity</u> <u>3b Construct validity</u>	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 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.	

Minimum Standard	Explanation	Notes/comments
3c Responsiveness	A PRO measure for use in longitudinal research studies 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.	
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.	
5. 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.	
6. Patient and investigator burden	PRO measures must not be overly burdensome for patients or investigators. 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 (i.e., 12-year-old or lower); however, it should be appropriately justified for the context of the proposed application.	

References

Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki D, Moinpour CM, McLeod LD, Lyons JC, Lenderking WR, Hinds PS, Hays RD, Greenhalgh J, Gerson R, Feeny D, Fayers PM, Cella D, Brundage M, Ahmed S, Aaronson NK, Butt Z. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22:1889-1905.

Further Reading

Crossnohere NL, Brundage M, Calvert MJ, et al. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. *Qual Life Res*. 2021;30(1):21-40. doi:10.1007/s11136-020-02625-z

Back to Table of Contents

Please Note: When referencing information included in this Chapter, we recommend citing the primary sources rather than this Handbook.