

The COSMIN checklist



Contact

CB Terwee, PhD
 VU University Medical Center
 Department of Epidemiology and Biostatistics
 EMGO Institute for Health and Care Research
 1081 BT Amsterdam
 The Netherlands
 Website: www.cosmin.nl, www.emgo.nl
 E-mail: cb.terwee@vumc.nl

Step 1. Evaluated measurement properties in the article

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I
	Interpretability	Box J

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models		yes	no	?
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	<input type="checkbox"/>	<input type="checkbox"/>	
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency			
	yes	no	?
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Design requirements</i>			
	yes	no	?
2 Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
3 Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
4 Was the sample size included in the internal consistency analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	<input type="checkbox"/>	<input type="checkbox"/>	
6 Was the sample size included in the unidimensionality analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8 Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>			
	yes	no	NA
9 for Classical Test Theory (CTT): Was Cronbach's alpha calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10 for dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11 for IRT: Was a goodness of fit statistic at a global level calculated? e.g. χ^2 , reliability coefficient of estimated latent trait value (index of (subject or item) separation)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)			
	yes	no	?
<i>Design requirements</i>			
	yes	no	?
1 Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2 Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3 Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 Were at least two measurements available?	<input type="checkbox"/>	<input type="checkbox"/>	
5 Were the administrations independent?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
7 Were patients stable in the interim period on the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8	Was the time interval appropriate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA ?
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	for ordinal scores: Was a weighted kappa calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box C. Measurement error: absolute measures				
<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Were at least two measurements available?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Were the administrations independent?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Were patients stable in the interim period on the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Was the time interval appropriate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	?
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	<input type="checkbox"/>	<input type="checkbox"/>	

Box D. Content validity (including face validity)

<i>General requirements</i>		yes	no	?
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	

Box E. Structural validity

1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Design requirements</i>		yes	no	?
2	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
4	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
6	for CTT: Was exploratory or confirmatory factor analysis performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box F. Hypotheses testing

<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		yes	no	NA
5	Was the expected <i>direction</i> of correlations or mean differences included in the hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Was the expected absolute or relative <i>magnitude</i> of correlations or mean differences included in the hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	for convergent validity: Was an adequate description provided of the comparator instrument(s)?	<input type="checkbox"/>	<input type="checkbox"/>	
8	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA
10	Were design and statistical methods adequate for the hypotheses to be tested?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box G. Cross-cultural validity

		yes	no	?
	<i>Design requirements</i>			
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	<input type="checkbox"/>	<input type="checkbox"/>	
6	Did the translators work independently from each other?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	Were items translated forward and backward?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Was there an adequate description of how differences between the original and translated versions were resolved?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Was the translation reviewed by a committee (e.g. original developers)?	<input type="checkbox"/>	<input type="checkbox"/>	
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	<input type="checkbox"/>	<input type="checkbox"/>	

11	Was the sample used in the pre-test adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
12	Were the samples similar for all characteristics except language and/or cultural background?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
14	for CTT: Was confirmatory factor analysis performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	for IRT: Was differential item function (DIF) between language groups assessed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box H. Criterion validity				
<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Can the criterion used or employed be considered as a reasonable 'gold standard'?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	for dichotomous scores: Were sensitivity and specificity determined?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box I. Responsiveness				
<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was a longitudinal design with at least two measurement used?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	

7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Design requirements for hypotheses testing</i>	yes	no	?
	For constructs for which a gold standard was not available:			
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		yes	no	NA
9	Was the expected <i>direction</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Was an adequate description provided of the comparator instrument(s)?	<input type="checkbox"/>	<input type="checkbox"/>	
12	Were the measurement properties of the comparator instrument(s) adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
13	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA
14	Were design and statistical methods adequate for the hypotheses to be tested?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<i>Design requirement for comparison to a gold standard</i>	yes	no	?
	For constructs for which a gold standard was available:			
15	Can the criterion for change be considered as a reasonable gold standard?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA
17	for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box J. Interpretability		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	

3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was the distribution of the (total) scores in the study sample described?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the percentage of the respondents who had the lowest possible (total) score described?	<input type="checkbox"/>	<input type="checkbox"/>	
6	Was the percentage of the respondents who had the highest possible (total) score described?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups? e.g. for normative groups, subgroups of patients, or the general population	<input type="checkbox"/>	<input type="checkbox"/>	
8	Was the minimal important change (MIC) or the minimal important difference (MID) determined?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	

Step 4: Determining the Generalisability of the results

Box Generalisability		yes	no	NA
	Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of:			
1	median or mean age (with standard deviation or range)?	<input type="checkbox"/>	<input type="checkbox"/>	
2	distribution of sex?	<input type="checkbox"/>	<input type="checkbox"/>	
3	important disease characteristics (e.g. severity, status, duration) and description of treatment?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	setting(s) in which the study was conducted? e.g. general population, primary care or hospital/rehabilitation care	<input type="checkbox"/>	<input type="checkbox"/>	
5	countries in which the study was conducted?	<input type="checkbox"/>	<input type="checkbox"/>	
6	language in which the HR-PRO instrument was evaluated?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Was the method used to select patients adequately described? e.g. convenience, consecutive, or random	<input type="checkbox"/>	<input type="checkbox"/>	
		yes	no	?
8	Was the percentage of missing responses (response rate) acceptable?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>