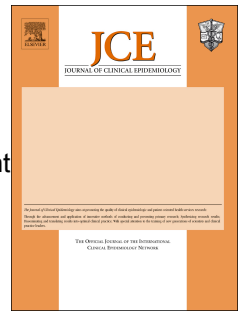

At the CORE of measurement – What do you want to measure? And how do you want to measure it? A COSMIN Perspective

L.B. Mokkink, S. Herbelet, C.B. Terwee, M. Boers

PII: S0895-4356(25)00169-6

DOI: https://doi.org/10.1016/j.jclinepi.2025.111836

Reference: JCE 111836

To appear in: *Journal of Clinical Epidemiology*

Received Date: 17 February 2025

Revised Date: 7 May 2025

Accepted Date: 14 May 2025

Please cite this article as: Mokkink L, Herbelet S, Terwee C, Boers M, At the CORE of measurement – What do you want to measure? And how do you want to measure it? A COSMIN Perspective, *Journal of Clinical Epidemiology* (2025), doi: https://doi.org/10.1016/j.jclinepi.2025.111836.

At the CORE of measurement – What do you want to measure? And how do you want to measure it? A COSMIN Perspective

LB Mokkink[1, 2], S Herbelet[1,3], CB Terwee[1, 2], M Boers[1,2]

[1] Amsterdam UMC, Vrije Universiteit Amsterdam, Departement of Epidemiology and Data Science, Amsterdam, Netherlands

[2] Amsterdam Public Health research institute, Methodology, Amsterdam, Netherlands

[3] Ghent University, Department of Head and Skin, Ghent, Belgium

Corresponding author:

LB (Wieneke) Mokkink, w.mokkink@amsterdamumc.nl. Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Location AMC, J1B-225, Meibergdreef 9, Amsterdam 1105 AZ, The Netherlands

Abstract

Any measurement starts with determining what to measure (i.e. which outcome) and how to measure it (i.e. which outcome measurement instrument).

This article is the first from a series of COSMIN Perspective in which we will explain measurements and how to evaluate measurement quality. In this first Perspective we explain the rationale for defining an outcome and operationalizing it into an outcome measurement instrument.

An outcome should first be defined based on theory, and next be operationalized into one or more observable items, questions, tasks or parameters. But a measurement instrument is more than its items or machine. Any type of measurement instrument consist of five components: (1) equipment, (2) preparatory actions, (3) unprocessed data collection *or* collection of the biological sample(s), (4) processing and storage, and (5) assignment of the score *or* determination of the value of the biological sample.

Thinking in terms of the components of measurement instruments can enhance the design and understanding of studies on measurement properties.

Key words: COSMIN, measurement, outcome, measurement instrument

Running title: At the CORE of measurement – a COSMIN perspective

Word count: 1513

Background

COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) was initiated by an international group of researchers aiming to improve the selection of outcome measurement instruments of health outcomes by developing and encouraging the use of transparent methodology and practical tools to select the most suitable outcome measurement instruments in research and clinical practice. This includes consensus-based guidelines for conducting and reporting studies of measurement properties and systematic reviews of outcome measurement instruments. By now, COSMIN has become more than a set of standards and tools. It is a way of thinking and asking the right questions about measurement instruments. In a series of papers, called COSMIN Perspectives, we will explain different measurement properties and measurement-related issues. In this first perspective, we will start with the core of measurement: determining what you want to measure (i.e. which outcome) and how you want to measure it (i.e. which outcome measurement instrument). We offer a perspective for defining and operationalizing outcomes, and guidance on how to develop and delineate or select a measurement instrument. In future perspectives we will reflect on how to check whether the operationalization fits the outcome, that is, evaluating the measurement properties of measurement instruments (see also (1)).

What is an outcome? Its definition.

An outcome refers to *what* is being measured. It is also called a construct, domain, concept, or factor of interest. Sometimes, an outcome is a latent construct, which is a phenomenon that cannot be directly measured, such as fatigue, social participation, or quality of life. Such outcomes need to be defined and operationalized to be measurable. Other outcomes can be directly measured, but need equipment to do so. Examples of such outcomes are blood pressure, muscle thickness, and hemoglobin levels.

People may have different ideas of what an outcome refers to. For example, quality of life can refer to different (sub)constructs (2), or pain can refer to current pain intensity, average pain last week, pain duration, etc. Therefore, it is important to further specify in more detail what one wants to measure. This can be done by defining the outcome, which is essentially a statement of one's understanding of the outcome. The definition of the outcome should be detailed enough to understand exactly what belongs to the outcome and what doesn't. This is important both in the development of an instrument and in the selection of an existing instrument for a study or clinical application.

In the process of defining an outcome, a conceptual model is essential. Examples of conceptual models are the model developed by Wilson and Cleary, who distinguished multiple levels of health (3), the International Classification of Functioning, Disability and Health (ICF), developed by the World Health Organization (4), or the OMERACT filter 2.0 framework (5). These models may help understand what the outcome of interest is, and what it is not. In the Wilson and Cleary model, for example, a distinction is being made between symptoms and functional status (among other levels of health). If the purpose is to measure physical function, which is an aspect of functional status, one should not mix it up with measures of symptoms that may influence physical function, such as fatigue or pain. So, references to symptoms should not be included in the definition of physical function. If an outcome is multi-dimensional, each sub-outcome should be defined separately. The choice for a specific conceptual model depends on what works best for the outcome of interest.

Defining the outcome is also very important for deciding how to measure it. For example, if one wants to measure physical function, one must first define physical function in more detail before an instrument can be chosen. Physical function can be defined as 'what a person can do in a standardized, controlled environment' (also called 'capacity') (6). This is best measured with a performance-based test e.g. in a hospital or lab. However, physical function can also be defined as 'what a person actually does in the usual circumstances of his/her everyday life' (which is – unfortunately – also called 'performance') (6), which is best measured with an accelerometer. Lastly,

one can define physical function as self-reported capability (i.e. 'what a person perceives or experiences they are capable of doing in their daily environment' (6)) rather than actual performance of physical activities (7), which is best measured with a PROM.

What is a measurement instrument? An operationalization of an outcome

Once the outcome has been defined, a measurement instrument should be developed or a suitable instrument should be selected. A measurement instrument can be considered an operationalization of the outcome. An operationalization of an outcome is the process by which outcome and theories are converted into concrete, measurable and executable terms. Depending on the choice of the type of measurement instrument, it can be operationalized into (one or more) observable items, questions, tasks or parameters.

The goal of measuring, i.e. applying a measurement instrument, is to quantify a well-defined outcome. Different types of measurement instruments exist, e.g. PROMs, clinical ratings, imaging-based instruments, performance-based test, and laboratory tests. All types of measurement instruments can be considered a set of components that make up (or delineate) the instrument. In a Delphi study (8) we developed two sets: one for outcome measurement instruments that do not involve biological sampling and one for outcomes measurement instruments that involve biological sampling.

The first set consists of the following components: (1) the equipment (all equipment used in preparation, administration, and assigning scores); (2) preparatory actions ('first time only' general preparatory actions, such as required expertise or training for professionals to prepare, administer, store or assign the scores; as well as specific preparatory actions for each measurement, such as preparations of equipment by professionals (e.g. calibration), preparations of the patient by the professional (e.g. giving instructions), and preparations undertaken by the patients (e.g. follow up dietary restrictions)); (3) unprocessed data collection (what the patient and/or professional(s)

actually do to obtain the unprocessed data); (4) data processing and storage (all actions undertaken on the unprocessed data to allow the assignment of the score); and (5) assignment of the score (methods used to transform processed data into a final score on the outcome measurement instrument) (8). In Figure 1 we give an example of the components of a PROM as well as for an MRI-based parameter.

----- Please, insert Figure 1 here ------

Figure 1. Components of measurement instruments of a PROM and an MRI-based parameter.

For instruments where biological sampling is involved (e.g. to measure laboratory values) the components are: (1) equipment, (2) preparatory actions preceding sample collection, (3) collection of the biological sample(s), (4) biological sampling processing and storage, and (5) determination of the value of the biological sample. More details are described elsewhere (8).

A measurement instrument is much more than its equipment: equipment is the head – the score is the tail. Some measurement instruments are very simple, e.g. containing only one question that a patient can answer with a five-point Likert scale, even without any instructions. Other measurement instruments are rather complex. For example, an MRI machine is not the measurement instrument, it is only the equipment. The measurement instrument could for example be 'MRI-determined muscle anatomical cross-sectional area'. In this case, the term 'measurement protocol' may be preferred over 'measurement instrument'. The protocol describes all of the above-mentioned components in detail (see Figure 1), including how the final score for the specific construct is finally determined. Therefore, each parameter measured with an MRI machine refers to a different measurement instrument. This may seem exaggerated, but when one wants to evaluate the quality of the measurement instruments

(that is, evaluating whether the outcome is well operationalized), each score must correspond to a (specific) outcome.

Checking the quality of the operationalization

The process of defining and operationalizing an outcome is an iterative process, based on theory, literature and expert (e.g. patient and health professional) opinion. Once the outcome has been operationalized (that is to say, the instrument has been developed or selected), it is necessary to return to the theory and definition of the outcome that is being measured in order to ascertain whether the instrument is aligned with the defined outcome. In other words, if the instrument is valid for measuring the outcome of interest. Hopefully, this has been properly done by others, and you can select an existing measurement instrument. If not, or if the outcome of interest (or the population) is (slightly) different from where the instrument was developed for, the measurement instrument needs to be modified or newly developed.

Instrument quality refers to the measurement properties (validity, reliability, and responsiveness) of the instrument. These should be assessed and documented before an instrument can be used in research or clinical practice. Thinking in terms of the components of measurement instruments can enhance the design and understanding of studies on measurement properties. For example, by considering how the components of the instrument under study are operationalized or what part of the measurement instrument is the focus of interest in a specific study (e.g. to evaluate reliability). In future COSMIN Perspectives we will explain this in more detail.

Further reading

Flake JK and Fried EI. Measurement Schmeasurement: Questionable Measurement Practices

and How to Avoid Them. Advances in Methods and Practices in Psychological Science 2020, Vol. 3(4) 456–465 DOI: 10.1177/251524592095239

In this paper the importance of defining psychological constructs is described, among other issues.

Lilienfeld SO, Strother AN. Psychological measurement and the replication crisis: Four sacred cows. Canadian Psychology / Psychologie Canadienne. 2020; 61:281-288. https://doi.org/10.1037/cap0000236

The authors reflect on the questions whether we can safely rely on the name of a measure to infer its content, among other interesting issues.

Beaton D, Maxwell L , Grosskleg S, Shea B, Tugwell P (editors). The OMERACT Handbook Version 2.1 [updated April 2021]. OMERACT. Available from https://omeract.org/handbook/

This handbook offers a detailed procedure for how to select outcomes.

References

1. Mokkink L, Terwee C, de Vet H. Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. J Clin Epidemiol. 2021;140:159-62.
2. Macefield RC, Jacobs M, Korfage IJ, Nicklin J, Whistance RN, Brookes ST, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). Trials. 2014;15:49.
3. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995;273(1):59-65.
4. Organization WH. ICF: international classification of functioning, disability and health. Geneva: World Health Organization; 2001.
5. Boers M, Beaton DE, Shea BJ, Maxwell LJ, Bartlett SJ, Bingham CO, 3rd, et al. OMERACT Filter 2.1: Elaboration of the Conceptual Framework for Outcome Measurement in Health Intervention Studies. J Rheumatol. 2019;46(8):1021-7.
6. Holsbeeke L, Ketelaar M, Schoemaker MM, Gorter JW. Capacity, capability, and performance: different constructs or three of a kind? Arch Phys Med Rehabil. 2009;90(5):849-55.
7. HealthMeasures.net. User Manual and Scoring instructions, PROMIS Physical Function. 2023. Accessed: 27-01-2025. Available from: https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Physical_Function_User_Manual_and_Scoring_Instructions_02202023.pdf.
8. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, 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;20(293).

	PROM	MRI-based parameter
Equipment	Questions in the questionnaire with response options	Magnet room, equipment, control room
Preparatory actions	Instructions for patients	Setting of settings, place RF coils over specific body part, checking safety issues, comforting patient
Unprocessed data collection	Completion of questionnaire	Moment between switching on & of the radiofrequency source
Data processing & storage	Answers are entered as e.g. 0-4 score	Signals converted to images (K-space & Fourier transformation and stored on an computer)
Assignment of the score	Sumscore is calculated	Selection of specific slice(s), adjusting brightness, type of image (T1, T2) clinical interpretation of image

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: