

# The COSMIN Risk of Bias

	Very good	Adequate	Doubtful	Inadequate	Not available	
	●	●	●	●	●	Box 10
	●	●	●	●	●	Box 9
	●	●	●	●	●	Box 8
	●	●	●	●	●	Box 7
	●	●	●	●	●	Box 6
	●	●	●	●	●	Box 5
	●	●	●	●	●	Box 4
	●	●	●	●	●	Box 3
	●	●	●	●	●	Box 2
	●	●	●	●	●	Box 1
Comrie et al. 2015	●	●	●	●	●	
Ekkekakis et al. 2005	●	●	●	●	●	
Hassan et al. 2013	●	●	●	●	●	
Hassan et al. 2018	●	●	●	●	●	
Hobbs et al. 2003	●	●	●	●	●	
Jennings et al. 2009	●	●	●	●	●	
Karim et al. 2006	●	●	●	●	●	
Mathew et al. 2019	●	●	●	●	●	
Martinez et al. 2002	●	●	●	●	●	
Michael et al. 2008	●	●	●	●	●	
Mosafa et al. 2019	●	●	●	●	●	

Henrik Hein Lauridsen  
 Associate Professor  
 Research unit for Clinical Biomechanics

## Quote

***“Clinical trials are only as good as the patient-reported outcome measures assessed”***

*Peter Tugwell, Journal of Clinical Epidemiology 2020;125: A7-A9*



# Eight step process

## What do you want to review?

- Step 1: Formulate the research question
- Step 2: Formulate the eligibility criteria
- Step 3: Develop the literature search

## Which studies have been done?

- Step 4: Conduct the literature search

## What PROMs are available?

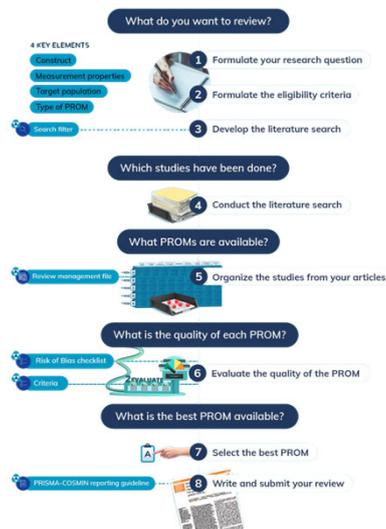
- Step 5: Extract the PROM data

## What is the quality of each PROM?

- Step 6: Evaluate the nine measurement properties per PROM

## What is the best PROM available?

- Step 7: Formulate recommendations
- Step 8: Report the systematic review



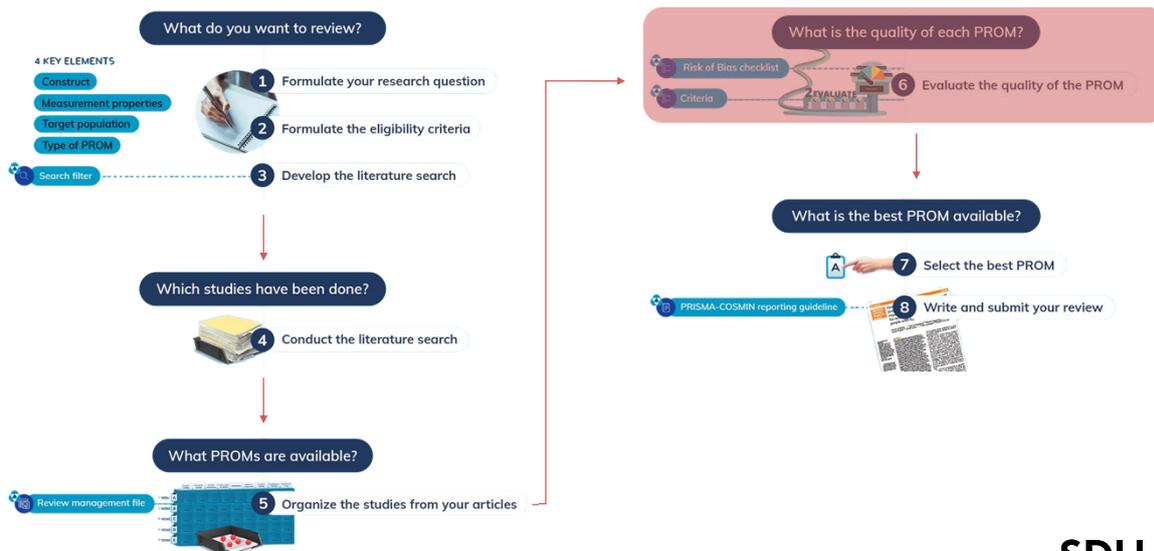
<https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/>  
<https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/>

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



3

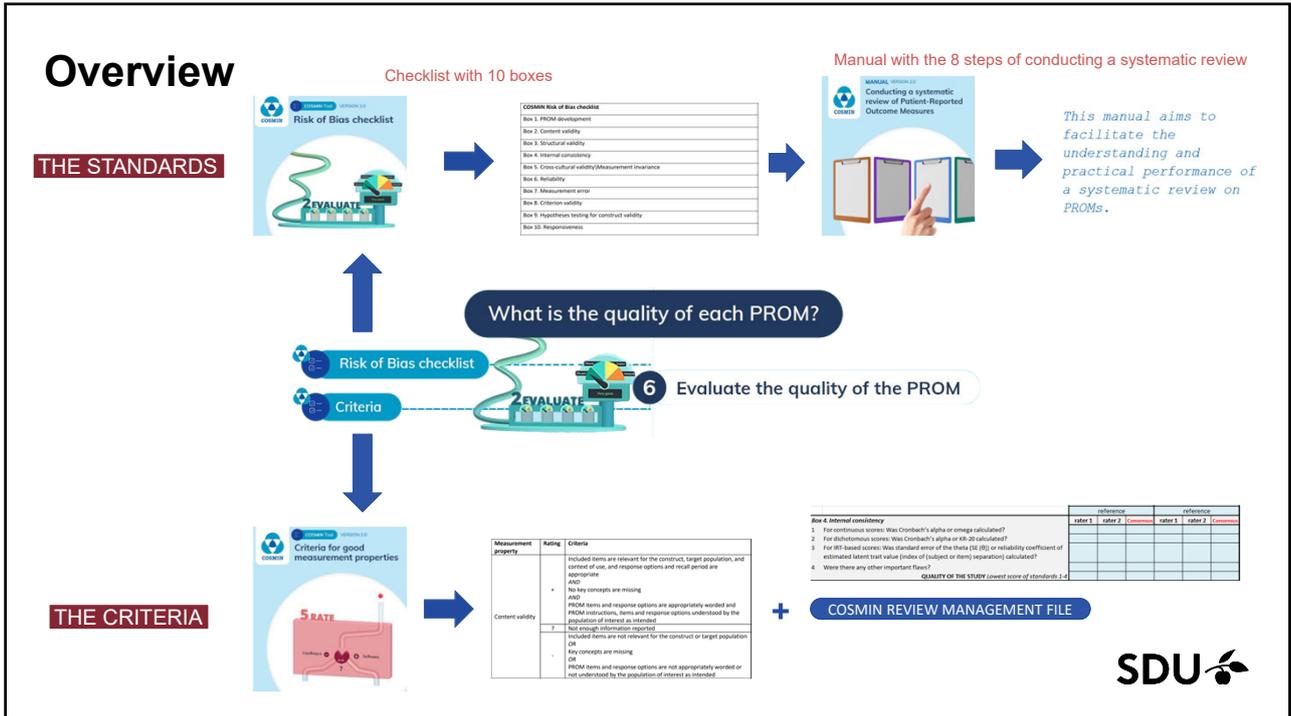
# Overview



DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



4



5

# Standards & Criteria

**THE STANDARDS**

- Refers to design requirements and preferred statistical methods for evaluating the **methodological quality** of studies on measurement properties

**THE CRITERIA**

- Refers to what constitutes good measurement properties – **quality of the PROM**

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS **SDU**

6

# Evaluate the quality of the PROM

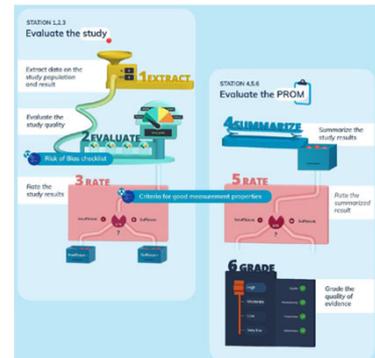
## The Marble Run

### Evaluate the study

- Station 1 – extract the data
- Station 2 – evaluate the study quality → Use *Risk of Bias checklist*
- Station 3 – rate the study results → Use *Criteria for good measurement properties*

### Evaluate the PROM

- Station 4 – summarize the result
  - Station 5 – rate the summarized results
  - Station 6 – grade the quality of evidence
- } Use *Review management file*



DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



7

## Station 2: Evaluate the quality of the study



- Each measurement property has a COSMIN box containing all standards
- 10 boxes for good methodological quality of development and measurement properties:

1. PROM development
2. Content validity
3. Structural validity
4. Internal consistency
5. Cross-cultural validity/Measurement invariance
6. Reproducibility (reliability)
7. Measurement error
8. Criterion validity
9. Hypotheses testing for construct validity
10. Responsiveness

} 10 measurement properties



DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



8

## Station 2: Evaluate the quality of the study

- **Study quality is rated using a four-point system**

- **Very good** → standard met or optimal method
- **Adequate** → assumably standard met or suboptimal method
- **Doubtful** → unclear if standard met or preferred method used
- **Inadequate** → standard not met or inadequate method

E.g. if for a reliability study one item in a box is rated as 'inadequate', the overall methodological quality of that reliability study is rated as 'inadequate'

"Worst score count" principle

- Overall rating for a *standard* is determined by taking the **lowest** rating of any standard for each measurement property

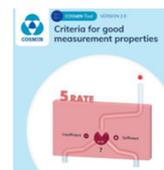
DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



9

## Station 3: Rate the study results

- Are the study results sufficient or not?
- Each result is rated:
  - Sufficient (+) → measurement property is of good quality
  - Insufficient (-) → measurement property is of poor quality
  - Indeterminate (?) → something was done, not enough information



DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



10

## Example: Internal consistency

MEASUREMENT: INTERDISCIPLINARY RESEARCH AND PERSPECTIVES  
<https://doi.org/10.1080/15366367.2024.2392059>

 Routledge  
Taylor & Francis Group

 OPEN ACCESS  Check for updates

### Measurement Properties of the Patient Assessment of Transitions in Healthcare Settings (PATH) Questionnaire

Sisse Walløe <sup>a,b</sup>, Lars Morse <sup>a,c</sup>, Esben Petersen <sup>d,e</sup>, Søren T. Skou <sup>b,f</sup>,  
Charlotte Simony <sup>b,g</sup>, Søren Möller <sup>g</sup>, and Henrik Lauridsen <sup>c</sup>

You have found a validation study, where internal consistency was reported to be 0.79 (Cronbach's Alpha). The authors conclude that the PROM is reliable.

You are unsure if you can trust their findings. What do you do?

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



11

## The COSMIN Risk of Bias checklist standards



	COSMIN Risk of Bias Checklist
	Box 1. PROM development
	Box 2. Content validity
	Box 3. Structural validity
✗	Box 4. Internal consistency
	Box 5. Cross-cultural validity/Measurement invariance
	Box 6. Reproducibility (reliability)
	Box 7. Measurement error
	Box 8. Criterion validity
	Box 9. Hypotheses testing for construct validity
	Box 10. Responsiveness

*Domain:* Reproducibility (reliability)

*Measurement property:* Internal consistency

*Definition:*

**"The degree of interrelatedness among the items"**

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



12

# Background PROM information



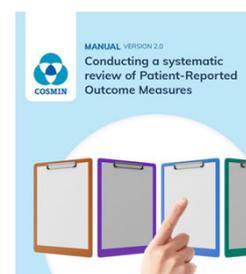
Scope of the PROM		
1	Is a clear description provided of the construct to be measured?	<input type="radio"/> Construct clearly described <input type="radio"/> Construct not clearly described
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?	<input type="radio"/> Origin of the construct clear <input type="radio"/> Origin of the construct not clear
3	Is a clear description provided of the target population for which the PROM was developed?	<input type="radio"/> Target population clearly described <input type="radio"/> Target population not clearly described
4	Is a clear description provided of the context of use	<input type="radio"/> Context of use clearly described <input type="radio"/> Context of use not clearly described
5	What is the measurement model on which the PROM is based?	<input type="radio"/> Reflective model <input type="radio"/> Formative model <sup>1</sup> <input type="radio"/> unclear
	Does the study concern unidimensionality or structural validity? <sup>2</sup>	<input type="radio"/> unidimensionality <input type="radio"/> structural validity

Information used in the "COSMIN review management file"

# Internal consistency – need to know

## Need to know

- Structural validity should be performed before internal consistency is determined
- Applied to unidimensional scale
- Cannot be higher than the quality of the structural validity
- Internal consistency applied to multidimensional scales should be ignored



### MEASUREMENT PROPERTY 3

#### 6.3 – Internal consistency

Internal consistency refers to the degree of interrelatedness among the items [124] and is often expressed in a Cronbach's alpha (CT), Omega [14], or an IRT- or Rasch-based reliability parameter based on single scores, such as the standard error (SE) of theta. For an appropriate interpretation of the internal consistency parameter, the items together should form a unidimensional scale. Internal consistency and unidimensionality are not the same. Internal consistency refers to the relatedness among the items, it tells you whether you have enough items in the scale to precisely measure the construct of interest. Unidimensionality refers to whether the items in a scale measure a single construct. Unidimensionality is a prerequisite for a clear interpretation of the internal consistency parameter (CT, Omega), and can be investigated by factor analysis or IRT/Rasch methods (see structural validity). Therefore, the evidence for structural validity should be taken into account when rating the evidence for internal consistency.

If you are evaluating the internal consistency of a PROM, you should first check the assumptions: is the PROM based on a reflective model (see step 3), and is there evidence for sufficient structural validity of the scale. In step 5 you have decided whether the PROM is based on a formative or reflective model. If the PROM is based on a formative model, you can ignore the studies on internal consistency, as this requires the PROM to be based on a reflective model. For an appropriate interpretation of the internal consistency results, the items together should form a unidimensional scale. If you do not have evidence for sufficient structural validity, you can rate the results for internal consistency of the PROM as 'indeterminate' (see also step 6.3.3).

## Internal consistency standards (box 4)



Box 4. Internal consistency					
Statistical methods	very good	adequate	doubtful	inadequate	NA
1 For continuous scores: Was Cronbach's alpha or omega calculated?	Cronbach's alpha, or Omega calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated	NA
2 For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated	NA
3 For IRT-based scores: Was standard error of the theta (SE(θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	SE(θ) or reliability coefficient calculated			SE(θ) or reliability coefficient NOT calculated	NA
<b>Other</b>					
4 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



15

## Quality of the PROM (the criteria)



Internal consistency	+	At least low evidence for sufficient unidimensionality <b>AND</b> Cronbach's alpha $\geq 0.70$
	?	Criteria for "at least low evidence for sufficient unidimensionality" not met <b>OR</b> Evidence for insufficient unidimensionality <b>OR</b> Not enough information reported
	-	At least low quality evidence for sufficient unidimensionality <b>AND</b> Cronbach's alpha $< 0.70$

+	= sufficient
?	= indeterminate
-	= insufficient

Alpha was 0.87 in the validation study

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



16

# Grading the quality of the evidence



## GRADE approach

### Summary of study quality (standards) AND PROM quality (criteria)

Quality level	Definition
High	We are <b>very confident</b> that the true measurement property lies close to that of the estimate* of the measurement property
Moderate	We are <b>moderately confident</b> in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different
Low	Our <b>confidence in the measurement property estimate is limited</b> ; the true measurement property may be substantially different from the estimate of the measurement property
Very low	We have <b>very little confidence in the measurement property estimate</b> ; the true measurement property is likely to be substantially different from the estimate of the measurement property

\* Estimate of the measurement property refers to the pooled or summarized result of the measurement property of a PROM. These definitions were adapted from the GRADE approach

# Applying GRADE



- The GRADE approach evaluates 4 areas

- Risk of Bias** - evaluates the extent to which the study design and execution could introduce bias
- Inconsistency** – inconsistency in the study results
- Imprecision** – total sample included in the studies
- Indirectness** – applicability of the results to the population and context of interest in your review



Risk of bias	Downgrading for Risk of Bias
<b>No</b>	There are multiple studies of at least adequate quality, or there is one study of very good quality available
<b>Serious</b>	There are multiple studies of doubtful quality available, or there is only one study of adequate quality
<b>Very serious</b>	There are multiple studies of inadequate quality, or there is only one study of doubtful quality available
<b>Extremely serious</b>	There is only one study of inadequate quality available

# Applying GRADE

E.g.: if criteria = + but RoB is 'very serious', the quality of the evidence is lowered by 2, i.e. from high to low

Risk of bias	Downgrading for Risk of Bias
No	There are multiple studies of at least adequate quality, or there is one study of very good quality available
Serious	There are multiple studies of doubtful quality available, or there is only one study of adequate quality
Very serious	There are multiple studies of inadequate quality, or there is only one study of doubtful quality available
Extremely serious	There is only one study of inadequate quality available



Quality of evidence	Lower if
High	Risk of bias -1 Serious -2 Very serious -3 Extremely serious
Moderate	
Low	Inconsistency -1 Serious -2 Very serious
Very low	Imprecision -1 Serious -2 Very serious  Indirectness -1 Serious -2 Very serious

# RoB – PROM versions and subscales



## COSMIN considers each version of a PROM separately

- The measurement properties should be rated separately for each version of a PROM
  - Version number
  - Language version
  - Paper vs. electronic
  - CAT forms

## COSMIN considers each subscale of a (multi-dimensional) PROM separately

- Each sum-score is assumed to represent a construct and is therefore considered a separate PROM
- If the subscale scores are also summarised into a total score, the measurement properties of the total score should also be rated separately

## Applications

1. Systematic review of measurement properties
2. Evaluation of single studies
3. Questionnaire selection
4. Identification of the need for further research on measurement properties
5. Designing a study on measurement properties
6. Reporting a study on measurement properties
7. Reviewing the quality of submitted manuscripts on measurement properties

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



21

## The COSMIN checklists and manuals

4 checklists:

NB: Outdated COSMIN checklist:  
• Mokkink et al. (2010)



1. **COSMIN Risk of Bias checklist for PROMs v. 3.0<sup>1</sup>**
  - A checklist to assess the **methodological quality of single studies included in systematic reviews of PROMs**
  - Contains standards referring to **design requirements and preferred statistical methods** of studies on measurement properties
2. **COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments<sup>2</sup>**
  - Standards to assess the quality of studies on **reliability and measurement error of any type of measurement instrument, which are typically more complex as PROMs, such as equipment**
3. **COSMIN Study Design checklist**
  - Recommended for **designing studies to evaluate measurement properties of existing PROMs**
4. **COSMIN Reporting Guideline for studies on measurement properties v. 2.0<sup>3</sup>**
  - Guidelines for the **reporting of studies on measurement properties of PROMs**

1. Mokkink LB et al. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res. 2024 Nov;33(11):2929–39.  
2. Mokkink LB et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Medical Research Methodology. 2020  
3. Gagnier JJ et al. COSMIN reporting guideline for studies on measurement properties of patient reported outcome measures: version 2.0. Qual Life Res. 2025 Mar 28;1–11.



22

**www.cosmin.nl**

Many important resources

Finding the right tool:

- I'm deciding what to measure
- I'm looking for available outcome measurement instruments
- I want to select the most suitable outcome measurement instrument
- I'm conducting a study on measurement properties
- I'm conducting a systematic review of outcome measurement instruments
- I'm developing a Core Outcome Set

COSMIN tools:

- 📄 COSMIN Taxonomy of Measurement Properties
- 🔍 COSMIN Database of Systematic Reviews
- 📄 Checklists for Assessing Study Qualities
- 🔍 Search Filters for Relevant Studies
- 📄 Guideline for Conducting Systematic Reviews
- 📄 Guideline for selecting instruments for a Core Outcome Set
- 📄 Book Measurement in Medicine

### COSMIN

A center of expertise for outcome measurement instruments in health research and clinical practice.

### JOIN OUR COSMIN CLUB

We hold regular online COSMIN Club meetings to discuss your ongoing studies or reviews. Sign up and get in touch.

[JOIN ONE OF THE COSMIN CLUBS >](#)

### COSMIN courses

Clinimetrics: Assessing measurement properties of health measurement instruments

Clinimetrics is the scientific discipline that aimed to develop and evaluate clinical and health measurement instruments and improve its quality.

📅 online course on Clinimetrics 6 half days June 11-18, 2025 (in English),

📅 3-day course on Clinimetrics February 10-12, 2025 (in Dutch),

📅 3-day online course on Item Response Theory March 4-13, 2025 (in English).

DEPARTMENT OF SPORTS SCIENCE AND CLINICAL BIOMECHANICS



23



24