

## HTA

### HTA and Evaluation Methods

**Table 2: Important properties of PROMs**

Property	Description	Explanatory notes
<b>RELIABILITY</b>	Measurements are repeatable and consistent, and must distinguish between changes in response and changes due to errors in administration.  Is a necessary (but not comprehensive) component of validity.	1) <b>Test-retest reliability</b> - a measure of the ability of e.g., a psychologic testing instrument to yield the same result for a single Patient at 2 different closely spaced test periods, so that any variation detected reflects reliability of the instrument rather than changes in the Patient's status. May be considered as intra-interviewer reliability for interviewer administered PROs. 2) <b>Internal consistency</b> - consistency of the results delivered in a test, ensuring that the various items measuring the different constructs deliver consistent scores. 3) <b>Inter-interviewer reliability</b> – determines the changes in the results when the instrument is administered by two or more interviewers.
<b>VALIDITY</b>		
Criterion validity	The extent to which the scores of a PRO instrument are related to a known gold standard measure of the same concept (correlate with another measure considered more accurate).	Practically, it is difficult to determine the criterion validity as there is no gold standard for most PROs. Two types: “predictive” and “concurrent”. In the former Check: the ability to predict something in the real world which is present theoretically, for instance, if slowness of movement in Parkinson's disease is measured, one could provide the respective patients with the result and determine if there is a high correlation between the scores and the degree of slowness. High correlation is evidence for “predictive validity”.  Check: in concurrent validity the ability of the item to distinguish between the groups in real world as theoretically expected. E.g. in the assessment of breathlessness, the item should be able to differentiate between asthma and COPD
Content validity “Face validity” (appears to measure concept of interest) “Content validity in context of use” (adequately covers concept/domain of interest)	The extent to which an instrument measures what it is intended to measure (the concept of interest).	A lack of content validity means that data gathered may appear useful but is in fact measuring a different effect. (off target) Check e.g. by: <ul style="list-style-type: none"> <li>- Literature review</li> <li>- Expert opinion</li> <li>- evaluation of how well items deal with the complete continuum of patient experiences</li> <li>- patient and clinician evaluation of the relevance and comprehensiveness of the content contained in the measures through qualitative research with the targeted patient population (essential)</li> <li>- Input from target population of patients to document understandability and comprehensiveness of measure</li> <li>- Diversity in demographic &amp; disease characteristics of target population</li> </ul>

Construct validity	Logical relationships should exist between assessment measures	<p>A lack of construct validity means that data gathered may appear useful but is in fact measuring a different effect.</p> <p>Check: Evidence that relationships among concepts, (domains) and items conform to a priori hypotheses concerning logical relationships that should exist with other measures or characteristics of patients and patient groups</p> <p>Example: In COPD patients: expect that patients with lower treadmill exercise capacity generally will have more dyspnoea (shortness of breath) in daily life than those with higher exercise capacity, and – <i>logical connection</i> - expect substantial correlations between a new measure of emotional function and existing emotional function questionnaires.</p>
<b>RESPONSIVENESS</b> Also: <b>sensitivity to change or ability to detect change</b>	Captures changes over time in the construct being measured – before and after an intervention OR– in different disease or treatment states	<p>If the instrument is not adequately responsive, it may fail to measure patient response in which the intervention improves how patients feel and return false negative results.</p> <p>Check: Evaluated within specific populations and not a fixed/inherent property of the item Determine the relevant, clinically meaningful effect size</p>
<b>PRACTICALITY and FEASIBILITY</b>	Measurements are easily obtained, and the instrument is easy to operate.	<p>If the instrument is difficult to operate, patient response may change but not be measured or be measured inappropriately.</p> <p>If the instrument requires inordinate time the compliance is jeopardised.</p> <p>Check: - Verify the acceptability of the PROM by patients Review the instructions, questions, response options and recall period of the PROM</p>
<b>INTERPRETABILITY</b>	The degree to which one can assign qualitative meaning (clinical or commonly understood connotations) to a PROM's quantitative scores or change in scores <sup>1</sup>	<p>Patients or clinicians or <b>payers</b> may not make best decisions if output of tool is difficult to understand. PROMs vary in scale, measurement, and interpretation, making it difficult for patients and physicians alike to reach conclusions when evaluating results from these measures, e.g., in clinical trials. Metrics can indicate a favorable result with higher values, whereas some indicate a favorable result with lower values. This can lead to a lack of patient or clinician acceptance and may limit its utility.</p> <p>Check: The meaningfulness of scores produced by the questionnaire –What does a score mean? –What is the minimal clinically important difference (MID)? (see box below) –Should an overall score be computed and presented to support a given claim?</p>



<sup>1</sup> Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745.

