The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument

Lidwine B. Mokkink[1], Cecilia A. C. Prinsen[1], Lex M. Bouter[1],
Henrica C. W. de Vet[1], Caroline B. Terwee[1]

ABSTRACT | Background: COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) is an initiative of an international multidisciplinary team of researchers who aim to improve the selection of outcome measurement instruments both in research and in clinical practice by developing tools for selecting the most appropriate available instrument. **Method:** In this paper these tools are described, i.e. the COSMIN taxonomy and definition of measurement properties; the COSMIN checklist to evaluate the methodological quality of studies on measurement properties; a search filter for finding studies on measurement properties; a protocol for systematic reviews of outcome measurement instruments; a database of systematic reviews of outcome measurement instruments; and a guideline for selecting outcome measurement instruments for Core Outcome Sets in clinical trials. Currently, we are updating the COSMIN checklist, particularly the standards for content validity studies. Also new standards for studies using Item Response Theory methods will be developed. Additionally, in the future we want to develop standards for studies on the quality of non-patient reported outcome measures, such as clinician-reported outcomes and performance-based outcomes. **Conclusions:** In summary, we plea for more standardization in the use of outcome measurement instruments, for conducting high quality systematic reviews on measurement instruments in which the best available outcome measurement instrument is recommended, and for stopping the use of poor outcome measurement instruments.

Keywords: COSMIN; measurement properties; outcome measures; systematic reviews of instruments; outcome selection.

BULLET POINTS

- COSMIN aims to improve instrument selection in research and clinical practice.
- Description of COSMIN tools for selecting most appropriate instrument.
- Call for standardization in instrument use.
- Call for conducting high quality systematic reviews on instruments.
- Call for stopping the use of poor measurement instruments.

HOW TO CITE THIS ARTICLE

Introduction

COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) is an initiative of an international multidisciplinary team of researchers with a background in epidemiology, psychometrics, qualitative research, and health care, who have expertise in the development and evaluation of outcome measurement instruments[1]. The COSMIN initiative aims to improve the selection of outcome measurement instruments both in research and in clinical practice by developing tools for selecting the

most appropriate instrument. The COSMIN Steering Committee (see Appendix 1), founded in 2005, was inspired by a lack of clarity in the literature about terminology and definitions of measurement properties. Moreover, there exists an impressive amount of outcome measurement instruments and there are even many instruments measuring the same construct, developed for the same patient population, and still new ones are being developed. So researchers and clinicians

have to choose the most suitable instrument for their application.

The process of selecting outcome measures for specific purposes is complex. Choices involve conceptual considerations, such as defining the construct and population; practical aspects, such as burden for patients and raters, and costs; and quality aspects assessed by nine different measurement properties clustered in the domains reliability, validity and responsiveness[2]. Selecting unsuitable or poor quality outcome measurement instruments may introduce bias in the conclusions of studies. This may lead to a waste of resources and be unethical because participating patients contribute little or nothing to the body of knowledge but still suffer from the burdens and risks of the study[3].

An additional problem is that in systematic reviews of clinical trials the results reported cannot be compared and statistically pooled when different instruments are used to measure the same construct of interest in each study. Moreover, in clinical trials evaluating the benefits and harms of health care interventions, often a great variety of outcomes are reported. This makes it even more difficult to compare and combine results. This hampers the usefulness of clinical trial evidence to inform clinicians, at the cost of the best possible care for patients. Standardization in outcomes and outcome measurement instruments in specific areas of research is therefore highly warranted.

The COSMIN initiative wants to improve the selection of outcome measurement instruments by developing methodological guidelines based on consensus reached in a broad international panel of experts. The initial focus was on patient-reported outcome measures (PROMs). Therefore, the focus of this paper is only on PROMs.

First, some conceptual considerations concerning the selection of an outcome measurement instrument are explained. Next, the tools yielded by the COSMIN initiative will be described. Finally, we describe our future plans for research.

Conceptual considerations when selecting outcome measurement instruments

It is important to understand the distinction between an outcome and an outcome measurement instrument. An outcome refers to the construct of interest. Since we talk about patient-reported outcomes, the outcome is often a phenomenon that cannot be observed directly,

for example fatigue or health-related quality of life. The outcome chosen defines *what* is being measured. An outcome measurement instrument refers to *how* the outcome is being measured. It refers to the specific outcome measurement instrument. For example, the Neurological Fatigue Index for multiple sclerosis (NFI-MS)[4] or the Skindex-29[5] to measure quality of life in dermatology.

When selecting an outcome measurement instrument for research or clinical practice, first the outcome to be measured should be clearly defined. That is, one should define what to measure. For example, when measuring a broad construct such as health-related quality of life, it should be clarified which subdomains are relevant for the target population in the specific context of interest. Sometimes several definitions exist for an outcome. There are, for instance, multiple definitions for the construct 'disability'. The World Health Organization (WHO) defines 'disability' as a broad concept: 'problems an individual may experience in functioning, namely impairments, activity limitations and participation restrictions'[6]. Nagi[7] defined disability more narrowly as 'a pattern of behaviour that evolves in situations of long-term or continued impairment that are associated with functioning limitations' (previously called 'handicap' in the International Classification of Functioning of the WHO[8]). Without explicitly defining or describing the intended outcome, people may have different ideas about it and interpret it differently.

Next, one has to choose a specific instrument. Often, for the same outcome multiple measurement instruments are available. To select the best available outcome measurement instrument the COSMIN initiative has yielded several tools.

Standardization of the selection of outcomes and outcome measurement instruments in specific areas of research will improve consistencies in reporting and decrease difficulties in comparing and combining the findings in systematic reviews and meta-analyses. This can be obtained by the development of Core Outcome Sets (COS). A COS is an agreed standardized set of outcomes that should be measured and reported, as a minimum, in all clinical trials in a specific disease or trial population (i.e. *what* to measure)[9]. Once the COS is defined, it is then important to achieve consensus on which outcome measurement instruments should be selected to measure the core outcomes, referring to Core Outcome Measurement Instruments (i.e *how* to measure)[10]. The existence or use of a core outcome set does not imply that outcomes in a particular trial should

be restricted to those in the relevant core outcome set. Rather, there is an expectation that the core outcomes will be collected and reported, making it easier for the results of trials to be compared, contrasted and combined as appropriate; while researchers continue to explore other outcomes as well[11].

COSMIN tools

The COSMIN initiative has developed the following tools to help researchers and clinicians choosing the most appropriate outcome measurement instrument:

1. COSMIN taxonomy and definitions of measurement properties;

2. COSMIN checklist to evaluate the methodological quality of studies on measurement properties;

3. Search filter for finding studies on measurement properties;

4. Protocol for systematic reviews of outcome measurement instruments;

5. Database of systematic reviews of outcome measurement instruments;

6. Guideline for selecting outcome measurement instruments for outcomes included in a Core Outcome Set.

We performed an international Delphi study aiming to develop consensus-based standards for assessing the methodological quality of studies on measurement properties[1,2,12-14]. Results from this study were the COSMIN taxonomy and definitions, and the COSMIN checklist.

COSMIN taxonomy and definitions

We first developed a taxonomy and reached consensus on definitions of the measurement properties (see Table 1)[2]. Nine measurement properties clustered within three domains, i.e. reliability, validity and responsiveness, were considered relevant in the evaluation of outcome measurement instruments (Figure 1).
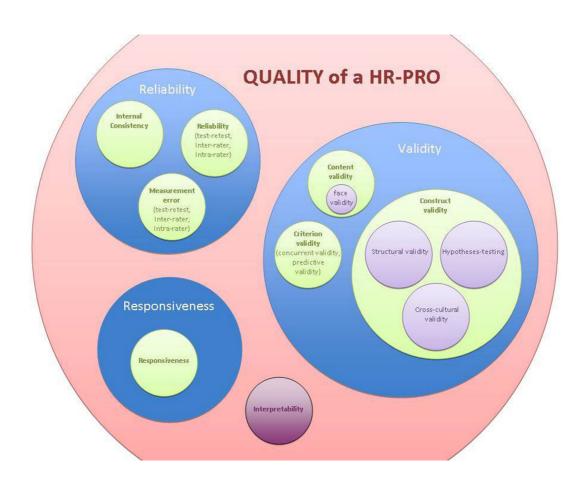
Table 1. Definitions of domains, measurement properties, and aspects of measurement properties.

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*			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[15]. *Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument.

COSMIN checklist

We developed a critical appraisal tool, i.e. the COSMIN checklist[12-14], containing standards for evaluating the methodological quality of studies on the measurement properties of outcome measurement instruments. The COSMIN checklist and a supplementary manual can be optained from the COSMIN website[16]. For each measurement property a box with standards was developed. These standards describe design requirements and preferred statistical methods. For example, in a high quality study of internal consistency, first a check for the unidimensionality of the (sub)scale should be done (Box A item 5 of the COSMIN checklist)[12]. Subsequently the internal consistency statistic should be calculated for the items of this unidimensional (sub) scale (Box A item 7)[12]. Other standards concern, for instance, using an appropriate time interval between test and retest administration when investigating test-retest reliability and measurement error (Box B and C item 8)[12], or formulating a priori hypotheses for hypotheses testing (a form of construct validity) (Box F item 4)[12].

When examining the interrater reliability and agreement of the items of the COSMIN checklist, we found that the reliability of the individual items was low (i.e. only 6% of the items had a Kappa statistic above 0.75), but that the agreement between raters was appropriate for 80% of the items[17]. When using the COSMIN checklist in a systematic review, we recommend getting some prior on-the-job training and experience, completing it by two independent raters, and reaching consensus about the ratings[17]. To use the COSMIN checklist in a systematic review of measurement instruments, we developed a four-point rating system for scoring the items of the COSMIN checklist[14]. With this version it is possible to calculate overall methodological quality scores per study on a measurement property. This is useful and enlightening in systematic reviews, as it allows to present conclusions on the quality of the instruments under study accompanied by various levels of evidence[14].

Search filter for finding studies on measurement properties

To facilitate the selection of outcome measurement instruments to be included in a systematic review of measurement instruments, a search filter was developed and validated in cooperation with clinical librarians for finding studies on measurement properties in PubMed[18]. In such a review, the filter can be combined with search terms for the outcome and the population of interest. The filter for finding studies on measurement properties showed to have a sensitivity of 97.4% and a positive predictive value of 4.4%. We translated this filter for EMBASE and CINAHL, and all filters are available from the COSMIN website[16].

Protocol for systematic reviews of outcome measurement instruments

Systematic reviews of outcome measurement instruments are important for the evidence-based selection of instruments. In such a review, the measurement properties of all outcome measurement instruments for a specific construct in a specific population are described and compared according to predefined criteria, and a conclusion is drawn about the most appropriate instrument.

We developed a protocol for performing systematic reviews of measurement instruments, including a 10-step procedure (available from the COSMIN website). In this protocol we describe how the COSMIN search filter[18] can be used to identify all relevant outcome measurement instruments, as well as how the COSMIN checklist[12] can be used to assess the quality of the included studies. In addition to the search filter for studies on measurement properties, and if the review concerns PROMs, a PROM filter developed by the University of Oxford can be used (available from the COSMIN website).

In addition, we describe the method of a best evidence synthesis in which the number of studies, their quality and (consistency of) results can be combined to determine the strength of the evidence for each measurement property. For example, strong evidence for a positive reliability is obtained when consistent positive results (ICCs or Kappas >0.70) are found in at least two studies of good quality or one study of excellent quality. The procedure is similar to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach[19] that is used in reviews of clinical trials. Previously developed cut-off values (such as ICC or Kappas >0.70) are used to determine whether an outcome measurement instrument has good measurement properties[20].

In 2009 we concluded, based on a review of systematic reviews of measurement instruments, that the quality of these reviews should and could be improved[21]. Recently, we updated this review, and concluded that the quality of published systematic reviews of measurement instruments has improved[22]. However,

there is still room for improvement with regards to the search strategy, and especially the quality assessment of the included studies and instruments as well as the data synthesis. Therefore, we are currently updating the protocol for performing systematic reviews of measurement instruments, aiming to publish it as a peer-reviewed guideline for systematic reviews of outcome measurement instruments (manuscript in preparation). In this way, we aim to contribute to the improvement of systematic reviews of measurement instruments.

Database for systematic review of outcome measurement instruments

The COSMIN initiative maintains an overview of published systematic reviews of outcome measurement instruments. This overview is presented in a searchable database available on the COSMIN website[16]. Currently, it contains 569 systematic reviews and we aim to update this overview yearly. The COSMIN database provides a good starting point to search for and select outcome measurement instruments.

Guideline for selecting outcome measurement instruments for outcomes included in a Core Outcome Set

COSMIN collaborated with the COMET (Core Outcome Measures in Effectiveness Trials) initiative to develop a guideline for the selection of outcome measurement instruments for outcomes included in a COS[10]. We reached consensus among a large group of experts on four main steps in the selection of outcome measurement instruments for COS: Step 1) conceptual considerations; Step 2) finding existing outcome measurement instruments; Step 3) quality assessment of outcome measurement instruments; and Step 4) generic recommendations on the selection of outcome measurement instruments for outcomes included in a COS. The resulting consensus-based guideline can be used by COS developers in defining *how* to measure core outcomes (submitted publication by Prinsen CA, et al. How to select outcome measurement instruments for outcomes included in a 'Core Outcome Set' – a practical guideline).

● **Ongoing and future studies**

At the moment, we work on updating the COSMIN checklist. Over the past years, users of the COSMIN checklist have identified gaps in the available standards. Recent regulatory guidelines on outcome measurement instruments development and evaluation call for an extension of the COSMIN checklist with respect to its standards for the quality of studies on content validity within the specific context of interest. Therefore, a Delphi study is underway which aims to reach consensus on new COSMIN standards and criteria for evaluating the content validity (including face validity) of PROMs. In these new standards, the quality of the development process of PROMs will be taken into account, and criteria for what constitutes good content validity will be developed.

In addition, a shift has taken place in recent years from the use of traditional statistical methods (i.e. Classical Test Theory (CTT)) to the recommended use of newer statistical methods (e.g. Item Response Theory (IRT)[23] and Rasch Measurement Theory[24]) analyses for developing and evaluating outcome measurement instruments. This requires an extension of the COSMIN standards for studies using IRT and Rasch methods. Clear methodological advantages of using IRT or other modern test theory methods over or in addition to CTT have been described[25]. Well-developed IRT-based instruments, have probably better measurement properties than CTT-based instruments[26,27]. In addition, IRT allows for Computer Adaptive Testing (CAT), a method of questionnaire administration in which a computer algorithm iteratively selects questions based on previous answers. Questionnaires that are completed by CAT dramatically decreases the burden for patients to complete questionnaires and improving precision[28-31]. Examples of IRT-based instruments are the Patient Reported Outcomes Measurement Information System (PROMIS) instruments, which are available as CAT instruments as well as static short forms[32]. A next step to be addressed is to achieve consensus among an international group of experts on standards for the methodological quality of studies using IRT and Rasch methods for evaluating measurement properties and to operationalize these standards into a user-friendly and easily applicable checklist to be used e.g. in systematic reviews of outcome measurement instruments.

The COSMIN standards were originally developed for evaluating the quality of studies on the measurement properties of PROMs. Although the COSMIN standards have also been used in systematic reviews of other types of outcome measurement instruments, adaptations are required to use the COSMIN standards for evaluating the quality of studies on the measurement properties of other patient-centered outcome measurement instruments, such as clinician-reported outcome measure

(e.g. a goniometer to measure range of motion), or a performance based test (e.g. a six minute walk test to measure walking speed). It is our ambition to develop new standards specific for other types of instruments.

Finally, we want to develop reporting guidelines for studies on measurement properties, and for systematic reviews on measurement properties.

● Need for high quality systematic reviews of outcome measurement instruments and Core Outcome Set development

By the development of the COSMIN tools described above and by generating awareness for the importance of selecting high quality instruments, COSMIN aims to accomplish that researchers and clinicians make their choices on outcomes and outcome measurement instruments more informed. We plea for more standardization in the use of outcomes and outcome measurement instruments. We support the aim of the COMET initiative to stimulate the development of COS. The use of COS will lead to more standardization in outcome reporting in specific areas of research, making it easier for the results of trials to be compared and combined as appropriate. COSMIN strongly encourages researchers to perform high quality systematic reviews of outcome measurement instruments. More high quality systematic reviews of outcome measurement instruments are needed to make an informed choice for the best instrument for a specific purpose and for stopping the use of poor outcome measurement instruments.

● References

1. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006;6(1):2. http://dx.doi.org/10.1186/1471-2288-6-2. PMid:16433905.

2. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45. http://dx.doi.org/10.1016/j.jclinepi.2010.02.006. PMid:20494804.

3. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383(9912):166-75. http://dx.doi.org/10.1016/S0140-6736(13)62227-8. PMid:24411645.

4. Mills RJ, Young CA, Pallant JF, Tennant A. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). Health Qual Life Outcomes. 2010;8(1):22. http://dx.doi.org/10.1186/1477-7525-8-22. PMid:20152031.

5. Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. J Invest Dermatol. 1996;107(5):707-13. http://dx.doi.org/10.1111/1523-1747.ep12365600. PMid:8875954.

6. World Health Organization – WHO. ICF: International Classification of Functioning, disability and health. Geneva: WHO; 2001.

7. Nagi SZ. Some conceptual issues in disability and rehabilitation. In: Sussman MB, editor. Sociology and rehabilitation. Columbus: Ohio State University Press; 1965. p. 100-13.

8. World Health Organization – WHO. International classification of International Classification of Impairments, Disabilities, and Handicaps. Geneva: WHO; 1980.

9. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13(1):132. http://dx.doi.org/10.1186/1745-6215-13-132. PMid:22867278.

10. Prinsen CA, Vohra S, Rose MR, King-Jones S, Ishaque S, Bhaloo Z, et al. Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set'. Trials. 2014;15(1):247. http://dx.doi.org/10.1186/1745-6215-15-247. PMid:24962012.

11. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO 3rd, Conaghan PG, et al. The OMERACT handbook. OMERACT; 2015.

12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49. http://dx.doi.org/10.1007/s11136-010-9606-8. PMid:20169472.

13. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10(1):22. http://dx.doi.org/10.1186/1471-2288-10-22. PMid:20298572.

14. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651-7. http://dx.doi.org/10.1007/s11136-011-9960-1. PMid:21732199.

15. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: University Press; 2008.

16. COnsensus-based Standards for the selection of health Measurement INstruments – COSMIN [Internet]. Amsterdam: COSMIN; 2015 [cited 2015 Sept. 28]. Available from: http://www.cosmin.nl

17. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability

of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Med Res Methodol. 2010;10(1):82. http://dx.doi.org/10.1186/1471-2288-10-82. PMid:20860789.

18. Terwee CB, Jansma EP, Riphagen II, Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115-23. http://dx.doi.org/10.1007/s11136-009-9528-5. PMid:19711195.

19. GRADE working group [Internet]. 2005 [cited 2015 Sept. 28]. Available from: http://www.gradeworkinggroup.org

20. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012. PMid:17161752.

21. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. 2009;18(3):313-33. http://dx.doi.org/10.1007/s11136-009-9451-9. PMid:19238586.

22. Terwee CB, Prinsen CA, Garotti MGR, Suman A, Vet HC, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res. 2015. http://dx.doi.org/10.1007/s11136-015-1122-4. PMid:26346986.

23. Lord FM. Applications of item response theory to practical testing problems. Mahwah: Erlbaum; 1980.

24. Andrich D. Rasch models for measurement. Beverly Hills: Sage Publications; 1988.

25. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates; 2000.

26. Fries JF, Bruce B, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. Ann Rheum Dis. 2006;65(Suppl 3):iii16-21. http://dx.doi.org/10.1136/ard.2006.059279. PMid:17038464.

27. Lai JS, Cella D, Choi S, Junghaenel DU, Christodoulou C, Gershon R, et al. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. Arch Phys Med Rehabil. 2011;92(10 Suppl):S20-7. http://dx.doi.org/10.1016/j.apmr.2010.08.033. PMid:21958919.

28. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther. 2011;13(5):R147. http://dx.doi.org/10.1186/ar3461. PMid:21914216.

29. Thissen D, Reeve BB, Bjorner JB, Chang CH. Methodological issues for building item banks and computerized adaptive scales. Qual Life Res. 2007;16(S1 Suppl 1):109-19. http://dx.doi.org/10.1007/s11136-007-9169-5. PMid:17294284.

30. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. J Clin Epidemiol. 2006;59(11):1174-82. http://dx.doi.org/10.1016/j.jclinepi.2006.02.010. PMid:17027428.

31. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. Qual Life Res. 2007;16(S1 Suppl 1):95-108. http://dx.doi.org/10.1007/s11136-007-9168-6. PMid:17530450.

32. PROMIS – Dynamic tools to measure health outcomes from the patient perspective [Internet]. Chapel Hill: PROMIS; 2015 [cited 2015 Sept. 28]. Available from: http://www.nihpromis.org

Correspondence
Lidwine B. Mokkink
VU University Medical Center
Institute for Health and Care Research
Department of Epidemiology and Biostatistics and EMGO
De Boelelaan 1089a, 1081 HV
Amsterdam, the Netherlands
e-mail: w.mokkink@vumc.nl

Appendix 1. COSMIN steering committee members.

Original COSMIN steering committee

The original COSMIN steering committee was responsible for the initiation and organisation of the Delphi study in which the COSMIN taxonomy and COSMIN checklist was developed. Members of the original steering committee were:

Lidwine B (Wieneke) Mokkink, Assistant professor in Clinimetrics, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Caroline B Terwee**, Associate professor in Measurement, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Jordi Alonso**, Head of the Health Services Research Group. Epidemiology and Public Health Program. Institut Municipal d'Investigació Mèdica (IMIM-Hospital del Mar), Barcelona, Spain. **Donald L Patrick**, Professor of Health Services and Director of the Social and Behavioral Sciences Program at the University of Washington, Director of the Seattle Quality of Life Group, Seattle, US. **Lex M Bouter**, Professor of Methodology and Integrity, VU University and Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Henrica CW de Vet**, Professor of Clinimetrics, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Dirk L Knol**, Statistician, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Paul W Stratford**, Professor School of Rehabilitation Science, Faculty of Health Sciences, and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada.

Current COSMIN steering committee

The current steering committee is responsible for updating the COSMIN checklist. Members of the current steering committee are:

Lidwine B (Wieneke) Mokkink, Assistant professor in Clinimetrics, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Cecilia AC (Sanna) Prinsen**, Assistant professor in Clinimetrics, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Jordi Alonso**, Head of the Health Services Research Group. Epidemiology and Public Health Program. Institut Municipal d'Investigació Mèdica (IMIM-Hospital del Mar), Barcelona, Spain. **Donald L Patrick**, Professor of Health Services and Director of the Social and Behavioral Sciences Program at the University of Washington, Director of the Seattle Quality of Life Group, Seattle, US. **Lex M Bouter**, Professor of Methodology and Integrity, VU University and Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Henrica CW de Vet**, Professor of Clinimetrics, Department of Epidemiology and Biostatistics, Emgo Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. **Marjan Westerman**, Assistan professor, expert in qualitative research, VU University, Amsterdam, the Netherlands. **Caroline B Terwee**, Associate professor in Measurement, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. Dirk Knol and Paul Stratford have retired from the COSMIN group in 2015.