VUmc VU University Medical Center Amsterdam

Developing a Core Outcome Set for (stage IV) melanoma trials

C.A.C. (Sanna) Prinsen, MSc, PhD
Clinical epidemiologist
VU University Medical Center
Amsterdam, The Netherlands

Annual Cochrane Skin Group Meeting
Dresden, Germany
17-Mar-2015

What is a 'Core Outcome Set'?

'Collection Of Style'
Berlin, October 2014

A 'core outcome set' (COS) is a recommendation of ***what*** should be measured and reported in all clinical trials

Once COS are defined, it is then important to achieve consensus on ***how*** these outcomes should be measured

What = outcome
How = outcome measurement instrument

What is the problem?

VUmc VU University Medical Center Amsterdam

There is lack of **standardization in outcome reporting** in melanoma clinical trials

This hampers the **usefulness of clinical trial evidence** to inform clinicians

At the cost of the best possible **care** for melanoma patients

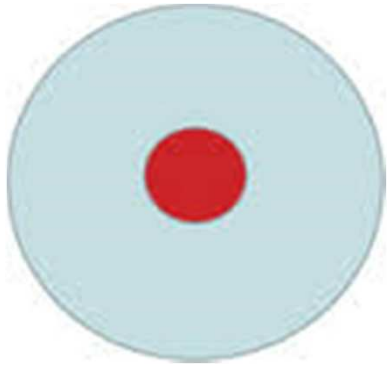
What is the solution?

Consensus:

International consensus among relevant stakeholders on the 'core outcomes', including recommendations on outcome measurement instruments that can be used to measure these core outcomes

Recommendation:

The core outcomes should be measured and reported as a *minimum* in all future melanoma clinical trials



Schmitt *et al*. (2015)

Objective

Overall aim:

To develop a multi-disciplinary, consensus-based set of core outcomes, including their relevant outcome measurement instruments, for melanoma clinical trials

Steering Committee

VUmc VU University Medical Center Amsterdam

Name	Background	Country
Sanna Prinsen	Clinical epidemiologist, postdoc COSMIN and COMET, member HOME research groups	The Netherlands
Caroline Terwee	Epidemiologist, co-founder of COSMIN	The Netherlands
Phyllis Spuls	Dermatologist, Executive Committee HOME	The Netherlands
Jochen Schmitt	Dermatologist, Executive Committee HOME, co-founder CSG-COUSIN	Germany
Stefanie Deckert	Scientific researcher, member HOME research group, member VAPAIN	Germany
Maarten Boers	Rheumatologist, clinical epidemiologist, co-founder OMERACT	The Netherlands
Marcel Bekkenk	Dermatologist, expertise in melanoma	The Netherlands
Robert Stern	Dermatologist	USA
Alexander van Akkooi	Surgical oncologist, expertise in melanoma	The Netherlands
Astrid Nollen	Patient research partner, chair Dutch Melanoma Foundation	The Netherlands
Cynthia Chauhan	Patient research partner	USA
Oncologist (vacancy)		
PhD student (vacancy)		

How are core outcomes agreed upon?

VU University Medical Center Amsterdam

The HOME roadmap to develop core sets of outcome measurements

Step 1: Define scope and applicability
Population (condition)
Intervention
Setting (e.g. trial, registry)
Geographical / regional scope
Stakeholders

↓

Step 2: Develop Core Set of Outcome Domains
Consensus study involving representatives of relevant stakeholders.

← *What* to measure

↓

Step 3: Develop Core Set of Outcome Measurements
Identification and recommendation of adequate measurement instrument(s) for each core outcome domain by a 5 stage process

← *How* to measure

	Stage 1 →	Stage 2 →	Stage 3 →			Stage 4 →	Stage 5
Task	**Identify all instruments** previously used to measure the domain.	Establish the **extent and quality of testing** of the identified instruments.	Determine **which instruments** are good enough quality **meet the requirements of the OMERACT filter** and be shortlisted for further consideration.			Carry out **validation studies** on shortlisted scales.	**Finalise core outcome(s)** for domain.
Methodology	Systematic review of outcome instruments used.	Systematic review of validation studies of the long-list of identified instruments. Highlight any gaps in validation.	**Apply OMERACT filter:** Truth, discrimination and feasibility:			Consensus discussion and voting to determine what validation studies will be conducted on short-listed instruments. Gaps in testing were highlighted in stage 2 (systematic review). Appropriate methods used to fill the gaps in validation.	Re-apply the OMERACT filter with the results of the completed validation studies. Consensus discussion and voting on core outcome to be recommended.
			Truth *"Is the measure truthful, does it measure what it intends to measure? Is the result unbiased and relevant?"*	**Discrimination** *"Does the measure discriminate between situations that are of interest?"*	**Feasibility** *"Can the measure be applied easily in it's intended setting given constraints of time, money, and interpretability?"*		
			Consensus discussion and voting on truth: 1. Face validity 2. Content validity 3. Construct validity 4. Criterion validity	Consensus discussion and voting on discrimination: 1. Reliability 2. Sensitivity to change	Consensus discussion and voting on feasibility: 1. Time taken 2. Cost 3. Interpretability		
Output	**Long-list of all instruments previously used to measure the domain.**	**Summary of which instruments have been tested and the quality, extent and results of any testing.**	**Short-list of potential instruments that meet the requirements of the OMERACT filter.**			**Short-list of fully tested instruments.**	**Recommended core outcome instrument for the domain.**

↓

Step 4: Disseminate, review, and possibly revise Core Set of Outcome Measurements

Schmitt *et al*. (2015)

1. Define scope and applicability

VUmc VU University Medical Center Amsterdam

Scope and applicability:

- ◇ Population (melanoma)
- ◇ Setting (clinical trials)
- ◇ Geographical scope (global)
- ◇ Stakeholders* (all relevant)

1. Define scope and applicability

VUmc VU University Medical Center Amsterdam

Stakeholders:

- ◇ Patient representatives
- ◇ Researchers (incl methodologists)
- ◇ Healthcare providers (incl dermatologists, oncologists, surgeons, nurses)
- ◇ Policy makers (incl regulators, payers)
- ◇ Representatives from pharmaceutical industry
- ◇ Representatives from drug regulatory authorities
- ◇ Journal editors

2. Define core set of outcomes

VU University Medical Center Amsterdam

Consensus-based method

Delphi study:
A structured, iterative process to achieve consensus among a group of stakeholders about a given issue

The group of stakeholders does not need to meet which confers **anonymity**; opinions are to be expressed **free from group pressure**; and **possible dominance** of individuals in face-to-face group meetings is being avoided

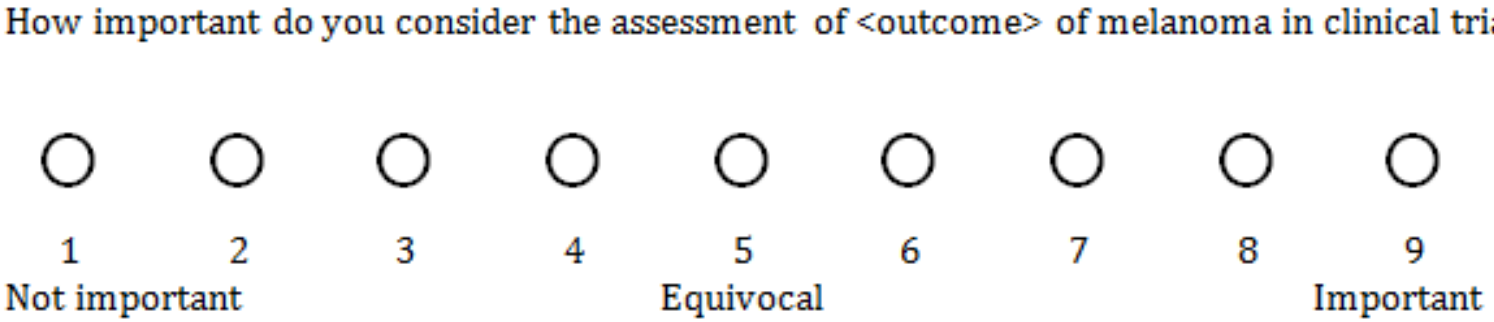
Delphic oracle's skills of interpretation and foresight

2. Define core set of outcomes

VUmc VU University Medical Center Amsterdam

- **Literature review:** to identify all outcomes that have been measured and reported in melanoma clinical trials*

- **Questionnaire survey:** to reach consensus on core outcomes



  - Consensus: ≥70% scoring 7-9 and ≤15% scoring 1-3
  - Approx. 3 rounds

- **Group discussions and voting:** to achieve consensus on final core set of outcomes

3. Define core set of OMIs

VUmc VU University Medical Center Amsterdam

To identify, validate, or develop an appropriate outcome measurement instrument (OMI) **for each core outcome**

5 stages*

Guideline on instrument selection: recommended to include **only one** outcome measurement instrument for each outcome in the COS

Prinsen *et al*. (manuscript in preparation)

3. Define core set of OMIs

VUmc VU University Medical Center Amsterdam

For each core outcome:

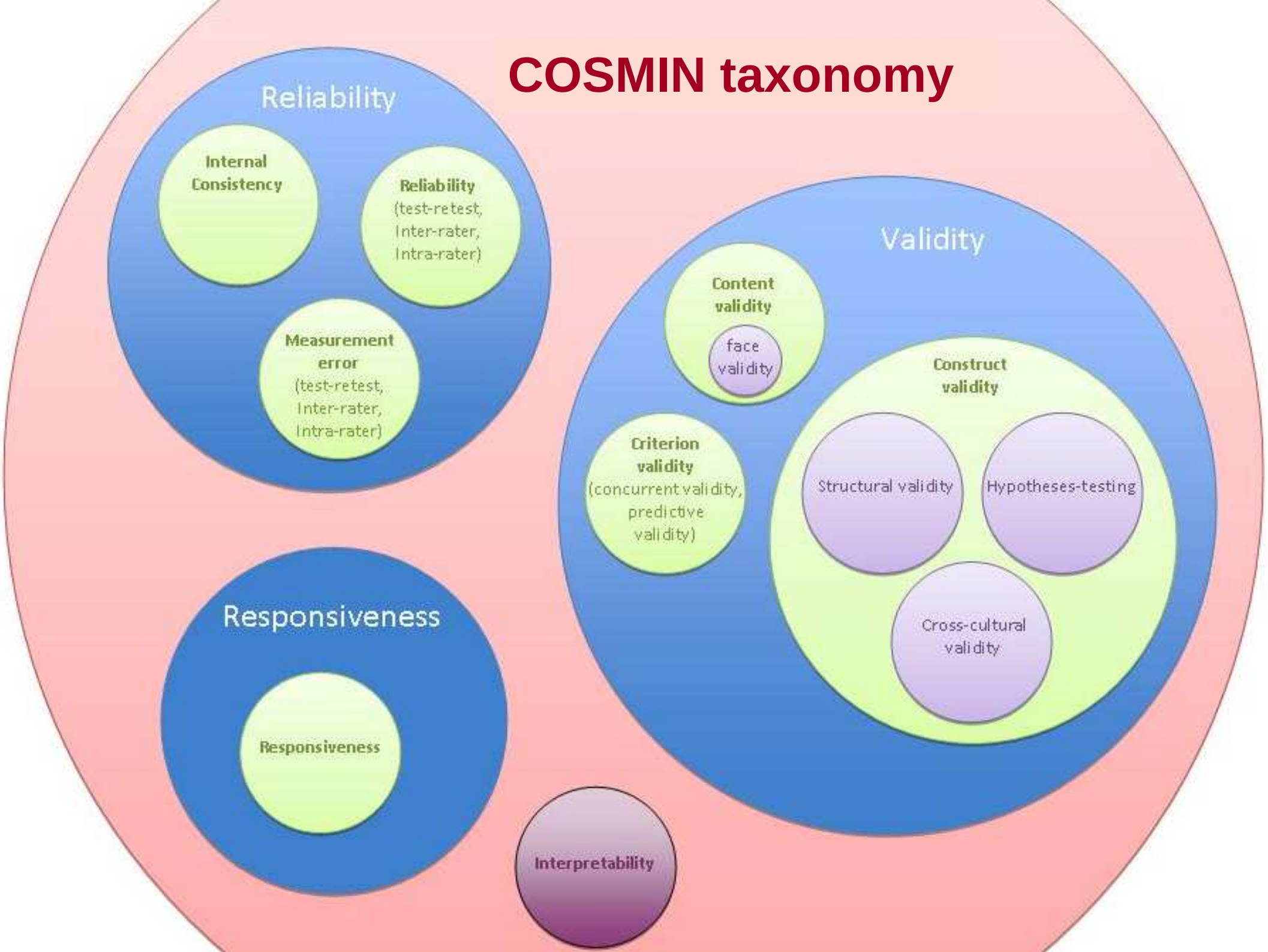
Stage 1: Systematic review to identify all OMIs used to assess the construct of interest → long list

Stage 2: Systematic review for each OMI to investigate the quality of the OMIs (COSMIN, qual criteria)

Stage 3: Determine whether OMIs are suitable for the assessment of the core outcome (reliability, validity, and feasibility) → short list

Stage 4: Additional validation studies may be needed

Stage 5: Delphi study to reach consensus on the core outcome measurement instrument → voting


COSMIN taxonomy
Reliability
Internal Consistency
Reliability (test-retest, Inter-rater, Intra-rater)
Measurement error (test-retest, Inter-rater, Intra-rater)
Validity
Content validity
face validity
Construct validity
Criterion validity (concurrent validity, predictive validity)
Structural validity
Hypotheses-testing
Cross-cultural validity
Responsiveness
Responsiveness
Interpretability

COSMIN definitions of domains, measurement properties, and aspects of measurement properties

Term			Definition
Domain	Measurement property	Aspect of a measurement property	
Reliability			The degree to which the measurement is free from measurement error
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health related-patient reported outcomes (HR-PRO) (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)
	Internal consistency		The degree of the interrelatedness among the items
	Reliability		The proportion of the total variance in the measurements which is due to 'true'† differences between patients
	Measurement error		The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Validity			The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
	Content validity		The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
		Face validity	The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of an HR-PRO instrument are consistent with hypotheses *(for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups)* based on the assumption that the HR-PRO instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
		Hypotheses testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
	Criterion validity		The degree to which the scores of an HR-PRO instrument are an adequate reflection of a 'gold standard'
Responsiveness			The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Responsiveness		Idem responsiveness
Interpretability*			Interpretability is the degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores.

† *The word 'true' must be seen in the context of the CTT, which states that any observation is composed of two components – a true score and error associated with the observation. 'True' is the average score that would be obtained if the scale were given an infinite number of times. It refers only to the consistency of the score, and not to its accuracy (ref Streiner & Norman)*
* *Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument*

Example

Domain	Measurement property	Definition
Reliability		The degree to which the measurment is free from measurement error
	Internal consistency (Box A)	The degree of the interrelatedness among the items

COSMIN checklist

VUmc VU University Medical Center Amsterdam

For each measurement property standards were developed for how this property should be evaluated

- Design requirements
- Requirements for the statistical methods

The standars refer to the **quality of a study** on measurement properties, not the quality of the instruments that are being evaluated.

COSMIN scoring system

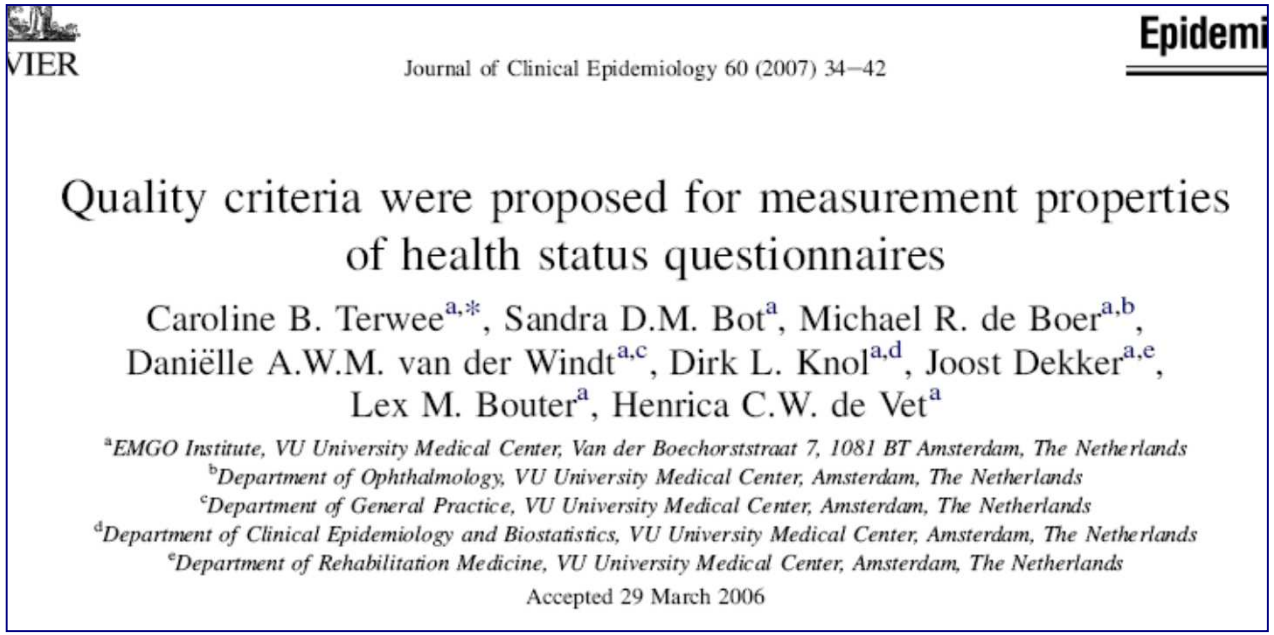
VUmc VU University Medical Center Amsterdam

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)				
	excellent	good	fair	poor
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

Quality of the OMI

VUmc VU University Medical Center Amsterdam

Use criteria for good measurement properties:



VIER

Journal of Clinical Epidemiology 60 (2007) 34–42

Epidemi

Quality criteria were proposed for measurement properties of health status questionnaires

Caroline B. Terwee[a,*], Sandra D.M. Bot[a], Michael R. de Boer[a,b], Daniëlle A.W.M. van der Windt[a,c], Dirk L. Knol[a,d], Joost Dekker[a,e], Lex M. Bouter[a], Henrica C.W. de Vet[a]

[a]*EMGO Institute, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands*
[b]*Department of Ophthalmology, VU University Medical Center, Amsterdam, The Netherlands*
[c]*Department of General Practice, VU University Medical Center, Amsterdam, The Netherlands*
[d]*Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands*
[e]*Department of Rehabilitation Medicine, VU University Medical Center, Amsterdam, The Netherlands*

Accepted 29 March 2006

Levels of Evidence

VUmc VU University Medical Center Amsterdam

Level	Rating	Criteria
strong	**+++ or ---**	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
moderate	**++ or --**	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
limited	**+ or -**	One study of fair methodological quality
conflicting	**+/-**	Conflicting findings
unknown	**?**	Only studies of poor methodological quality

3. Define core set of OMIs

VUmc VU University Medical Center Amsterdam

For each core outcome:

Stage 1: Systematic review to identify all OMIs used to assess the construct of interest → long list

Stage 2: Systematic review for each OMI to investigate the quality of the OMIs (COSMIN, qual criteria)

Stage 3: Determine whether OMIs are suitable for the assessment of the core outcome (reliability, validity, and feasibility) → short list

Stage 4: Additional validation studies may be needed

Stage 5: Delphi study to reach consensus on the core outcome measurement instrument → voting

4. Dissemination

VUmc VU University Medical Center Amsterdam

Dissemination and implementation:

◇ To enhance the use of the core set in clinical trials

◇ Involvement of all relevant stakeholders

◇ Publications in leading journals

◇ Presentations at relevant meetings

◇ Dissemination to journal editors and reviewers

◇ Dissemination to other stakeholders

→ Guidance materials

→ Monitoring to detect barriers

Potential importance of the results

VUmc VU University Medical Center Amsterdam

Standardization in outcome reporting will:

- ◇ Allow comparisons across clinical trials
- ◇ Improve the usefulness of clinical trial evidence to inform healthcare providers in decision making
- ◇ Limit outcome reporting bias

Idzerda *et al*. (2014)

If we will be awarded…

VUmc VU University Medical Center Amsterdam

Announcement: April/May 2015

Proposed project will be embedded within **Cochrane Skin Group Outcomes Research Initiative** (CSG-COUSIN)

Projected study time lines:

01-Oct-2015 to 30-Sep-2018

pcori
PATIENT-CENTERED OUTCOMES RESEARCH INSTITUTE

Guidelines

- **COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN): a guideline for systematic reviews of outcome measurement instruments**

  CAC Prinsen, LB Mokkink, HCW de Vet, CB Terwee

- **How to select outcome measurement instruments for outcomes included in a 'Core Outcome Set' – a practical guideline**

  CAC Prinsen, S Vohra, MR Rose, M Boers, P Tugwell, M Clark, PR Williamson, CB Terwee

www.cosmin.nl

www.comet-initiative.org

Thank you

VUmc VU University Medical Center Amsterdam

c.prinsen@vumc.nl

UniversitätsCentrum Evidenzbasierte Gesundheitsversorgung

Universitätsklinikum Carl Gustav Carus DIE DRESDNER.

Developing a core outcome set for (stage IV) melanoma trials – results of a systematic review

Stefanie Deckert, Melanie Schubert, Sanna Prinsen, Marlene Garzarolli, Jochen Schmitt