



# Methodological quality of 100 recent systematic reviews of health-related outcome measurement instruments: an overview of reviews

Ellen B. M. Elsmann<sup>1,2</sup> · Lidwine B. Mokkink<sup>1</sup> · Inger L. Abma<sup>3</sup> · Olalekan Lee Aiyegbusi<sup>4</sup> · Alessandro Chiarotto<sup>5</sup> · Kirstie L. Haywood<sup>6</sup> · Karen Matvienko-Sikar<sup>7</sup> · Daniella M. Oosterveer<sup>8</sup> · Jan J. M. Pool<sup>9</sup> · Ilse E. J. Swinkels-Meewisse<sup>10</sup> · Martin Offringa<sup>2</sup> · Caroline B. Terwee<sup>1</sup>

Accepted: 29 May 2024 / Published online: 3 July 2024  
© The Author(s) 2024

## Abstract

**Purpose** Systematic reviews evaluating and comparing the measurement properties of outcome measurement instruments (OMIs) play an important role in OMI selection. Earlier overviews of review quality (2007, 2014) evidenced substantial concerns with regards to alignment to scientific standards. This overview aimed to investigate whether the quality of recent systematic reviews of OMIs lives up to the current scientific standards.

**Methods** One hundred systematic reviews of OMIs published from June 1, 2021 onwards were randomly selected through a systematic literature search performed on March 17, 2022 in MEDLINE and EMBASE. The quality of systematic reviews was appraised by two independent reviewers. An updated data extraction form was informed by the earlier studies, and results were compared to these earlier studies' findings.

**Results** A quarter of the reviews had an unclear research question or aim, and in 22% of the reviews the search strategy did not match the aim. Half of the reviews had an incomprehensive search strategy, because relevant search terms were not included. In 63% of the reviews (compared to 41% in 2014 and 30% in 2007) a risk of bias assessment was conducted. In 73% of the reviews (some) measurement properties were evaluated (58% in 2014 and 55% in 2007). In 60% of the reviews the data were (partly) synthesized (42% in 2014 and 7% in 2007); evaluation of measurement properties and data syntheses was not conducted separately for subscales in the majority. Certainty assessments of the quality of the total body of evidence were conducted in only 33% of reviews (not assessed in 2014 and 2007). The majority (58%) did not make any recommendations on which OMI (not) to use.

**Conclusion** Despite clear improvements in risk of bias assessments, measurement property evaluation and data synthesis, specifying the research question, conducting the search strategy and performing a certainty assessment remain poor. To ensure that systematic reviews of OMIs meet current scientific standards, more consistent conduct and reporting of systematic reviews of OMIs is needed.

**Keywords** Systematic reviews · Outcome measurement instruments · Measurement properties · Reliability · Validity · COSMIN

## Plain English summary

Instruments that measure health outcomes are important for making treatment decisions and understanding diseases. Systematic reviews are used to compare different instruments and help select the best one for a specific situation. Previous studies have shown that the quality of

these reviews can vary and may not always meet scientific standards. Since then, new tools and methods have been developed to help systematic review authors in improving the quality of their work. This study looked into the quality of recent systematic reviews of instruments. The study identified important improvements over time. For example, risk of bias is more often evaluated, and the data is analyzed in a better way. However, the study also shows that there are still areas that need improvement. These include formulating a clear research question, and

Extended author information available on the last page of the article

creating a comprehensive search strategy. Ongoing efforts are needed to improve the quality of systematic reviews of instruments. This can be achieved by developing new and accessible resources.

## Introduction

Outcome measurement instruments (OMIs) are used to evaluate the impact of disease and treatment [1–3]. When many different OMIs that measure similar constructs are available [1, 4, 5], the choice for an OMI depends on various aspects, including its quality (i.e., the sufficiency of measurement properties) [6]. Systematic reviews in which the measurement properties of OMIs are critically evaluated and compared are important tools for the selection of an OMI [4], for example in core outcome sets used in research projects or clinical practice [7]. With these systematic reviews, gaps in knowledge about the measurement properties of OMIs can also be identified.

Only well-designed, well-conducted, and comprehensively reported systematic reviews can provide a complete and balanced overview of the measurement properties of OMIs [4]. High-quality systematic reviews have: a well-defined research question; a comprehensive search strategy in multiple databases; independent abstract and full-text article selection; a risk of bias assessment of included studies; a systematic evaluation and syntheses of the results; and a certainty assessment of the body of evidence [8].

Previous overviews appraising the quality of systematic reviews of OMIs identified major limitations in the search strategy, the risk of bias assessment, and the evaluation and synthesis of the measurement properties' results [9, 10]. These limitations preclude systematic reviews to provide a complete and unbiased overview of the measurement properties of OMIs. This has consequences for knowledge users, who rely on the findings of these systematic reviews and might select suboptimal OMIs to use in their research or clinical practice [11]. This in turn impacts the measurements conducted on patients, which might be invalid and unreliable, and possibly even lead to incorrect healthcare decisions.

Various methodologies and practical tools have been developed to guide authors in conducting high-quality systematic reviews of OMIs [4, 12, 13]. The methodology and tools developed by the COSMIN (CONsensus-based Standards for the selection of health Measurement INSTRUMENTS) initiative are the most comprehensive and most widely used (Fig. 1) [14]. Since the most recent overview that assessed the quality of systematic reviews of OMIs, published in 2016 [10], the COSMIN guideline for systematic reviews has been

developed [4] and the COSMIN risk of bias checklist has been updated [15, 16]. Other methodologies and tools for critical appraisal of OMIs have also been developed and updated since then [12, 17]. When reading or reviewing such systematic reviews, even those that claim to have used these guidelines, we often observe flaws in the design, conduct, and reporting. The aim of this overview of reviews was therefore to investigate whether the quality of recent systematic reviews of OMIs lives up to the current scientific standards. As a secondary aim, we explored which aspects have notably improved over time.

## Methods

The study protocol was registered in the PROSPERO database, number CRD42022320675 [18]. There were no important deviations from the protocol. The study was reported according to the preferred reporting items for overviews of reviews (PRIOR) statement [19]. Consistent with the previous overview [10], we randomly selected 100 out of 136 most recent systematic reviews from the COSMIN database of systematic reviews [20]. These reviews were identified while updating the COSMIN database through a systematic literature search performed on March 17, 2022 in MEDLINE (through PubMed) and EMBASE (through [www.embase.com](http://www.embase.com)), and concerned systematic reviews of OMIs published from June 1, 2021 onwards. The search strategy consisted of search terms for systematic reviews, search terms for OMIs, and a validated search filter for measurement properties [21]. The full search strategy can be found in Supplementary File 1. Table 1 contains inclusion and exclusion criteria for the COSMIN database [20]. We defined systematic reviews of OMIs as peer-reviewed studies with a systematic search in at least one electronic database which aimed to summarize evidence on the measurement properties of all OMIs of interest to the review.

Eligibility for inclusion in the COSMIN database was determined by one reviewer (IS). All reviewers confirmed that each review appraised in the current study complied with the inclusion and exclusion criteria. If a review was selected from the COSMIN database that should have been excluded (false-positive), this review was replaced by a randomly selected new review after confirming exclusion by a third reviewer (LM).

A study-specific data extraction form (Supplementary File 2) was developed to appraise the quality of systematic reviews of OMIs, which includes both methodological quality and reporting quality—two aspects that cannot be considered separately when appraising the quality of published OMI systematic reviews. The data extraction form was based on criteria used in previous studies [9, 10], which were updated for this study. The data extraction form

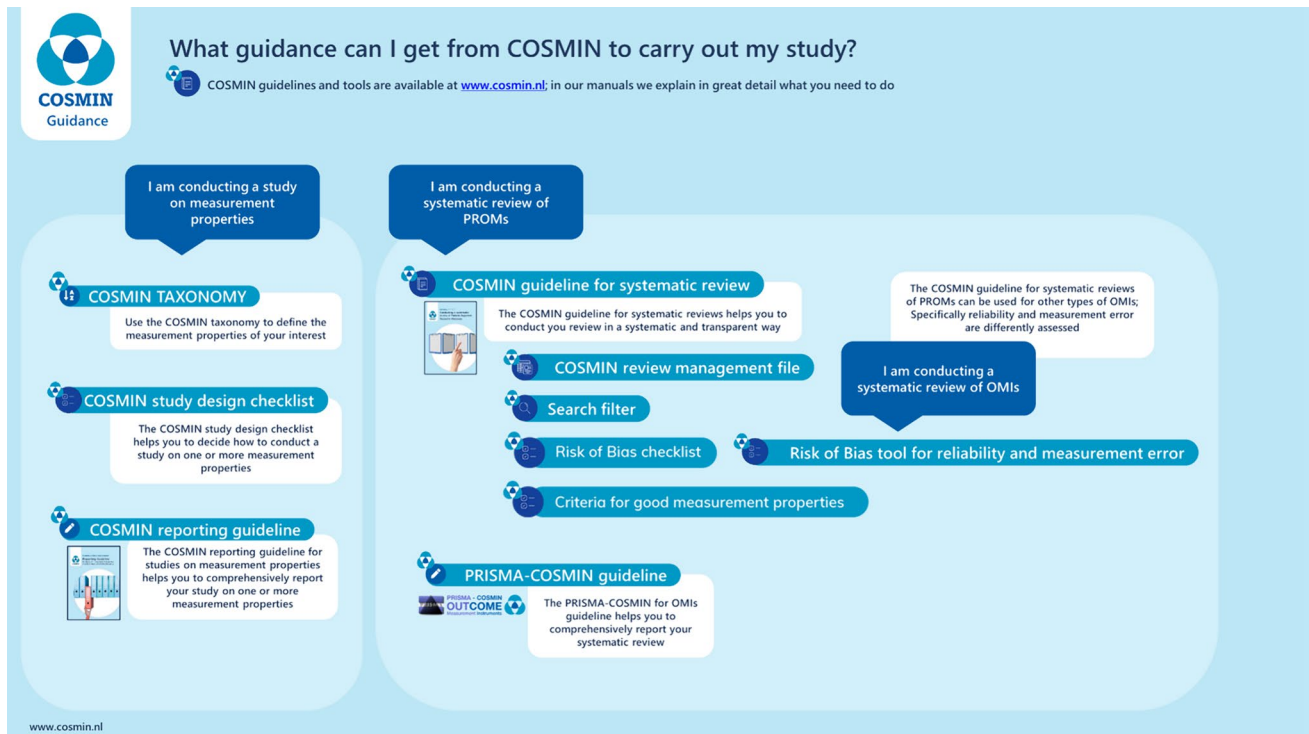


Fig. 1 COSMIN tools and methodology

Table 1 Inclusion and exclusion criteria for the COSMIN database [20]

Inclusion criteria	Exclusion criteria
Search performed in at least one electronic database	It concerns a prognostic review (i.e. aiming to predict an outcome using multivariable analyses)
Aim is to identify all OMI's of interest and summarize evidence on their measurement properties	Aim is to evaluate one or the most commonly used OMI's
Construct of interest is (aspect of) health status, based on Wilson & Cleary model [22]: Biological and physiological processes Symptoms Physical functioning Social/psychological functioning General health perceptions Health-related quality of life	Reviews that include only randomized controlled trials
Study population is humans (patients or general population)	Non-English reviews
Instrument of interest is OMI's, i.e., instruments which can be applied in longitudinal studies to monitor changes in health over time	Instrument is a diagnostic or screening instrument
At least one measurement property of the included OMI's is evaluated and reported	

contained items on the key elements of the review (i.e., construct, population, type of OMI, and measurement properties of interest), search strategy, eligibility criteria, article selection, data extraction, risk of bias assessment, evaluation of measurement properties, data synthesis, certainty assessment, presentation of results, instrument recommendation, and elements of open science). Specifically, the appropriateness of the search for the construct, population, type of OMI

and measurement properties was based on published search filters [21, 23], search terms found at [blocks.bmi-online.nl](http://blocks.bmi-online.nl), and the reviewers own knowledge. For each item, two independent reviewers extracted information on whether this was done/reported in the included reviews. No attempts were made to verify information with study authors. Reviewers also noted any major methodological and reporting flaws for each of these aspects.

The data extraction form was pilot tested with six OMI systematic reviews [24–29] by two independent reviewers (different pairs of EE, CT, and LM). A subsequent update was done after training the other reviewers, who were instructed to extract data for one of these six reviews [25]. Discrepancies were discussed during two 90-min Zoom meetings intended to standardize the data extraction process. After these meetings, the data extraction form and instructions on how to appraise each systematic review were finalized, and five pairs of reviewers were formed (EE&JP/IS, LM&DO, CT&IA, KH&KM, AC&OA). Each reviewer pair subsequently appraised the quality of 18–19 systematic reviews independently. Reviews were not appraised by a reviewer who was a co-author or had a potential conflict of interest. Discrepancies between the pair of reviewers were resolved through discussion. Appraisals of reviewers were descriptively synthesized by review counts and a qualitative comparison of the results was made to the results of previous studies [9, 10], if possible.

## Results

Characteristics of the 100 systematic reviews are presented in Table 2. Half of the included reviews focused on patient-reported outcomes, 30% focused on non-patient-reported outcomes, and 20% on a combination of both. The aspect of health of the construct of interest in the reviews was mostly functional status (62%), symptom status (56%), and/or general health perceptions (36%). Reviews focused on a variety of populations, such as children and (older) adults with a variety of diseases and conditions. Questionnaires (77%), clinical rating scales (41%), and/or performance-based tests (24%) were the OMI types most often included.

Syntheses of the quality appraisal of the 100 systematic reviews of OMIs [24–123] are presented in Table 3. Supplementary File 2 contains the completed data extraction form, whereas Supplementary File 3 contains the data from Table 3 in comparison with the results of the two previous studies [9, 10].

## Key elements

Only 11% of the reviews had a title that included all four key elements (i.e., construct, population, type of OMI, and measurement properties of interest) and the fact that it concerned a systematic review. In titles of the remaining reviews, often no reference to measurement property evaluation was made. 47% of the reviews had a title that omitted at least 2 key elements and/or the fact that it concerned a systematic review. The term ‘scoping review’ was used in 7% of the reviews. In 45% of the reviews all 4 key elements were included in the aim, whereas

**Table 2** Characteristics of systematic reviews of outcome measurement instruments (n = 100)

Characteristic	Reviews (n)
Construct of interest	
Patient-reported outcome	51
Non-patient-reported outcome	29
Patient- and non-patient-reported outcome	20
Level of health of the construct of interest*	
Biological and physiological variables	29
Symptom status	56
<i>Physical state</i>	42
<i>Emotional state</i>	19
<i>Cognitive/mental state</i>	13
Functional status	63
<i>Physical functioning</i>	48
<i>Social functioning</i>	22
<i>Cognitive/mental functioning</i>	21
<i>Role functioning</i>	9
General health perceptions	37
<i>Health-related quality of life</i>	36
<i>Self-rated health</i>	3
Overall quality of life	9
Age of the population of interest	
Children 0–18	20
Adults 18+	58
Both	22
Condition of the population of interest*	
Circulatory system	11
Congenital and genetic conditions	8
Digestive, endocrine and metabolic system	16
Ear, eye, and respiratory system	14
Genitourinary and reproductive system	9
Infections and parasites	5
Injuries and external causes	13
Mental and behavioral health	19
Musculoskeletal system	30
Neoplasms	15
Nervous system	16
Perinatal and pediatric health	4
Skin and subcutaneous tissue	10
Factors/symptoms influencing health status/contact with health services	42
OMI type included*	
Questionnaires	41
Clinical rating scales	41
Performance-based tests	24
Observations	7
Interviews	7
Imaging tests	6
Laboratory tests	6
Diaries	5
Videos	1

**Table 2** (continued)

Characteristic	Reviews (n)
Measurement properties considered per review, median [range]*	[1–9]
Content validity	64
Structural validity	46
Internal consistency	69
Cross-cultural validity/measurement invariance	37
Reliability	87
Measurement error	47
Construct validity	82
Criterion validity	50
Responsiveness	62

\*Multiple characteristics could be reported for the same review

in 18% of the review aims at least 2 key elements were not reported. Major flaws identified in the aim were often that the aim was unclear or vague, for example by stating that the aim was “to discuss validity” [121], or “to provide information about frailty instruments” [94].

### Search strategy

In 78% of the reviews the search strategy matched the research aim. When there was a mismatch between the aim and the search strategy, often the aim was to identify all available OMIs, whereas search terms for measurement properties were included. Hence, only OMIs with evidence for the measurement properties were identified.

Only 27% of the reviews had an appropriate search strategy with respect to search terms used for both the construct, population, OMI type and measurement properties. Search terms for OMI type were not appropriate for 40% of the reviews because relevant synonyms or search terms were not included. Search terms for measurement properties were deemed inappropriate for 34% of the reviews.

The number of databases searched ranged from 1–14, with a median of 4. MEDLINE was searched in 98% of the included reviews, whereas EMBASE was searched in 56%. Only 66% of the reviews performed reference checking of included articles.

### Eligibility criteria and article selection

In 75% of the reviews the eligibility criteria were clearly defined, and in 83% the eligibility criteria matched the research aim. Mismatches often concerned that the aim was to identify all available or used OMIs, whereas eligibility criteria included that the study should report on measurement properties, hence resulting in including

only OMIs that were validated to at least some extent. In 42% of the reviews other notable eligibility criteria were used, such as only including OMIs that were reported in at least a certain number of articles, only including validation studies of original OMIs or certain (language) versions, excluding studies of low quality, or excluding OMIs that were described in previously published systematic reviews.

In 65% and 69% of the reviews, respectively abstract and full text selection was (partly) done by at least 2 independent reviewers, compared to 41% and 38% in 2014. Data extraction was (partly) done by at least 2 independent reviewers in 42%, compared to 25% in both 2014 and 2007. In most other cases it was unclear whether two independent reviewers were involved.

### Risk of bias assessment

The methodological quality (i.e., risk of bias) of the studies was evaluated in 63% of the reviews, compared to 41% in 2014 and 30% in 2007. In 62% of those reviews, the quality assessment was done by at least two reviewers independently. For 33% of the reviews this was unclear.

### Measurement property evaluation

In 73% of the reviews (some) measurement properties of the included OMIs were evaluated, compared to 58% in 2014 and 55% in 2007. This means that in these reviews a judgement was made about the sufficiency of the measurement properties, rather than providing only the results of measurement properties. For those reviews in which (some) measurement properties were evaluated, (a reference to) criteria for measurement properties were provided in 81% of the reviews; in 19% of the reviews it was not clear on what criteria judgements were based. In those reviews in which measurement properties were evaluated and that included multidimensional OMIs, only 18% evaluated each subscale separately. In 22% of the reviews the evaluation of measurement properties was (partly) done by at least two independent reviewers.

### Data synthesis and certainty assessment

Data synthesis, in which results from multiple studies on the same OMI were combined, was (partly) performed in 60% of the reviews, compared to 44% in 2014 and 7% in 2007. In those reviews in which data synthesis was performed and that included multidimensional OMIs, synthesis was performed for each subscale separately in only 13% of the cases. Methods for data syntheses were clearly described in 47% of the reviews. In 84% of the reviews data synthesis was performed for each measurement property separately. Data

**Table 3** Quality appraisal of systematic reviews of outcome measurement instruments

Quality aspect	% Reviews (n= 100)
<b>Key elements</b>	
Key elements included in title	
Construct	80
Population	82
Type of OMI	66
Measurement properties	34
Systematic review	80
Key elements included in aim	
Construct	87
Population	81
Type of OMI	76
Measurement properties	76
<b>Search strategy</b>	
Search strategy matched aim	78
Search syntax for at least 1 database provided	70
Search appropriate for Construct	
Yes	49
Unclear	31
No	20
Population	
Yes	59
Unclear	21
No	20
Type of OMI	
Yes	48
Unclear	12
No	40
Measurement properties	
Yes	56
Unclear	10
No	34
Number of databases searched, median [range]	[1–14]
MEDLINE	98
EMBASE	56
Reference checking used	66
No time limits used in search or arguments provided for used of time limits	77
No language restrictions used in search	66
No other notable restrictions used in search	77
<b>Eligibility criteria</b>	
Inclusion and exclusion criteria clearly defined	75
Eligibility criteria matched aim	83
No other notably criteria used in eligibility	58
<b>Article selection</b>	
Abstract selection by at least 2 independent reviewers	
Yes	62
Partly	3
Unclear	26
No	9
Full-text selection by at least 2 independent reviewers	
Yes	67
Partly	2

**Table 3** (continued)

Quality aspect	% Reviews (n= 100)
Unclear	27
No	4
<b>Data extraction</b>	
Data extraction by at least 2 independent reviewers	
Yes	39
Partly	3
Unclear	44
No	14
<b>Risk of bias assessment</b>	
Methodological quality assessment of included studies	63
Methodological quality assessment by at least 2 independent reviewers	
Yes	62
Partly	1
Unclear	33
No	3
<b>Measurement property evaluation</b>	
Quality of the OMI (measurement properties) evaluated	
Yes	59
Some measurement properties	14
No	27
Criteria for measurement properties specified	
Yes	67
For some measurement properties	14
No	19
Evaluation of each subscale (if multidimensional)	
Yes	18
Partly	5
Unclear	26
No	51
Measurement properties evaluated by at least 2 independent reviewers	
Yes	21
Partly	1
Unclear	70
No	8
<b>Data synthesis</b>	
Data synthesis performed (if possible)	
Yes	31
Partly	57
Unclear	3
No	8
Data synthesis performed for each subscale (if multidimensional)	
Yes	13
Unclear	50
No	37
Data synthesis methods clearly described	47
Data synthesis performed at the level of	
Measurement properties	84
Only domains of measurement properties	13
Only subscales or instruments	4
Data synthesis performed by at least 2 independent reviewers	
Yes	18

**Table 3** (continued)

Quality aspect	% Reviews (n= 100)
Unclear	75
No	7
<b>Certainty assessment</b>	
Quality of the evidence graded	33
Quality of the evidence graded for each subscale (if multidimensional)	
Yes	15
Unclear	19
No	67
Quality of the evidence graded by at least 2 independent reviewers	
Yes	27
Unclear	70
No	3
<b>Presentation of results</b>	
Flow chart provided	96
Reasons for excluding full text articles reported	
Full information (numbers for each reason)	65
Some information (reasons, but not specifying numbers)	20
No	15
Included instruments in accordance with inclusion criteria	
Yes	86
Unclear	12
No	2
Results of measurement properties reported as raw data	
Yes	42
For some measurement properties	30
No	28
<b>Instrument recommendation</b>	
Recommendations for instruments made	42
Recommendations made for each construct of interest	25
OMI recommendation consistent with evidence appraisal	
Yes	55
Partly	7
Unclear	24
No	14

synthesis was performed by at least 2 independent reviewers for 18% of the reviews.

In 33% of the reviews, a certainty assessment was done in which the quality of the evidence was graded. Quality of the evidence was graded by at least 2 independent reviewers in 27% of the reviews.

### Results and instrument recommendation

A flowchart was provided in 96% of the reviews, often with reasons for excluding full texts (85% vs. 55% in 2014). Included instruments were in 86% of the reviews in accordance with the inclusion criteria. In 72% of the reviews, the results of (some) measurement properties were reported as raw data.

In almost half of the reviews (42%) recommendations on which instrument (not) to use were made. In 25% of the reviews, recommendations were made for each construct of interest. In 62% of the reviews the recommendations made were consistent with the evidence appraisal.








A summary of the main results with recommendations for future OMI systematic reviews is provided in Table 4.

### Discussion

This overview of reviews aimed to investigate whether the quality of recent systematic reviews of OMIs lives up to the current scientific standards and which aspects have notably improved over time. Compared to previous studies [9, 10],



**Table 4** Overview of main findings and recommendations for future OMI systematic reviews

Features of high-quality systematic reviews of outcome measurement instruments	Main findings	Recommendations
 Well-defined research question/aim	Over half of the reviews had an unclear research question or aim	Formulate a research question/aim based on the four key elements: construct, population, (type of) instrument, measurement properties
 Comprehensive search strategy	Three-quarters of the reviews had an inappropriate or incomprehensive search strategy	If search terms for (type of) instrument and measurement properties are needed, use validated search filters (e.g., [21, 23])
 Independent abstract and full-text selection	For one-third of the reviews it was unclear whether article selection was done independently	Independently select articles and report how article selection took place
 Risk of bias assessment of included studies	One-third of the reviews had no risk of bias assessment of included studies	Conduct a risk of bias assessment using an appropriate tool, for example the COSMIN risk of bias checklist [16] or tool [15]
 Measurement property evaluation	In a quarter of the reviews no measurement properties were evaluated	Evaluate the measurement properties of included instruments using established criteria, for example criteria for good measurement properties [4]
 Synthesis of results, including instrument recommendations	In almost half of the reviews evidence from multiple studies for the same measurement property and instrument were not synthesized; recommendations on what instrument (not) to use were made in less than half of the reviews	Synthesize study results on the same measurement property of the same instrument and make recommendations on what instrument (not) to use
 Certainty assessment of the body of evidence	Two-thirds of the review did not include a certainty assessment of the body of evidence	Make a certainty assessment of the body of evidence using for example the (modified) GRADE system [124]

we found marked improvements in the conduct of risk of bias assessments, evaluation of measurement properties, and performance of formal data syntheses. Despite this, further improvements in these areas are necessary, as well as with respect to the research question and search strategy.

Over half of the reviews included in this study had an unclear research question or aim, for example with respect to the population of interest, the measurement properties that were evaluated, or the type of OMIs that were included. Including the four key elements, analogue to the PICO (population, intervention, comparison, outcome) format in systematic reviews of interventions [4, 8, 125], helps to formulate a well-defined research question and facilitates the development of an appropriate search strategy. Without a clear research question, it is not possible to assess the comprehensiveness of the search strategy.

Almost three-quarters of the reviews had an inappropriate or incomprehensive search strategy, often because inappropriate search terms for OMI type or measurement properties were included. It is preferred not to use search terms for OMI type to avoid missing any studies; however, if search terms are needed because of too many results, a search filter exists for PROMs [23]. A highly sensitive search filter also exists for measurement properties [21], but it was used in only 14 reviews. While searching both MEDLINE and EMBASE is recommended as a minimum by Cochrane [126], almost half of the reviews included in this study did not search EMBASE. Similarly, whilst reference checking is recommended [126], this was not reported by a third of the reviews. Through reference checking, one can also confirm the comprehensiveness of the search strategy: if many relevant articles were found through reference checking, the search was probably not comprehensive and important studies may have been missed [126].

In almost half of the reviews poorly justified eligibility criteria were used, e.g., only including OMIs in a certain language, or excluding OMIs that were included in previous systematic reviews. Such unintuitive eligibility criteria might negatively impact the inclusion of relevant studies or OMIs, hampering a complete synthesis of the body of available evidence. The number of reviews in which article selection and data extraction was conducted by at least 2 independent reviewers increased compared to previous overviews [9, 10].

Whilst a marked increase in the number of reviews that included a risk of bias assessment was found (63% currently compared to 41% in 2014 and 30% in 2007 [9, 10]), opportunities for improvement remain. Evaluating risk of bias in empirical studies on measurement properties is important, because results might not be valid if a study has bias. For example, relevant items might be missing in a PROM if patients were not involved in its development,

or the reliability of an OMI might be underestimated if the time interval between test and retest is (too) long. The COSMIN risk of bias checklist [16] or tool [15] were specifically developed for this purpose and were used in 47 reviews. Other risk of bias tools reported in the reviews [43, 70, 82, 101, 120] included, for example, the QUADAS-2 [127], QAREL [128], ROBINS-I [129], and Newcastle–Ottawa quality assessment scale [130]. These tools are, however, not specifically developed to assess the methodological quality of empirical measurement property studies and may not identify important bias.

The number of reviews in which measurement properties were formally evaluated has notably increased since 2007 (73% currently compared to 58% in 2014 and 55% in 2007 [9, 10]). In 14 reviews, however, it was not clear which criteria were used. In several reviews, authors mistakenly used risk of bias or certainty assessment ratings as a measure of OMI quality. However, these ratings refer to the quality of the study and the quality of the evidence, respectively, and not to the quality of the OMI (i.e., its measurement properties).

A clear increase in the number of reviews in which a data synthesis was performed was also observed (60% currently compared to 42% in 2014 and 7% in 2007 [9, 10]). However, the methods for data synthesis were often unclearly described and only in a third of the reviews a certainty assessment of the body of evidence was conducted. Potentially, the publication of the COSMIN guideline for systematic reviews of PROMs [4] in 2018 increased the number of reviews in which a data synthesis was performed. This guideline details how to synthesize multiple studies on the same measurement property for the same OMI, although more guidance might be necessary.

Each subscale in a multidimensional instrument should be considered a separate instrument as it represents a unique construct with measurement properties often varying between subscales [4]. However, we observed that few studies separately evaluated measurement properties or conducted an evidence synthesis at the subscale level. By not evaluating each subscale separately, a review therefore presents an incomplete picture of the measurement properties for the given scale.

Less than half of the reviews made recommendations about which OMI (not) to use. The conclusions of systematic reviews will be used by other researchers and clinicians who need to select an OMI for their purpose, although the selection of the most appropriate OMI may depend on the context and situation. Clear, evidence-based recommendations on which OMI (not) to use will help others in their OMI selection and contribute to the standardization of OMIs.

Although two-thirds of the reviews purport to include an evaluation of content validity, there is doubt over the

thoroughness of these evaluations. Whilst 25 reviews reported application of the COSMIN guideline for evaluating content validity, only 13 appear to have applied it correctly. One of the steps in the assessment of content validity according to the COSMIN guideline is the evaluation of the content by reviewers themselves. This step was often lacking. Other flaws included not distinguishing between development and content validity studies, and only conducting a risk of bias assessment without evaluating the content validity of the OMI.

Other major flaws that we observed in some reviews were confusing the quality of the study (i.e., risk of bias) with the quality of the OMI (i.e., its measurement properties) or making recommendations based on certainty assessment rather than the sufficiency of measurement properties.

### Towards high quality OMI systematic reviews

Systematic reviews of OMIs are difficult to conduct, and this study shows that the availability of methodology and tools that guide authors in the conduct of their systematic review does not translate automatically into high-quality systematic reviews. Besides more and better resources, behavioral change techniques [131], implementation strategies, and knowledge translation activities are needed to improve systematic review quality. Several of these have recently been developed or are being considered. First, the COSMIN guideline for systematic reviews has recently been updated and made more user-friendly to better facilitate reviewers [132]. Second, a newly developed animated video explains the key steps of conducting a systematic review of OMIs (available at <https://www.cosmin.nl/>). Third, a reporting guideline for OMI systematic reviews has recently been developed [133], and knowledge translation activities have been implemented to increase its uptake. Last, a course on how to conduct OMI systematic reviews is being developed to educate reviewers more thoroughly. To alert systematic review authors to the various tools available, an automated email can be sent to authors registering their review in PROSPERO. PROSPERO is a database for registering systematic reviews of health related outcomes [134], and although less than half of the included reviews reported prospective registration, such an email alert might increase the uptake of tools and improve the quality of future OMI systematic reviews.

### Limitations

An important limitation of the current study is the potential subjectivity in appraising the quality of systematic reviews. We attempted to use a rigorous and standardized data extraction process, in which we pilot tested and improved the data extraction form, provided training to reviewers

who were already experts in systematic reviews of OMIs, and assigned systematic reviews to reviewer pairs who independently appraised their quality and reached consensus about any discrepancies. However, because of large variations in the systematic reviews included, some degree and variation of subjective judgement in appraising the quality of systematic reviews could not be avoided. Second, some of the included reviews might not have been systematic reviews by definition, as the inclusion criteria were not stringent in that respect. We decided to include a review if at least one measurement property was evaluated (i.e., some degree of judgement was made about the sufficiency of a measurement property, as opposed to only providing an overview of the measurement properties). Third, we were unable to compare all quality aspects historically, because not all aspects were rated in the studies conducted in 2014 and 2007 [9, 10]. Compared to the previous studies, the current appraisal is the most comprehensive, and new elements were added, such as inclusion of key elements in the title, specification of criteria for measurement properties, evaluation of subscales, and assessment of certainty. Fourth, we randomly selected 100 recent reviews that fulfilled the eligibility criteria, out of a set of 136 reviews that were identified while updating the COSMIN database [20]. Our aim was not to include all available systematic reviews but rather to appraise and compare the quality of a random sample of the most recently published reviews with a set of reviews published respectively 8 and 15 years ago. We believe that the inclusion of additional reviews would not have altered our findings. Lastly, the appraisal of the reviews' quality was hampered by poor reporting, for example with respect to the process of data synthesis or the number of independent reviewers involved in each of the steps of the review process. The recently developed PRISMA-COSMIN for OMIs reporting guideline could improve the reporting of OMI systematic reviews [133]. Although the current study is not a one-to-one baseline assessment of reporting aspects required by PRISMA-COSMIN for OMIs, most reporting items have been included in the current quality appraisal. Because our aim was to assess whether the quality of recent systematic reviews lived up to the current scientific standards, including reporting quality, we have not contacted the authors of the included systematic reviews to provide additional information.

### Conclusion

In conclusion, this overview of 100 reviews published after June 2021 found, compared to previous overviews of reviews, a clear improvement in the number of OMI systematic reviews that conducted a risk of bias assessment, evaluated the measurement

properties of included OMIs, and conducted a data synthesis. However, room for improvement in these areas remains. Improvements regarding the research question and search strategy are urgently needed, as more than half of the reviews likely missed important studies. To ensure that systematic reviews of OMIs meet current scientific standards, more consistent conduct and reporting of systematic reviews of OMIs is needed.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11136-024-03706-z>.

**Author contributions** CB conceptualized and designed the study. Material preparation was performed by EE. Data extraction was performed by EE, LM, OA, AC, KH, KMS, DO, JP, ISM, and CB. Data analysis was performed by EE. The first draft of the manuscript was written by EE and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

**Data availability** All data supporting the findings of this study are available within the paper and its Supplementary Information.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose. Drs. Terwee and Mokkink are the founders of COSMIN.

**Ethical approval** This is a methodological study for which no ethical approval is required. There were no individual participants included in the study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Devlin, N. J., & Appleby, J. (2010). *Getting the most out of PROMs. Putting health outcomes at the heart of NHS decision-making*. The King's Fund.
- Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*. <https://doi.org/10.3310/hta2140>
- Greenhalgh, J. (2009). The applications of PROs in clinical practice: What are they, do they work, and why? *Quality of Life Research*, *18*, 115–123.
- Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157.
- Mokkink, L. B., Prinsen, C. A., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based standards for the selection of health measurement instruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, *20*, 105–113.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745.
- Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a “core outcome set”—a practical guideline. *Trials*, *17*(1), 1–10.
- Higgins Jpt, T., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V., (2022) *Cochrane Handbook for systematic reviews of interventions version 6.3 (updated February 2022)*, Cochrane.
- Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research*, *18*(3), 313–333.
- Terwee, C. B., Prinsen, C., Garotti, M. R., Suman, A., De Vet, H., & Mokkink, L. B. (2016). The quality of systematic reviews of health-related outcome measurement instruments. *Quality of Life Research*, *25*(4), 767–779.
- McKenna, S. P., & Heaney, A. (2021). Setting and maintaining standards for patient-reported outcome measures: Can we rely on the COSMIN checklists? *Journal of Medical Economics*, *24*(1), 502–511.
- OMERACT. (2021). *The OMERACT handbook for establishing and implementing core outcomes in clinical trials across the spectrum of rheumatologic conditions*. OMERACT.
- Aromataris, E. & Munn, Z. (2020). *JBI manual for evidence synthesis*. 2020: JBI.
- COSMIN (2023). Retrieved, [www.cosmin.nl](http://www.cosmin.nl).
- Mokkink, L. B., Boers, M., van der Vleuten, C., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2020). 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*, *20*(1), 1–13.
- Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1171–1179.
- Aromataris, E. & Munn, Z., *JBI Manual for Evidence Synthesis*. 2020, JBI.
- Elsman, E. B., Mokkink, L. B., Butcher, N. J., Offringa, M., & Terwee, C. B. (2022). *The reporting and methodological quality of systematic reviews of health-related outcome measurement instruments*. PROSPERO.
- Gates, M., Gates, A., Pieper, D., Fernandes, R. M., Tricco, A. C., Moher, D., Brennan, S. E., Li, T., Pollock, M., Lunny, C., Sepulveda, D., McKenzie, J. E., Scott, S. D., Robinson, K. A., Matthias, K., Bougioukas, K. I., Fusar-Poli, P., Whiting, P., Moss, S. J., & Hartling, L. (2022). Reporting guideline for overviews of reviews of healthcare interventions: Development of the PRIOR statement. *BMJ*, *378*, e070849.
- COSMIN (2022) *COSMIN database of systematic reviews*. Retrieved 30, Jan, 2022, from [www.cosmin.nl/tools/database-systematic-reviews/](http://www.cosmin.nl/tools/database-systematic-reviews/).

21. Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. (2009). Development of a methodological pubmed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, *18*, 1115–1123.
22. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *JAMA*, *273*(1), 59–65.
23. Mackintosh, A., Casañas i Comabella, C., Hadi, M., Gibbons, E., Fitzpatrick, R., & Roberts, N. (2010). *PROM GROUP CONSTRUCT & INSTRUMENT TYPE FILTERS*. Retrieved 2023, May, from <https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf>.
24. Alamrani, S., Rushton, A. B., Gardner, A., Bini, E., Falla, D., & Heneghan, N. R. (2021). Physical functioning in adolescents with idiopathic scoliosis: A systematic review of outcome measures and their measurement properties. *Spine*, *46*(18), E985–E997.
25. Barni, L., Ruiz-Munoz, M., Gonzalez-Sanchez, M., Cuesta-Vargas, A. I., Merchan-Baeza, J., & Freddolini, M. (2021). Psychometric analysis of the questionnaires for the assessment of upper limbs available in their Italian version: A systematic review of the structural and psychometric characteristics. *Health and Quality of Life Outcomes*, *19*(1), 259.
26. Meilani, E., Zanudin, A., & Mohd Nordin, N. A. (2022). Psychometric properties of quality of life questionnaires for patients with breast cancer-related lymphedema: A systematic review. *International Journal of Environmental Research and Public Health*, *19*(5), 2519.
27. Proctor, K. B., Tison, K. H., Park, J., Scahill, L., Vickery, B. P., & Sharp, W. G. (2022). A systematic review of parent report measures assessing the psychosocial impact of food allergy on patients and families. *Allergy*, *77*(5), 1347–1359.
28. Oluchi, S. E., Manaf, R. A., Ismail, S., Kadir Shahar, H., Mahmud, A., & Udeani, T. K. (2021). Health related quality of life measurements for diabetes: A systematic review. *International Journal of Environmental Research and Public Health*, *18*(17), 9245.
29. Sellitto, G., Morelli, A., Bassano, S., Conte, A., Baione, V., Galeoto, G., & Berardi, A. (2021). Outcome measures for physical fatigue in individuals with multiple sclerosis: A systematic review. *Expert Review of Pharmacoeconomics & Outcomes Research*, *21*(4), 625–646.
30. Ai, X., Yang, X., Fu, H. Y., Xu, J. M., & Tang, Y. M. (2022). Health-related quality of life questionnaires used in primary biliary cholangitis: A systematic review. *Scandinavian Journal of Gastroenterology*, *57*(3), 333–339.
31. Alaqeel, S., Alfakhri, A., Alkherb, Z., & Almeshal, N. (2022). Patient-reported outcome measures in Arabic-speaking populations: A systematic review. *Quality of Life Research*, *31*(5), 1309–1320.
32. Arts, J., Gubbels, J. S., Verhoeff, A. P., Chinapaw, M. J. M., Lettink, A., & Altenburg, T. M. (2022). A systematic review of proxy-report questionnaires assessing physical activity, sedentary behavior and/or sleep in young children (aged 0–5 years). *International Journal of Behavioral Nutrition and Physical Activity*, *19*(1), 18.
33. Baamer, R. M., Iqbal, A., Lobo, D. N., Knaggs, R. D., Levy, N. A., & Toh, L. S. (2022). Utility of unidimensional and functional pain assessment tools in adult postoperative patients: A systematic review. *British Journal of Anaesthesia*, *128*(5), 874–888.
34. Baqays, A., Zenke, J., Campbell, S., Johannsen, W., Rashid, M., Seikaly, H., & El-Hakim, H. (2021). Systematic review of validated parent-reported questionnaires assessing swallowing dysfunction in otherwise healthy infants and toddlers. *Journal of Otolaryngology—Head & Neck Surgery*, *50*(1), 68.
35. Bawua, L. K., Miaskowski, C., Hu, X., Rodway, G. W., & Pelter, M. M. (2021). A review of the literature on the accuracy, strengths, and limitations of visual, thoracic impedance, and electrocardiographic methods used to measure respiratory rate in hospitalized patients. *Annals of Noninvasive Electrocardiology*, *26*(5), e12885.
36. Beshara, P., Anderson, D. B., Pelletier, M., & Walsh, W. R. (2021). The reliability of the microsoft kinect and ambulatory sensor-based motion tracking devices to measure shoulder range-of-motion: a systematic review and meta-analysis. *Sensors (Basel)*, *21*(24), 8186.
37. Bik, A., Sam, C., de Groot, E. R., Visser, S. S. M., Wang, X., Tataranno, M. L., Benders, M., van den Hoogen, A., & Dudink, J. (2022). A scoping review of behavioral sleep stage classification methods for preterm infants. *Sleep Medicine*, *90*, 74–82.
38. Blanco-Orive, P., del Corral, T., Martín-Casas, P., Ceniza-Bordallo, G., & López-de-Uralde-Villanueva, I. (2022). Quality of life and exercise tolerance tools in children/adolescents with cystic fibrosis: Systematic review. *Medicina Clínica*, *158*(11), 519–530.
39. Bortolani, S., Brusa, C., Rolle, E., Monforte, M., De Arcangelis, V., Ricci, E., Mongini, T. E., & Tasca, G. (2022). Technology outcome measures in neuromuscular disorders: A systematic review. *European Journal of Neurology*, *29*(4), 1266–1278.
40. Bru-Luna, L. M., Marti-Vilar, M., Merino-Soto, C., & Cervera-Santiago, J. L. (2021). Emotional intelligence measures: A systematic review. *Healthcare*, *9*(12), 1696.
41. Bruyneel, A. V., & Dube, F. (2021). Best quantitative tools for assessing static and dynamic standing balance after stroke: A systematic review. *Physiotherapy Canada*, *73*(4), 329–340.
42. Buck, B., Gagen, E. C., Halverson, T. F., Nagendra, A., Ludwig, K. A., & Fortney, J. C. (2022). A systematic search and critical review of studies evaluating psychometric properties of patient-reported outcome measures for schizophrenia. *Journal of Psychiatric Research*, *147*, 13–23.
43. Chamorro-Moriana, G., Perez-Cabezas, V., Espuny-Ruiz, F., Torres-Enamorado, D., & Ridao-Fernandez, C. (2022). Assessing knee functionality: Systematic review of validated outcome measures. *Annals of Physical and Rehabilitation Medicine*, *65*(6), 101608.
44. Chen, B. S., Galus, T., Archer, S., Tadic, V., Horton, M., Pesudovs, K., Braithwaite, T., & Yu-Wai-Man, P. (2022). Capturing the experiences of patients with inherited optic neuropathies: A systematic review of patient-reported outcome measures (PROMs) and qualitative studies. *Graefes Archive for Clinical and Experimental Ophthalmology*, *260*(6), 2045–2055.
45. Collins, K. C., Burdall, O., Kassam, J., Firth, G., Perry, D., & Ramachandran, M. (2022). Health-related quality of life and functional outcome measures for pediatric multiple injury: A systematic review and narrative synthesis. *Journal of Trauma and Acute Care Surgery*, *92*(5), e92–e106.
46. Cuenca-Garcia, M., Marin-Jimenez, N., Perez-Bey, A., Sanchez-Oliva, D., Camiletti-Moiron, D., Alvarez-Gallardo, I. C., Ortega, F. B., & Castro-Pinero, J. (2022). Reliability of field-based fitness tests in adults: A systematic review. *Sports Medicine*, *52*(8), 1961–1979.
47. Daskalakis, I., Sperlakis, I., Sidiropoulou, B., Kontakis, G., & Tosounidis, T. (2021). Patient-reported outcome measures (PROMs) relevant to musculoskeletal conditions translated and validated in the greek language: A COSMIN-based systematic review of measurement properties. *Mediterr J Rheumatol*, *32*(3), 200–217.
48. de Assis Brasil, M. L., Zakhour, S., Figueira, G. L., Pires, P. P., Nardi, A. E., & Sardinha, A. (2022). Sexuality assessment of the Brazilian population: An integrative review of the available instruments. *Journal of Sex and Marital Therapy*, *48*(8), 757–774.
49. de Oliveira, M. P. B., da Silva Serrao, P. R. M., de Medeiros Takahashi, A. C., Pereira, N. D., & de Andrade, L. P. (2022). Reproducibility of assessment tests addressing body structure

- and function and activity in older adults with dementia: A systematic review. *Physical Therapy*. <https://doi.org/10.1093/ptj/pzab263>
50. de Witte, M., Kooijmans, R., Hermanns, M., van Hooren, S., Biesmans, K., Hermsen, M., Stams, G. J., & Moonen, X. (2021). Self-report stress measures to assess stress in adults with mild intellectual disabilities—a scoping review. *Frontiers in Psychology*, *12*, 742566.
  51. Essiet, I. A., Lander, N. J., Salmon, J., Duncan, M. J., Eyre, E. L. J., Ma, J., & Barnett, L. M. (2021). A systematic review of tools designed for teacher proxy-report of children's physical literacy or constituting elements. *International Journal of Behavioral Nutrition and Physical Activity*, *18*(1), 131.
  52. Etkin, R. G., Shimshoni, Y., Lebowitz, E. R., & Silverman, W. K. (2021). Using evaluative criteria to review youth anxiety measures, part I: self-report. *Journal of Clinical Child and Adolescent Psychology*, *50*(1), 58–76.
  53. Fan, Y., Shu, X., Leung, K. C. M., & Lo, E. C. M. (2021). Patient-reported outcome measures for masticatory function in adults: A systematic review. *BMC Oral Health*, *21*(1), 603.
  54. Galan-Gonzalez, E., Martinez-Perez, G., & Gascon-Catalan, A. (2021). Family functioning assessment instruments in adults with a non-psychiatric chronic disease: a systematic review. *Nurs Rep*, *11*(2), 341–355.
  55. Gallo, L., Gallo, M., Augustine, H., Leveille, C., Murphy, J., Copeland, A. E., & Thoma, A. (2022). Assessing patient frailty in plastic surgery: A systematic review. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, *75*(2), 579–585.
  56. Gard, E., Lyman, A., & Garg, H. (2022). Perinatal incontinence assessment tools: A psychometric evaluation and scoping review. *Journal of Women's Health (2002)*, *31*(8), 1208–1218.
  57. Giga, L., Petersone, A., Cakstina, S., & Berzina, G. (2021). Comparison of content and psychometric properties for assessment tools used for brain tumor patients: A scoping review. *Health and Quality of Life Outcomes*, *19*(1), 234.
  58. Goodney, P., Shah, S., Hu, Y. D., Suckow, B., Kinlay, S., Armstrong, D. G., Geraghty, P., Patterson, M., Menard, M., Patel, M. R., & Conte, M. S. (2022). A systematic review of patient-reported outcome measures patients with chronic limb-threatening ischemia. *Journal of Vascular Surgery*, *75*(5), 1762–1775.
  59. Guallar-Bouloc, M., Gomez-Bueno, P., Gonzalez-Sanchez, M., Molina-Torres, G., Lomas-Vega, R., & Galan-Mercant, A. (2021). Spanish questionnaires for the assessment of pelvic floor dysfunctions in women: A systematic review of the structural characteristics and psychometric properties. *International Journal of Environmental Research and Public Health*, *18*(23), 12858.
  60. Gutierrez-Vilahu, L., & Guerra-Balic, M. (2021). Footprint measurement methods for the assessment and classification of foot types in subjects with down syndrome: A systematic review. *Journal of Orthopaedic Surgery and Research*, *16*(1), 537.
  61. Halvorsen, M. B., Helvershou, S. B., Axelsdottir, B., Brondbo, P. H., & Martinussen, M. (2023). General measurement tools for assessing mental health problems among children and adolescents with an intellectual disability: A systematic review. *Journal of Autism and Developmental Disorders*, *53*(1), 132–204.
  62. Huang, E. Y., & Lam, S. C. (2021). Review of frailty measurement of older people: Evaluation of the conceptualization, included domains, psychometric properties, and applicability. *Aging Med (Milton)*, *4*(4), 272–291.
  63. Kitamura, K., van Hooff, M., Jacobs, W., Watanabe, K., & de Kleuver, M. (2022). Which frailty scales for patients with adult spinal deformity are feasible and adequate? A Systematic Review. *Spine*, *22*(7), 1191–1204.
  64. Kusi-Mensah, K., Nuamah, N. D., Wemakor, S., Agorinya, J., Seidu, R., Martyn-Dickens, C., & Bateman, A. (2022). Assessment tools for executive function and adaptive function following brain pathology among children in developing country contexts: A scoping review of current tools. *Neuropsychology Review*, *32*(3), 459–482.
  65. Lambert, K., & Stanford, J. (2022). Patient-reported outcome and experience measures administered by dietitians in the outpatient setting: Systematic review. *Canadian Journal of Dietetic Practice and Research*, *83*(1), 11.
  66. Le, C. Y., Truong, L. K., Holt, C. J., Filbay, S. R., Dennett, L., Johnson, J. A., Emery, C. A., & Whittaker, J. L. (2021). Searching for the holy grail: A systematic review of health-related quality of life measures for active youth. *Journal of Orthopaedic & Sports Physical Therapy*, *51*(10), 478–491.
  67. Lear, M. K., Lee, E. B., Smith, S. M., & Luoma, J. B. (2022). A systematic review of self-report measures of generalized shame. *Journal of Clinical Psychology*, *78*(7), 1288–1330.
  68. Levasseur, M., Lussier-Therrien, M., Biron, M. L., Dubois, M. F., Boissy, P., Naud, D., Dubuc, N., Coallier, J. C., Calve, J., & Audet, M. (2022). Scoping study of definitions and instruments measuring vulnerability in older adults. *Journal of the American Geriatrics Society*, *70*(1), 269–280.
  69. Lim-Watson, M. Z., Hays, R. D., Kingsberg, S., Kallich, J. D., & Murimi-Worstell, I. B. (2022). A systematic literature review of health-related quality of life measures for women with hypoactive sexual desire disorder and female sexual interest/arousal disorder. *Sexual Medicine Reviews*, *10*(1), 23–41.
  70. Liu, X., McNally, T. W., Beese, S., Downie, L. E., Solebo, A. L., Faes, L., Husain, S., Keane, P. A., Moore, D. J., & Denniston, A. K. (2021). Non-invasive instrument-based tests for quantifying anterior chamber flare in uveitis: A systematic review. *Ocular Immunology and Inflammation*, *29*(5), 982–990.
  71. Lodge, M. E., Moran, C., Sutton, A. D. J., Lee, H. C., Dhesi, J. K., Andrew, N. E., Ayton, D. R., Hunter-Smith, D. J., Srikanth, V. K., & Snowdon, D. A. (2022). Patient-reported outcome measures to evaluate postoperative quality of life in patients undergoing elective abdominal surgery: A systematic review. *Quality of Life Research*, *31*(8), 2267–2279.
  72. Long, M., Stansfeld, J. L., Davies, N., Crellin, N. E., & Moncrieff, J. (2022). A systematic review of social functioning outcome measures in schizophrenia with a focus on suitability for intervention research. *Schizophrenia Research*, *241*, 275–291.
  73. Luck-Sikorski, C., Rossmann, P., Topp, J., Augustin, M., Sommer, R., & Weinberger, N. A. (2022). Assessment of stigma related to visible skin diseases: A systematic review and evaluation of patient-reported outcome measures. *Journal of the European Academy of Dermatology and Venereology*, *36*(4), 499–525.
  74. Madso, K. G., Flo-Groeneboom, E., Pachana, N. A., & Nordhus, I. H. (2021). Assessing momentary well-being in people living with dementia: A systematic review of observational instruments. *Frontiers in Psychology*, *12*, 742510.
  75. Marshall, A. N., Root, H. J., Valovich McLeod, T. C., & Lam, K. C. (2022). Patient-reported outcome measures for pediatric patients with sport-related injuries: A systematic review. *Journal of Athletic Training*, *57*(4), 371–384.
  76. Marsico, P., Meier, L., van der Linden, M. L., Mercer, T. H., & van Hedel, H. J. A. (2022). Psychometric properties of lower limb somatosensory function and body awareness outcome measures in children with upper motor neuron lesions: A systematic review. *Developmental Neurorehabilitation*, *25*(5), 314–327.
  77. Master, H., Coleman, G., Dobson, F., Bennell, K., Hinman, R. S., Jakiela, J. T., & White, D. K. (2021). A narrative review on measurement properties of fixed-distance walk tests up to 40 meters for adults with knee osteoarthritis. *Journal of Rheumatology*, *48*(5), 638–647.
  78. Menegol, N. A., Ribeiro, S. N. S., Okubo, R., Gulonda, A., Sonza, A., Montemezzo, D., & Sanada, L. S. (2022). Quality












- assessment of neonatal pain scales translated and validated to brazilian portuguese: A systematic review of psychometric properties. *Pain Management Nursing*, 23(4), 559–565.
79. Mlynczyk, J., Abramowicz, P., Stawicki, M. K., & Konstantynowicz, J. (2022). Non-disease specific patient-reported outcome measures of health-related quality of life in juvenile idiopathic arthritis: A systematic review of current research and practice. *Rheumatology International*, 42(2), 191–203.
  80. Nascimento, D., Carmona, J., Mestre, T., Ferreira, J. J., & Guimaraes, I. (2021). Drooling rating scales in Parkinson's disease: A systematic review. *Parkinsonism & Related Disorders*, 91, 173–180.
  81. Ngwayi, J. R. M., Tan, J., Liang, N., Sita, E. G. E., Obie, K. U., & Porter, D. E. (2021). Systematic review and standardised assessment of Chinese cross-cultural adapted hip patient reported outcome measures (PROMs). *PLoS ONE*, 16(9), e0257081.
  82. Nolet, P. S., Yu, H., Cote, P., Meyer, A. L., Kristman, V. L., Sutton, D., Murnaghan, K., & Lemeunier, N. (2021). Reliability and validity of manual palpation for the assessment of patients with low back pain: A systematic and critical review. *Chiropr Man Therap*, 29(1), 33.
  83. Orth, Z., & van Wyk, B. (2021). Measuring mental wellness among adolescents living with a physical chronic condition: A systematic review of the mental health and mental well-being instruments. *BMC Psychol*, 9(1), 176.
  84. Overman, M. J., Leeworthy, S., & Welsh, T. J. (2021). Estimating premorbid intelligence in people living with dementia: A systematic review. *International Psychogeriatrics*, 33(11), 1145–1159.
  85. Parati, M., Ambrosini, E. B., Gallotta, M., Dalla Vecchia, L. A., Ferriero, G., & Ferrante, S. (2022). The reliability of gait parameters captured via instrumented walkways: a systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine*, 58(3), 363–377.
  86. Pereira, T., & Lousada, M. (2023). Psychometric properties of standardized instruments that are used to measure pragmatic intervention effects in children with developmental language disorder: A systematic review. *Journal of Autism and Developmental Disorders*, 53(5), 1764–1780.
  87. Phillips, S. M., Summerbell, C., Ball, H. L., Hesketh, K. R., Saxena, S., & Hillier-Brown, F. C. (2021). The validity, reliability, and feasibility of measurement tools used to assess sleep of pre-school aged children: A systematic rapid review. *Frontiers in Pediatrics*, 9, 770262.
  88. Piscitelli, D., Ferrarello, F., Ugolini, A., Verola, S., & Pellicciari, L. (2021). Measurement properties of the gross motor function classification system, gross motor function classification system-expanded & revised, manual ability classification system, and communication function classification system in cerebral palsy: A systematic review with meta-analysis. *Developmental Medicine and Child Neurology*, 63(11), 1251–1261.
  89. Pizzinato, A., Liguoro, I., Pusiolo, A., Cogo, P., Palese, A., & Vidal, E. (2022). Detection and assessment of postoperative pain in children with cognitive impairment: A systematic literature review and meta-analysis. *European Journal of Pain*, 26(5), 965–979.
  90. Ploumen, R. L. M., Willems, S. H., Jonkman, R. E. G., Nolte, J. W., & Becking, A. G. (2023). Quality of life after orthognathic surgery in patients with cleft: An overview of available patient-reported outcome measures. *Cleft Palate-Craniofacial Journal*, 60(4), 405–412.
  91. Qian, M., Shi, Y., Lv, J., & Yu, M. (2021). Instruments to assess self-neglect among older adults: A systematic review of measurement properties. *International Journal of Nursing Studies*, 123, 104070.
  92. Ramanathan, D., Chu, S., Prendes, M., & Carroll, B. T. (2021). Validated outcome measures and postsurgical scar assessment instruments in eyelid surgery: A systematic review. *Dermatologic Surgery*, 47(7), 914–920.
  93. Ratti, M. M., Gandaglia, G., Alleva, E., Leardini, L., Sisca, E. S., Der-evianko, A., Furnari, F., Mazzoleni Ferracini, S., Beyer, K., Moss, C., Pellegrino, F., Sorce, G., Barletta, F., Scuderi, S., Omar, M. I., MacLennan, S., Williamson, P. R., Zong, J., MacLennan, S. J., ... Cornford, P. (2022). Standardising the assessment of patient-reported outcome measures in localised prostate cancer A Systematic Review. *European Urology Oncology*, 5(2), 153–163.
  94. Roopsawang, I., Zaslavsky, O., Thompson, H., Aree-Ue, S., Kwan, R. Y. C., & Belza, B. (2022). Frailty measurements in hospitalised orthopaedic populations age 65 and older: A scoping review. *Journal of Clinical Nursing*, 31(9–10), 1149–1163.
  95. Roos, M., Dagenais, M., Pflieger, S., & Roy, J. S. (2022). Patient-reported outcome measures of musculoskeletal symptoms and psychosocial factors in musicians: A systematic review of psychometric properties. *Quality of Life Research*, 31(9), 2547–2566.
  96. Roth, E. M., Lubitz, C. C., Swan, J. S., & James, B. C. (2020). Patient-reported quality-of-life outcome measures in the thyroid cancer population. *Thyroid*, 30(10), 1414–1431.
  97. Rymer, J. A., Narcisse, D., Cosiano, M., Tanaka, J., McDermott, M. M., Treat-Jacobson, D. J., Conte, M. S., Tuttle, B., Patel, M. R., & Smolderen, K. G. (2022). Patient-reported outcome measures in symptomatic, non-limb-threatening peripheral artery disease: A state-of-the-art review. *Circulation. Cardiovascular Interventions*, 15(1), e011320.
  98. Sabah, S. A., Hedge, E. A., Abram, S. G. F., Alvand, A., Price, A. J., & Hopewell, S. (2021). Patient-reported outcome measures following revision knee replacement: A review of PROM instrument utilisation and measurement properties using the COSMIN checklist. *British Medical Journal Open*, 11(10), e046169.
  99. Salas, M., Mordin, M., Castro, C., Islam, Z., Tu, N., & Hackshaw, M. D. (2022). Health-related quality of life in women with breast cancer: A review of measures. *BMC Cancer*, 22(1), 66.
  100. Schulz, J., Vitt, E., & Niemier, K. (2022). A systematic review of motor control tests in low back pain based on reliability and validity. *Journal of Bodywork and Movement Therapies*, 29, 239–250.
  101. Schuttert, I., Timmerman, H., Petersen, K. K., McPhee, M. E., Arendt-Nielsen, L., Reneman, M. F., & Wolff, A. P. (2021). The definition, assessment, and prevalence of (human assumed) central sensitisation in patients with chronic low back pain: A systematic review. *Journal of Clinical Medicine*, 10(24), 5931.
  102. Shakshouk, H., Tkaczyk, E. R., Cowen, E. W., El-Azhary, R. A., Hashmi, S. K., Kenderian, S. J., & Lehman, J. S. (2021). Methods to assess disease activity and severity in cutaneous chronic graft-versus-host disease: A critical literature review. *Transplantation and Cellular Therapy*, 27(9), 738–746.
  103. Slavych, B. K., Zraick, R. I., & Ruleman, A. (2021). A systematic review of voice-related patient-reported outcome measures for use with adults. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2021.09.032>
  104. Smeele, H. P., Dijkstra, R. C. H., Kimman, M. L., van der Hulst, R., & Tuinder, S. M. H. (2022). Patient-reported outcome measures used for assessing breast sensation after mastectomy: Not fit for purpose. *Patient*, 15(4), 435–444.
  105. Strong, A., Arumugam, A., Tengman, E., Roijezon, U., & Hager, C. K. (2022). Properties of tests for knee joint threshold to detect passive motion following anterior cruciate ligament injury: A systematic review and meta-analysis. *Journal of Orthopaedic Surgery and Research*, 17(1), 134.
  106. Su, Y., Cochrane, B. B., Yu, S. Y., Reding, K., Herting, J. R., & Zaslavsky, O. (2022). Fatigue in community-dwelling older adults: A review of definitions, measures, and related factors. *Geriatric Nursing*, 43, 266–279.
  107. Thoen, A., Steyaert, J., Alaerts, K., Evers, K., & Van Damme, T. (2023). A systematic review of self-reported stress questionnaires in people on the autism spectrum. *Rev J Autism Dev Disord*, 10(2), 295–318.

108. Tinti, S., Parati, M., De Maria, B., Urbano, N., Sardo, V., Falcone, G., Terzoni, S., Alberti, A., & Destrebecq, A. (2022). Multi-dimensional dyspnea-related scales validated in individuals with cardio-respiratory and cancer diseases: A systematic review of psychometric properties. *Journal of Pain and Symptom Management*, *63*(1), 46–58.
109. Tu, J. Y., Jin, G., Chen, J. H., & Chen, Y. C. (2022). Caregiver burden and dementia: A systematic review of self-report instruments. *Journal of Alzheimer's Disease*, *86*(4), 1527–1543.
110. Ullrich, P., Werner, C., Abel, B., Hummel, M., Bauer, J. M., & Hauer, K. (2022). Assessing life-space mobility : A systematic review of questionnaires and their psychometric properties. *Zeitschrift für Gerontologie und Geriatrie*, *55*(8), 660–666.
111. Valenti, G. D., & Faraci, P. (2022). Instruments measuring fatalism: A systematic review. *Psychological Assessment*, *34*(2), 159–175.
112. van Dokkum, N. H., Reijneveld, S. A., de Best, J., Hamoen, M., Te Wierike, S. C. M., Bos, A. F., & de Kroon, M. L. A. (2022). Criterion validity and applicability of motor screening instruments in children aged 5–6 years: A systematic review. *International Journal of Environmental Research and Public Health*, *19*(2), 781.
113. van Krugten, F. C. W., Feskens, K., Busschbach, J. J. V., Hakkaart-van Roijen, L., & Brouwer, W. B. F. (2021). Instruments to assess quality of life in people with mental health problems: A systematic review and dimension analysis of generic, domain- and disease-specific instruments. *Health and Quality of Life Outcomes*, *19*(1), 249.
114. van Raath, M. I., Chohan, S., Wolkerstorfer, A., van der Horst, C., Limpens, J., Huang, X., Ding, B., Storm, G., van der Hulst, R., & Heger, M. (2021). Treatment outcome measurement instruments for port wine stains: A systematic review of their measurement properties. *Dermatology*, *237*(3), 416–432.
115. Wang, C., Chen, H., Qian, M., Shi, Y., Zhang, N., & Shang, S. (2022). Balance function in patients with COPD: A systematic review of measurement properties. *Clinical Nursing Research*, *31*(6), 1000–1013.
116. Wang, X., Cao, X., Li, J., Deng, C., Wang, T., Fu, L., & Zhang, Q. (2021). Evaluation of patient-reported outcome measures in intermittent self-catheterization users: A systematic review. *Archives of Physical Medicine and Rehabilitation*, *102*(11), 2239–2246.
117. Wen, H., Yang, Z., Zhu, Z., Han, S., Zhang, L., & Hu, Y. (2022). Psychometric properties of self-reported measures of health-related quality of life in people living with HIV: A systematic review. *Health and Quality of Life Outcomes*, *20*(1), 5.
118. Wessels, M. D., van der Putten, A. A. J., & Paap, M. C. S. (2021). Inventory of assessment practices in people with profound intellectual and multiple disabilities in three European countries. *Journal of Applied Research in Intellectual Disabilities*, *34*(6), 1521–1537.
119. Wiitavaara, B., & Florin, J. (2022). Content and psychometric evaluations of questionnaires for assessing physical function in people with arm-shoulder-hand disorders: A systematic review of the literature. *Disability and Rehabilitation*, *44*(24), 7575–7586.
120. Wohlleber, K., Heger, P., Probst, P., Engel, C., Diener, M. K., & Mihaljevic, A. L. (2021). Health-related quality of life in primary hepatic cancer: A systematic review assessing the methodological properties of instruments and a meta-analysis comparing treatment strategies. *Quality of Life Research*, *30*(9), 2429–2466.
121. Wu, J. W., Pepler, L., Maturi, B., Afonso, A. C. F., Sarmiento, J., & Haldenby, R. (2022). Systematic review of motor function scales and patient-reported outcomes in spinal muscular atrophy. *American Journal of Physical Medicine and Rehabilitation*, *101*(6), 590–608.
122. Yanez Touzet, A., Bhatti, A., Dohle, E., Bhatti, F., Lee, K. S., Furlan, J. C., Fehlings, M. G., Harrop, J. S., Zipser, C. M., Rodrigues-Pinto, R., Milligan, J., Sarewitz, E., Curt, A., Rahimi-Movaghar, V., Aarabi, B., Boerger, T. F., Tetreault, L., Chen, R., Guest, J. D., ... Committee, A.O.S.R.-D.S. (2022). Clinical outcome measures and their evidence base in degenerative cervical myelopathy: A systematic review to inform a core measurement set (AO Spine RECODE-DCM). *British Medical Journal Open*, *12*(1), e057650.
123. Zanin, E., Aiello, E. N., Diana, L., Fusi, G., Bonato, M., Niang, A., Ognibene, F., Corvaglia, A., De Caro, C., Cintoli, S., Marchetti, G., Vestri, A., Italian Working Group on t.-n. (2022). Tele-neuropsychological assessment tools in Italy: A systematic review on psychometric properties and usability. *Neurological Sciences*, *43*(1), 125–138.
124. Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926.
125. Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*, *18*, 1–9.
126. Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Marshall, C., Metzendorf, M.-I., Noel-Storr, A., Paynter, R., Rader, T., Thomas, J., & Wieland, L., *Chapter 4: Searching for and selecting studies.*, in *Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022)*, J. Higgins, Editors. 2022, Cochrane.
127. Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., Group\*, Q.-. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*, *155*(8), 529–536.
128. Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of clinical epidemiology*, *63*(8), 854–861.
129. Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., & Boutron, I. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, *355*, i4919.
130. Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2014). *Newcastle-Ottawa quality assessment scale cohort studies*. University of Ottawa.
131. Matvienko-Sikar, K., Hussey, S., Mellor, K., Byrne, M., Clarke, M., Kirkham, J. J., Kottner, J., Quirke, F., Saldanha, I. J., & Smith, V. (2024). Using behavioral science to increase core outcome set use in trials. *Journal of Clinical Epidemiology*, *168*, 111285.
132. Mokkink, L.B., Elsman, E.B.M., & Terwee, C.B., *COSMIN guideline 2.0 for systematic reviews of patient-reported outcome measures*. Submitted for publication.
133. Elsman, E.B.M., Mokkink, L.B., Terwee, C.B., Beaton, D., Gagnier, J.J., Tricco, A.C., Baba, A., Butcher, N.J., Smith, M., Hofstetter, C., Aiyegbusi, O.L., Berardi, A., Farmer, J., Haywood, K.L., Krause, K.R., Markham, S., Mayo-Wilson, E., Mehdipour, A., Ricketts, J., Szatmari, P., Touma, Z., Moher, D., & Offringa, M., *Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024*. Accepted for publication.
134. Schiavo, J. H. (2019). PROSPERO: An international register of systematic review protocols. *Medical Reference Services Quarterly*, *38*(2), 171–180.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



## Authors and Affiliations

Ellen B. M. Elsmann<sup>1,2</sup>  · Lidwine B. Mekkink<sup>1</sup>  · Inger L. Abma<sup>3</sup>  · Olalekan Lee Aiyegbusi<sup>4</sup>  ·  
Alessandro Chiarotto<sup>5</sup>  · Kirstie L. Haywood<sup>6</sup>  · Karen Matvienko-Sikar<sup>7</sup>  · Daniella M. Oosterveer<sup>8</sup>  ·  
Jan J. M. Pool<sup>9</sup>  · Ilse E. J. Swinkels-Meewisse<sup>10</sup> · Martin Offringa<sup>2</sup>  · Caroline B. Terwee<sup>1</sup> 

✉ Caroline B. Terwee  
cb.terwee@amsterdamumc.nl

<sup>1</sup> Department of Epidemiology & Data Science, Amsterdam UMC, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, De Boelelaan 1089a, 1081 HV Amsterdam, The Netherlands

<sup>2</sup> Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada

<sup>3</sup> IQ Health, Radboud Institute of Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>4</sup> Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, UK

<sup>5</sup> Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

<sup>6</sup> Warwick Applied Health, Warwick Medical School, University of Warwick, Coventry, UK

<sup>7</sup> School of Public Health, University College Cork, Cork, Ireland

<sup>8</sup> Basalt, Leiden/The Hague, The Netherlands

<sup>9</sup> University of Applied Sciences, Utrecht, The Netherlands

<sup>10</sup> Amsterdam UMC, Amsterdam, The Netherlands