Qual Life Res (2016) 25:767–779
DOI 10.1007/s11136-015-1122-4

REVIEW

The quality of systematic reviews of health-related outcome measurement instruments

C. B. Terwee[1] · C. A. C. Prinsen[1] · M. G. Ricci Garotti[2] · A. Suman[3] · H. C. W. de Vet[1] · L. B. Mokkink[1]

Accepted: 29 August 2015 / Published online: 7 September 2015

Abstract
Background Systematic reviews of outcome measurement instruments are important tools for the selection of instruments for research and clinical practice. Our aim was to assess the quality of systematic reviews of health-related outcome measurement instruments and to determine whether the quality has improved since our previous study in 2007.
Methods A systematic literature search was performed in MEDLINE and EMBASE between July 1, 2013, and June 19, 2014. The quality of the reviews was rated using a study-specific checklist.
Results A total of 102 reviews were included. In many reviews the search strategy was considered not comprehensive; in only 59 % of the reviews a search was performed in EMBASE and in about half of the reviews there was doubt about the comprehensiveness of the search terms used for type of measurement instruments and measurement properties. In 41 % of the reviews, compared to 30 % in our previous study, the methodological quality of the included studies was assessed. In 58 %, compared to 55 %, the quality of the included instruments was assessed. In 42 %, compared to 7 %, a data synthesis was performed in which the results from multiple studies on the same instrument were somehow combined.
Conclusion Despite a clear improvement in the quality of systematic reviews of outcome measurement instruments in comparison with our previous study in 2007, there is still room for improvement with regard to the search strategy, and especially the quality assessment of the included studies and the included instruments, and the data synthesis.

Keywords Systematic review · Outcome measurement instruments · Measurement properties · Reliability · Validity

✉ C. B. Terwee
cb.terwee@vumc.nl

[1] Department of Epidemiology and Biostatistics and the EMGO+ Institute for Health and Care Research, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

[2] Department of Psychology, University of Bologna, Bologna, Italy

[3] Department of Public Health and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands

Introduction

Health-related outcome measurement instruments are used to evaluate the effects of disease and treatment over time. Systematic reviews of health-related outcome measurement instruments are important tools for the selection of instruments for research and clinical practice and for identifying gaps in knowledge on the quality of outcome measurement instruments, i.e., their measurement properties [1]. Systematic reviews of outcome measurement instruments are being used for a number of purposes: (1) for selecting outcome measurement instruments for monitoring patients in clinical practice; (2) for selecting outcome measurement instruments in the design of new research projects; (3) as a source for evidence on the measurement properties of the outcome measurement instruments used in clinical trials and other (submitted) studies; and (4) for selecting outcome measurement instruments for outcomes included in Core Outcome Sets (COS, i.e., an agreed set of outcomes that should be

measured and reported in all clinical trials of a specific condition [2]).

Systematic reviews should be of high methodological quality to provide a comprehensive and unbiased overview of the measurement properties of the available outcome measurement instruments. In general, a high-quality systematic review consists of a comprehensive search strategy in multiple databases, a selection of abstracts and full-text articles by at least two independent reviewers, a methodological quality assessment of the included studies, and a systematic evaluation and interpretation of the results of the included studies (www.cochrane-handbook.org).

In 2007, we assessed the methodological quality of 148 systematic reviews of health-related outcome measurement instruments published up to March 2007 [1]. Three major limitations were identified. First, the search strategy was often of low quality: in 22 % of the reviews a search was performed in only one database, the search strategy was often poorly described, and in more than 70 % of the reviews it was not reported whether the article selection and data extraction was done by two independent reviewers. Second, in only 30 % of the reviews the methodological quality of the included studies on measurement properties was (partly) evaluated. Third, in only 55 % of the reviews (some) criteria were used to evaluate the quality of the included instruments.

Since this study, the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative developed tools to improve the quality of systematic reviews of outcome measurement instruments. COSMIN is an international group of researchers that aim to improve the selection of health measurement instruments for research and clinical practice. By means of an international Delphi study, COSMIN developed consensus-based standards for assessing the methodological quality of studies on measurement properties, including design requirements and preferred statistical methods [3]. The standards were operationalized into a user-friendly checklist that can be used in systematic reviews of outcome measurement instruments to evaluate the quality of the included studies on measurement properties [4]. The COSMIN checklist was published in 2010 and has been used in more than 60 systematic reviews of outcome measurement instruments since then. In addition, COSMIN researchers developed a protocol for systematic reviews of outcome measurement instruments, made available through the COSMIN website (www.cosmin.nl).

The aim of this study was to assess the current state of the quality of systematic reviews of health-related outcome measurement instruments and to determine whether the methodological quality of these reviews has been improved over time.

Methods

A systematic literature search was performed on June 19, 2014, in MEDLINE (using PubMed) and EMBASE (using www.embase.com) to identify all systematic reviews of health-related outcome measurement instruments published between July 1, 2013, and June 19, 2014. We aimed to identify about 100 reviews to make a comparison with the reviews from our study in 2007. We included reviews published from 2013 onwards that had the potential to have incorporated the COSMIN checklist, which was published in 2010 [3].

The search strategy consisted of search terms for systematic reviews, search terms for measurement instruments, and a validated methodological search filter for measurement properties [5]. References of included reviews were checked for additional relevant reviews. The full search strategy is provided in 'Appendix.'

The following inclusion criteria were used: (1) the study should be a systematic review (we considered a review to be systematic if at least one search in an electronic database was performed); (2) the aim of the review should be to identify all outcome measurement instruments of interest and to summarize the evidence on their measurement properties; (3) the construct of interest of the review should be (aspects of) health status, defined as (a) biological and physiological processes, OR (b) symptoms, OR (c) physical functioning, OR (d) social/psychological functioning, OR (e) general health perceptions, OR (f) health-related quality of life (based on the model of Wilson & Cleary [6]); (4) the study population should contain humans (patients or general population); (5) the instruments of interest should be outcome measurement instruments, defined as instruments which can be/are applied in longitudinal studies to monitor changes in health over time (the outcome measure is the dependent variable); and (6) the study should evaluate and report on at least one or more measurement properties of the included instruments.

The following exclusion criteria were used: (1) reviews of diagnostic or screening instruments which are not used to evaluate the effects of disease and treatment over time; (2) prognostic reviews, i.e., reviews of prognostic studies (prediction models) which aim to predict an outcome using multi-variable analysis; (3) non-English articles; and (4) reviews of only one, or only the most commonly used measurement instruments, or reviews that only included randomized controlled trials (RCTs).

Titles and abstracts were screened by two reviewers independently (CT and MR or CP and MR), and consensus was reached. Full-text articles were screened by two reviewers independently (different couples of CT, CP, MR, and LM), and consensus was reached by discussion among the two reviewers.

A study-specific checklist was developed for data extraction and to evaluate the quality of the systematic reviews of health-related outcome measurement instruments, based on criteria used in our previous study [1], existing guidelines for the appraisal of systematic reviews of clinical trials (Cochrane handbook (http://www.cochrane-handbook.org) [7] and diagnostic studies (Cochrane handbook (http://srdta.cochrane.org/handbook-dta-reviews), and a checklist for assessing the methodological quality of systematic reviews (AMSTAR) [8]). Our checklist contains items on the quality of the research question (inclusion of the construct, target population, measurement instruments and measurement properties of interest), the search strategy (number and kind of databases searched, use of a time window, use of search terms for measurement properties and measurement instruments, language limitations and reference checking), whether the inclusion and exclusion criteria were clearly described, whether and how quality assessment of the included studies was performed, whether and how quality assessment of the included outcome measurement instruments was performed, whether and how data synthesis was performed, whether article selection, data extraction, and quality assessment was done by two reviewers independently, whether evidence-based recommendations were provided, and whether conflict of interest statements were included (Table 1). The quality assessment criteria were similar to those used in our study in 2007. However, in the current study we used additional criteria that were not assessed in our study in 2007. This is indicated in Table 1.

We also counted the number of measurement properties reported in the reviews, using the COSMIN taxonomy (nine measurement properties) [9]. The data extraction and quality appraisal was done by two reviewers independently (different couples of CT, CP, MR, and LM), and consensus was reached by discussion among the two reviewers. If consensus was not reached, a third reviewer was included and the final decision was made based on consensus among the three reviewers.

The results of the study were compared to the results of our previous study, which used a similar search strategy (although not exactly the same search terms for measurement properties as the search filter for measurement properties did not exist yet), the same inclusion criteria, and included all systematic reviews of health-related outcome measurement instruments ($n = 148$) published up to March 2007 [1].

Results

The search strategy yielded 1703 unique records, of which 157 abstracts were selected for retrieving the full-text articles. From these 157 articles, 55 articles were excluded because they were about non-health-related constructs or did not provide information on the measurement properties. The remaining 102 systematic reviews of outcome measurement instruments were included [10–59, 60–109, 110, 111]. A flow chart of the abstract and article selection process is provided in Fig. 1.

The results of the quality appraisal of the systematic reviews of outcome measurement instruments are presented in Table 1 and compared with the results of our previous study for items for which this was possible. The construct, target population, and measurement properties of interest of the review were clearly described in the research aim in more than 80 % of the reviews. However, in only 52 % of the reviews the type of measurement instruments of interest was described in the research aim.

The search strategy was described at least to some extent in 93 % of the reviews, as compared to 84 % in 2007. In 54 % of the reviews, no search terms for measurement properties were used. In 25 % of the reviews no search terms for type of measurement instruments of interest were used.

The median number of databases searched was four (range 1–15), and in 92 % of the reviews a search in at least two databases was performed. This percentage was 76 % in 2007. MEDLINE or PubMed was used in 92 % of the reviews (93 % in 2007), and EMBASE was used in 59 % of the reviews (35 % in 2007).

The selection of abstracts and full-text articles was performed by at least two reviewers independently in 29 % of the reviews, as compared to 22 % in 2007. In 59 % of the reviews (75 % in 2007), it was unclear or not described.

In 41 % of the reviews, the methodological quality of the included studies was assessed, as compared to 30 % in 2007. In 60 % ($n = 25$) of these reviews [10, 18, 21, 26, 29–31, 39, 50, 52, 54, 57, 63, 64, 68, 78, 84, 87, 91, 96, 102, 104, 107, 108, 110] the COSMIN checklist [3, 4] was used. In 60 % ($n = 25$) of the reviews that assessed the quality of the studies, the quality assessment was performed by at least two reviewers independently.

In 58 % of the reviews the quality of the included outcome measurement instruments (i.e., their measurement properties) was assessed, as compared to 55 % in 2007. In 36 % ($n = 21$) of these reviews [10, 15, 18, 26, 29, 35, 39, 50, 54, 57, 62, 63, 68, 78, 84, 85, 91, 93, 97, 100, 108], quality criteria published by Terwee et al. [112] were used. In 32 % ($n = 19$) of these reviews, the quality assessment was performed by at least two reviewers independently.

In 42 % of the reviews some kind of data synthesis was performed in which the results from multiple studies on the same instrument were combined according to predefined criteria, as compared to 7 % in 2007. However, in only about half of these reviews ($n = 20$) [10, 12, 13, 29, 39, 50, 54, 63, 68, 70, 77, 78, 81, 82, 84, 95, 102, 104, 108, 109], it

Table 1 Quality assessment of systematic reviews of outcome measurement instruments

Quality aspects	% current study	% study of 2007 [1][a]
Elements included in the research aim		
Construct of interest	94	
Population of interest	88	
Type of measurement instrument of interest	52	
Measurement properties of interest	81	
All available instruments included	52	
Only instruments included that have at least some evidence of measurement properties	48	
Search strategy described	93	84
No search terms or validated search filter used for		
Measurement properties	64	
Type of instrument	25	
Number of databases searched [median (range)]	4 (1–15)	
Search in at least 2 databases	92	76
MEDLINE/PubMed	92	93
EMBASE	59	35
Additional databases	87	57
Reference checking used	65	
No time limits used or good arguments for a time limit	72	
No language restrictions used	26	79
Inclusion and exclusion criteria clearly described	86	72
Reasons for excluding articles reported	55	
Abstract selection by at least 2 reviewers?		
Yes	41	
No	21	
Unclear	38	
Full-text article selection by at least 2 reviewers?		
Yes	38	
No	13	
Unclear	48	
Abstract and full-text article selection by at least 2 reviewers?		
Yes	29	22
No	12	3
Unclear	59	75
Methodological quality of studies assessed	41	30
Quality assessment of studies done by at least 2 reviewers		
Yes	60	
No	12	
Unclear	28	
Data on measurement properties extracted by at least 2 reviewers		
Yes	25	25
No	13	4
Unclear	62	71
Quality of the instrument (measurement properties) assessed	58	55
Quality assessment of the instrument by at least 2 reviewers		
Yes	33	
No	5	
Unclear	62	
Results from multiple studies on the same instrument somehow combined (e.g., best evidence synthesis or pooling)		
Yes, clearly described	20	7[b]
Yes, but unclear how	22	
No	58	
Data synthesis was performed…		
Per measurement property	79	
Only for domains (reliability, validity, responsiveness)	9	
Only for the whole instrument	12	
Recommendation provided for the best instrument		
One instrument is recommended per construct	23	
More instruments are recommended per construct	26	
No recommendation for the best instrument	51	
Results for the measurement properties reported as raw data		
Yes	56	
Partly	13	
No	31	
Number of measurement properties reported [median (range)]	6 (1–9)	
Conflict of interest or funding source declared	81	
One of the authors of the review is also the developer of one of the instruments evaluated in the review	9	

[a] Not all items were evaluated in the study in 2007

[b] Yes (clearly described or unclear how combined)

was clearly described how this was done. In 81 % ($n = 34$) of the reviews that performed data synthesis, the data synthesis was performed per measurement property separately. In 13 reviews [10, 13, 29, 39, 50, 54, 57, 63, 68, 78, 91, 104, 108], a best evidence syntheses was used based on

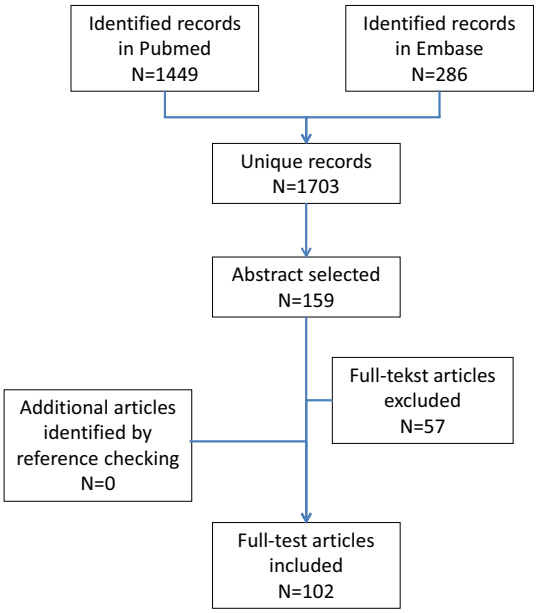

Fig. 1 Flow chart of abstract and article selection

methods used in Cochrane reviews of clinical trials, in which the number, methodological quality, and consistency of the results of the included studies are taken into account.

In 49 % of the reviews clear recommendations were provided for either one or multiple outcome measurement instruments per construct that were considered the best. In about half of these reviews ($n = 23$) [14, 16, 24, 29, 30, 37, 40, 42, 45, 50, 52–54, 63, 69, 72, 87, 90, 96, 101, 102, 110, 111] a recommendation was provided for one best outcome measurement instrument per construct of interest.

In 81 % of the reviews a conflict of interest statement was provided. In nine of these reviews [31, 36, 48, 51, 53, 59, 76, 79, 93] (one of) the authors of the review was involved in the development of one of the instruments included in the review; in only one review [31] this was explicitly stated and the instrument for which this was the case was rated by an independent reviewer. In four reviews [36, 48, 59, 79] the instrument that was recommended as (one of) the best instruments was developed by (one of) the authors of the review, but the involvement of the authors in the development of the instrument was not mentioned in the conflict of interest statement.

Discussion

Despite some clear improvements in the quality of systematic reviews of outcome measurement instruments since our study in 2007 [1], there is still room for improvement with regard to the search strategy, and especially with regard to the quality assessment of the included studies, the quality assessment of the included instruments, and the data synthesis.

In many cases the search strategy was likely to be incomprehensive, for several reasons: First, in only 59 % of the reviews a search was performed in EMBASE (35 % in 2007 [1]). In several systematic reviews of outcome measurement instruments that we performed, we found two or three relevant articles in EMBASE that were not found in MEDLINE [113–115]. Therefore we recommend reviewers to always search at least MEDLINE and EMBASE. Second, in only 25 % of the reviews, no search terms for type of measurement instruments were used. It is understandable that many reviews use search terms for measurement instruments because without search terms for type of measurement instruments, often too many abstracts need to be screened. However, in about half of the reviews we had doubts about the comprehensiveness of the search terms used for type of measurement instruments. For example, a review on questionnaires for assessment of gastroesophageal reflux disease used the search terms: ('Questionnaires'[Mesh] OR questionnaire*[ti] OR scale*[ti]) [16] for type of instruments. We consider it doubtful whether all questionnaires will be found with these terms, because authors may use other terms like 'instrument,' 'outcome measure' or 'patient-reported outcome.' In general it is preferred to use no search terms for measurement instruments, to avoid missing studies. However, if this leads to too many search results, a comprehensive block of search terms need to be developed. For reviews on patient-reported outcome measures, a search filter developed by a group from Oxford University (available from www.cosmin.nl) could be used. Third, in 28 % of the reviews we had doubts about the comprehensiveness of the search terms used for measurement properties. For example, a review on outcome measurement instruments for upper limb function in multiple sclerosis used the search terms (psychometric properties OR psychometrics OR validity OR reliability OR test–retest OR responsiveness) [51] for measurement properties. We consider it doubtful whether all studies on measurement properties will be found with these terms, because authors may use other terms like 'measurement properties,' 'clinimetric properties,' or sensitivity to change.' There is large variation in terminology used for measurement properties, and studies on measurement properties are poorly indexed in databases like MEDLINE [5]. Again, it is preferred to use no search terms for measurement properties to avoid missing studies. However, if this is not feasible, a highly sensitive search filter for finding studies on measurement properties could be used, which was published in 2009 [5] (also available from www.cosmin.nl). This filter was used in only 10 reviews. Fifty-five reviews used no search terms for measurement properties,

which is considered to be a comprehensive strategy because all studies will be screened, but time consuming. Using the search filter for finding studies on measurement properties reduces the number needed to read with about 75 % and has a high sensitivity of about 95–97 % [5]. Fourth, we could not rate the comprehensiveness of the search terms for the construct and target population of interest, because this requires expertise and knowledge of the construct and target population of interest. In many cases it is useful to consult a clinical librarian, which was not reported in most of these reviews. A recent study showed that librarian and information specialist authorship was associated with better reported systematic review search quality [116]. Finally, in 36 % of the reviews reference checking was not performed. It is generally recommended for systematic reviews to check references of included articles. If several relevant articles are found with reference checking, the search strategy was likely incomprehensive and should be adapted.

There is important room for improvement with regard to the quality assessment of the included studies, which was performed in only 41 % of the reviews. This percentage was 30 % in our previous study [1], so it has improved, but not satisfactory. If the methodological quality of a study on the measurement properties of an instrument is inadequate, the results may be biased and the quality of the instrument may be underestimated or overestimated. The COSMIN checklist was used in 25 reviews. An additional 17 reviews used other checklists or recommendations such as QUADAS or ad hoc developed standards. Many of these standards seemed incomplete, for example do not include standards for all measurement properties, or were unclearly described. We recommend to use the COSMIN checklist because it is the only consensus-based checklist containing detailed standards for the preferred design characteristics and statistical methods of studies on measurement properties and includes a standardized rating system for scoring the quality of studies on measurement properties [3, 4, 117]. QUADAS was developed for rating the quality of studies on diagnostic measurement instruments [118, 119], not outcome measurement instruments, so it is less applicable for these kind of reviews.

There is also room for improvement in the quality assessment of the included outcome measurement instruments, which was performed in only 58 % of the reviews. This was 55 % in our previous study, so has not improved much. There is wide variation in how the quality of the instruments was assessed and which criteria for what constitutes good measurement properties were used. The most often used quality criteria were those published by Terwee et al. [112], which were used in 21 reviews. These criteria were not developed using a consensus procedure, but recently, international consensus was reached on these criteria (with minor modifications) in a collaborative study of the COSMIN and the Core Outcome Measures in Effectiveness Trials (COMET) initiative regarding the development of a guideline for selecting outcome measurement instruments for Core Outcome Sets [120].

In less than half of the reviews (42 %) a data synthesis was performed in which the results from multiple studies on the same instrument were somehow combined. In our previous study this was only 7 %. Data synthesis is an important step in a systematic review to develop evidence-based and transparent recommendations for the best instrument for a given context of use. The methodology of data synthesis of studies on measurement properties is not yet as thoroughly developed as it is for reviews of clinical trials, where GRADE recommendations are being used [121–123]. The data synthesis of studies on measurement properties is complex because it is different for each measurement property. For example, to rate the evidence for internal consistency, methods and results of factor analyses (dimensionality) as well as methods and results of internal consistency analyses (Cronbach's alpha) should be considered and combined and this information may come from different studies. To rate the evidence for reliability, statistical pooling of intraclass correlation coefficients might be considered. It is not yet clear how the results of different construct validity or responsiveness studies can be combined, taking into account the strength and the number of hypotheses tested, the constructs being measured with the comparison instruments used, the quality of the comparison instruments, and the kind of subgroups being compared. It remains to be examined if the GRADE recommendations can also be applied, or perhaps in modified form, in systematic reviews of outcome measurement instruments.

Although it is not easy to perform a data synthesis, a review requires a transparent conclusion. Only 49 % of the reviews provided recommendations for the use of one or multiple outcome measurement instruments. We think it is important that reviews provide clear recommendations for the use of instruments because researchers and clinicians need to choose an instrument for their study or use in clinical practice, even when the information on certain measurement properties is scarce or lacking. A recommendation for the use of one instrument per construct and population of interest will facilitate uniformity in outcome reporting and, as a consequence, meta-analyses of studies. It is also important to discourage the use of instruments with evidence for poor measurement properties and to indicate which instruments need further research on their measurement properties.

Finally, it is important that reviewers clearly indicate their involvement in the development or validation of one of more of the included instruments in the review in a conflict of interest statement because this may have

influenced their ratings of the included instruments and their recommendations.

Some limitations of this review should be acknowledged. First, no validated search filter was used for finding systematic reviews, such as the recommended health-evidence.ca systematic review search filter [124], because this filter includes the terms 'meta-analysis' and 'intervention', which were not relevant for our review. Our search terms were, however, quite similar to the remaining search terms of this filter, such as 'systematic review.tw,' so we believe our search was sufficiently sensitive. Second, in our previous review we also searched in PsycINFO but this yielded only 3 of the 148 included reviews. In this study we therefore decided to only use MEDLINE and EMBASE. Our aim was not to find all available systematic reviews but to compare the quality of a set of the most recently published reviews with a set of reviews published six or more years ago. Third, as in many of the included articles in our review, our search terms for type of measurement instruments may also not have been comprehensive, because many different terms are being used in the literature for measurement instruments. Ideally, no search terms should be used for type of measurement instruments but in our review that would have increased the number of records found in PubMed alone to more 160.000. Fourth, the quality assessment of the included reviews was hampered by poor reporting of methods used in the review, especially whether the abstracts and articles were selected by two independent reviewers, and how the data synthesis was performed. We had no time to contact the authors of the reviews for more information. Therefore, we may have underestimated the quality of some reviews.

We conclude that despite a clear improvement in the quality of systematic reviews of outcome measurement instruments in comparison with our previous study in 2007, there is still room for improvement regarding the search strategy, and especially with regard to the quality assessment of the included studies, the quality assessment of the included instruments, and the data synthesis. We recommend reviewers to use the tools developed by the COSMIN group, such as the search filter for finding studies on measurement properties and the COSMIN checklist for assessing the quality of the included studies. A protocol for performing systematic reviews of outcome measurement instruments is available from the authors (CT). The methodology of systematic reviews of outcome measurement instruments need to be further developed. The COSMIN group is currently working on a guideline for systematic reviews of outcome measurement instruments (an update of the currently available protocol). There is also room for improvement with regard to the reporting of systematic reviews of outcome measurement instruments. The development of guidelines for reporting systematic reviews of outcome measurement instruments merits attention in future research.

Compliance with ethical standards

Conflict of interest Three authors of this study (Terwee, de Vet, Mokkink) are members of the COSMIN steering committee and developed the recommended COSMIN checklist. Two authors (Terwee, de Vet) participated in the development of the recommended PubMed search filter for finding studies on measurement properties [5]. None of the authors were co-author on any of the included articles in this review.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix: Search strategy

Pubmed search June 19, 2014

#1: (instruments[tiab] OR scales[tiab] OR Questionnaires[tiab] OR measures[ti] OR methods[ti] OR outcome measurements[tiab] OR (tests[tiab] AND review[tiab]) OR Questionnaires[MeSH] OR interview[MeSH]).

#2 (systematic[sb] OR (literature AND search*) OR (Medline AND search*) OR review[ti]).

#3 (instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test–retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab]

OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]).

(#1 AND #2 AND #3) NOT ('delphi-technique'[ti] OR cross-sectional[ti] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]).

Filters: Publication date from 2013/07/01.

Embase search June 19, 2014

#1 instruments:ti,ab OR scales:ti,ab OR questionnaires:ti,ab OR measures:ti OR methods:ti OR outcome-measurements:ti,ab OR (tests:ti,ab AND review:ti,ab) OR 'outcomes research'/de OR 'treatment outcome'/de OR 'psychologic test'/de OR 'measurement'/de OR 'functional assessment'/de OR 'pain assessment'/de OR 'questionnaire'/de OR 'rating scale'/de.

#2 review:ti OR (literature AND search*) OR (medline AND search*) OR 'systematic review'/exp.

#3 'intermethod comparison'/exp OR 'data collection method'/exp OR 'validation study'/exp OR 'feasibility study'/exp OR 'pilot study'/exp OR 'psychometry'/exp OR 'reproducibility'/exp OR reproducib*:ab,ti OR 'audit':ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR 'observer variation'/exp OR 'observer variation':ab,ti OR 'discriminant analysis'/exp OR 'validity'/exp OR reliab*:ab,ti OR valid*:ab,ti OR 'coefficient':ab,ti OR 'internal consistency':ab,ti OR (cronbach*:ab,ti AND ('alpha':ab,ti OR 'alphas':ab,ti)) OR 'item correlation':ab,ti OR 'item correlations':ab,ti OR 'item selection':ab,ti OR 'item selections':ab,ti OR 'item reduction':ab,ti OR 'item reductions':ab,ti OR 'agreement':ab,ti OR 'precision':ab,ti OR 'imprecision':ab,ti OR 'precise values':ab,ti OR 'test–retest':ab,ti OR ('test':ab,ti AND 'retest':ab,ti) OR (reliab*:ab,ti AND ('test':ab,ti OR 'retest':ab,ti)) OR 'stability':ab,ti OR 'interrater':ab,ti OR 'inter-rater':ab,ti OR 'intrarater':ab,ti OR 'intra-rater':ab,ti OR 'intertester':ab,ti OR 'inter-tester':ab,ti OR 'intratester':ab,ti OR 'intra-tester':ab,ti OR 'interobeserver':ab,ti OR 'inter-observer':ab,ti OR 'intraobserver':ab,ti OR 'intra-observer':ab,ti OR 'intertechnician':ab,ti OR 'inter-technician':ab,ti OR 'intratechnician':ab,ti OR 'intra-technician':ab,ti OR 'interexaminer':ab,ti OR 'inter-examiner':ab,ti OR 'intraexaminer':ab,ti OR 'intra-examiner':ab,ti OR 'interassay':ab,ti OR 'inter-assay':ab,ti OR 'intraassay':ab,ti OR 'intra-assay':ab,ti OR 'interindividual':ab,ti OR 'inter-individual':ab,ti OR 'intraindividual':ab,ti OR 'intra-individual':ab,ti OR 'interparticipant':ab,ti OR 'inter-participant':ab,ti OR 'intraparticipant':ab,ti OR 'intra-participant':ab,ti OR 'kappa':ab,ti OR 'kappas':ab,ti OR 'coefficient of variation':ab,ti OR repeatab*:ab,ti OR (replicab*:ab,ti OR 'repeated':ab,ti AND ('measure':ab,ti OR 'measures':ab,ti OR 'findings':ab,ti OR 'result':ab,ti OR 'results':ab,ti OR 'test':ab,ti OR 'tests':ab,ti)) OR generaliza*:ab,ti OR

generalisa*:ab,ti OR 'concordance':ab,ti OR ('intra-class':ab,ti AND correlation*:ab,ti) OR 'discriminative':ab,ti OR 'known group':ab,ti OR 'factor analysis':ab,ti OR 'factor analyses':ab,ti OR 'factor structure':ab,ti OR 'factor structures':ab,ti OR 'dimensionality':ab,ti OR subscale*:ab,ti OR 'multitrait scaling analysis':ab,ti OR 'multitrait scaling analyses':ab,ti OR 'item discriminant':ab,ti OR 'interscale correlation':ab,ti OR 'interscale correlations':ab,ti OR ('error':ab,ti OR 'errors':ab,ti AND (measure*:ab,ti OR correlat*:ab,ti OR evaluat*:ab,ti OR 'accuracy':ab,ti OR 'accurate':ab,ti OR 'precision':ab,ti OR 'mean':ab,ti)) OR 'individual variability':ab,ti OR 'interval variability':ab,ti OR 'rate variability':ab,ti OR 'variability analysis':ab,ti OR ('uncertainty':ab,ti AND ('measurement':ab,ti OR 'measuring':ab,ti)) OR 'standard error of measurement':ab,ti OR sensitiv*:ab,ti OR responsive*:ab,ti OR ('limit':ab,ti AND 'detection':ab,ti) OR 'minimal detectable concentration':ab,ti OR interpretab*:ab,ti OR (small*:ab,ti AND ('real':ab,ti OR 'detectable':ab,ti) AND ('change':ab,ti OR 'difference':ab,ti)) OR 'meaningful change':ab,ti OR 'minimal important change':ab,ti OR 'minimal important difference':ab,ti OR 'minimally important change':ab,ti OR 'minimally important difference':ab,ti OR 'minimal detectable change':ab,ti OR 'minimal detectable difference':ab,ti OR 'minimally detectable change':ab,ti OR 'minimally detectable difference':ab,ti OR 'minimal real change':ab,ti OR 'minimal real difference':ab,ti OR 'minimally real change':ab,ti OR 'minimally real difference':ab,ti OR 'ceiling effect':ab,ti OR 'floor effect':ab,ti OR 'item response model':ab,ti OR 'irt':ab,ti OR 'rasch':ab,ti OR 'differential item functioning':ab,ti OR 'dif':ab,ti OR 'computer adaptive testing':ab,ti OR 'item bank':ab,ti OR 'cross-cultural equivalence':ab,ti.

(#1 AND #2 AND #3) NOT 'Delphi technique':ti OR Cross-sectional:ti OR 'case report'/de OR letter:it OR animal/exp OR 'animal model'/exp OR 'animal experiment'/exp.

Filters: Publication date from 2013/07/01.

References

1. Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., et al. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research, 18*, 313–333.
2. Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., et al. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials, 13*, 132.
3. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*, 539–549.
4. Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research, 21*, 651–657.
5. 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.
6. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA, 273*, 59–65.
7. Higgins, J. P. T. & Greens, S. The Cochrane Collaboration (Ed.). (2008). Cochrane handbook for systematic reviews of interventions. Version 5.0.1 (update September 2008). www.cochrane-handbook.org
8. Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., et al. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology, 7*, 10.
9. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (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*, 737–745.
10. Ammann-Reiffer, C., Bastiaenen, C. H., de Bie, R. A., & van Hedel, H. J. (2014). Measurement properties of gait-related outcomes in youth with neuromuscular diagnoses: A systematic review. *Physical Therapy.*
11. Ashford, S., & Turner-Stokes, L. (2013). Systematic review of upper-limb function measurement methods in botulinum toxin intervention for focal spasticity. *Physiotherapy Research International, 18*, 178–189.
12. Barrett, A., Clark, M., Demuro, C., & Esser, D. (2013). Proxy-reported questionnaires for young children with asthma: A structured review. *European Respiratory Journal, 42*, 513–526.
13. Barrett, E., McCreesh, K., & Lewis, J. (2014). Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: A systematic review. *Manual Therapy, 19*, 10–17.
14. Bassi, F., Carr, A. B., Chang, T. L., Estafanous, E. W., Garrett, N. R., Happonen, R. P., et al. (2013). Functional outcomes for clinical evaluation of implant restorations. *The International Journal of Prosthodontics, 26*, 411–418.
15. Bialocerkowski, A., O'shea, K., & Pin, T. W. (2013). Psychometric properties of outcome measures for children and adolescents with brachial plexus birth palsy: A systematic review. *Developmental Medicine and Child Neurology, 55*, 1075–1088.
16. Bolier, E. A., Kessing, B. F., Smout, A. J., & Bredenoord, A. J. (2013). Systematic review: Questionnaires for assessment of gastroesophageal reflux disease. *Diseases of the Esophagus.*
17. Bowling, A., Rowe, G., Adams, S., Sands, P., Samsi, K., Crane, M. et al. (2015). Quality of life in dementia: A systematically conducted narrative review of dementia-specific measurement scales. *Aging and Mental Health, 19*, 13–31.
18. Chandratre, P., Roddy, E., Clarson, L., Richardson, J., Hider, S. L., & Mallen, C. D. (2013). Health-related quality of life in gout: A systematic review. *Rheumatology (Oxford), 52*, 2031–2040.
19. Chang, K. W., Justice, D., Chung, K. C., & Yang, L. J. (2013). A systematic review of evaluation methods for neonatal brachial plexus palsy. *Journal of Neurosurgery: Pediatrics.*
20. Chien, C. W., Rodger, S., Copley, J., & McLaren, C. (2014). Measures of participation outcomes related to hand use for 2- to

12-year-old children with disabilities: A systematic review. *Child: Care, Health and Development, 40*, 458–471.
21. Chow, M. Y., Morrow, A. M., Cooper Robbins, S. C., & Leask, J. (2013). Condition-specific quality of life questionnaires for caregivers of children with pediatric conditions: A systematic review. *Quality of Life Research, 22*, 2183–2200.
22. Conway, A., Page, K., Rolley, J. X., & Worrall-Carter, L. (2014). A review of sedation scales for the cardiac catheterization laboratory. *Journal of PeriAnesthesia Nursing, 29*, 191–212.
23. Coombs, T., Nicholas, A., & Pirkis, J. (2013). A review of social inclusion measures. *Australian and New Zealand Journal of Psychiatry, 47*, 906–919.
24. Crosta, Q. R., Ward, T. M., Walker, A. J., & Peters, L. M. (2014). A review of pain measures for hospitalized children with cognitive impairment. *Journal for Specialists in Pediatric Nursing, 19*, 109–118.
25. Dalbeth, N., Zhong, C. S., Grainger, R., Khanna, D., Khanna, P. P., Singh, J. A., et al. (2014). Outcome measures in acute gout: A systematic literature review. *Journal of Rheumatology, 41*, 558–568.
26. de Almeida, J. R., Witterick, I. J., Gullane, P. J., Gentili, F., Lohfeld, L., Ringash, J., et al. (2013). Quality of life instruments for skull base pathology: Systematic review and methodologic appraisal. *Head and Neck, 35*, 1221–1231.
27. de Boer, M. K., Castelein, S., Wiersma, D., Schoevers, R. A., & Knegtering, H. (2014). A systematic review of instruments to measure sexual functioning in patients using antipsychotics. *The Journal of Sex Research, 51*, 383–389.
28. Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health, 8*, 14.
29. Dekkers, K. J., Rameckers, E. A., Smeets, R. J., & Janssen-Potten, Y. J. (2014). Upper extremity strength measurement for children with cerebral palsy: A systematic review of available instruments. *Physical Therapy, 94*, 609–622.
30. Dorfman, T. L., Sumamo, S. E., Rempel, G. R., Scott, S. D., & Hartling, L. (2014). An evaluation of instruments for scoring physiological and behavioral cues of pain, non-pain related distress, and adequacy of analgesia and sedation in pediatric mechanically ventilated patients: A systematic review. *International Journal of Nursing Studies, 51*, 654–676.
31. Field, D., & Livingstone, R. (2013). Clinical tools that measure sitting posture, seated postural control or functional abilities in children with motor impairments: A systematic review. *Clinical Rehabilitation, 27*, 994–1004.
32. Gor-Garcia-Fogeda, M. D., Molina-Rueda, F., Cuesta-Gomez, A., Carratala-Tejada, M., Alguacil-Diego, I. M., & Miangolarra-Page, J. C. (2014). Scales to assess gross motor function in stroke patients: A systematic review. *Archives of Physical Medicine and Rehabilitation, 95*, 1174–1183.
33. Gorecki, C., Nixon, J., Lamping, D. L., Alavi, Y., & Brown, J. M. (2014). Patient-reported outcome measures for chronic wounds with particular reference to pressure ulcer research: A systematic review. *International Journal of Nursing Studies, 51*, 157–165.
34. Grubbs, J. R, Jr, Tolleson-Rinehart, S., Huynh, K., & Davis, R. M. (2014). A review of quality of life measures in dry eye questionnaires. *Cornea, 33*, 215–218.
35. Hamoen, E. H., De, R. M., Witjes, J. A., Barentsz, J. O., & Rovers, M. M. (2014). Measuring health-related quality of life in men with prostate cancer: A systematic review of the most used questionnaires and their validity. *Urologic Oncology.*
36. Han, H. R., Song, H. J., Nguyen, T., & Kim, M. T. (2014). Measuring self-care in patients with hypertension: A systematic review of literature. *Journal of Cardiovascular Nursing, 29*, 55–67.
37. Harrington, S., Michener, L. A., Kendig, T., Miale, S., & George, S. Z. (2014). Patient-reported upper extremity outcome measures used in breast cancer survivors: A systematic review. *Archives of Physical Medicine and Rehabilitation, 95*, 153–162.
38. Hawkins, A. T., Henry, A. J., Crandell, D. M., & Nguyen, L. L. (2014). A systematic review of functional and quality of life assessment after major lower extremity amputation. *Annals of Vascular Surgery, 28*, 763–780.
39. Haywood, K. L., Collin, S. M., & Crawley, E. (2014). Assessing severity of illness and outcomes of treatment in children with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (CFS/ME): A systematic review of patient-reported outcome measures (PROMs). *Child: Care, Health and Development.*
40. Hoey, L. M., Fulbrook, P., & Douglas, J. A. (2014). Sleep assessment of hospitalised patients: A literature review. *International Journal of Nursing Studies*.
41. Horton, L., Duffy, T., Hollins, M. C., & Martin, C. R. (2014). Comprehensive assessment of alcohol-related brain damage (ARBD): Gap or chasm in the evidence? *Journal of Psychiatric and Mental Health Nursing*.
42. Ikeda, E., Hinckson, E., & Krageloh, C. (2014). Assessment of quality of life in children and youth with autism spectrum disorder: A critical review. *Quality of Life Research, 23*, 1069–1085.
43. Izumi, K. (2014). The measures to evaluate constipation: A review article. *Gastroenterology Nursing, 37*, 137–146.
44. Jabir, S. (2013). Assessing improvement in quality of life and patient satisfaction following body contouring surgery in patients with massive weight loss: A critical review of outcome measures employed. *Plastic Surgery International, 2013*, 515737.
45. James, S., Ziviani, J., & Boyd, R. (2014). A systematic review of activities of daily living measures for children and adolescents with cerebral palsy. *Developmental Medicine and Child Neurology, 56*, 233–244.
46. Janaudis-Ferreira, T., Beauchamp, M. K., Robles, P. G., Goldstein, R. S., & Brooks, D. (2014). Measurement of activities of daily living in patients with COPD: A systematic review. *Chest, 145*, 253–271.
47. Karazsia, B. T., & Brown Kirschman, K. J. (2013). Evidence-based assessment of childhood injuries and physical risk-taking behaviors. *Journal of Pediatric Psychology, 38*, 829–845.
48. Khadka, J., McAlinden, C., & Pesudovs, K. (2013). Quality assessment of ophthalmic questionnaires: Review and recommendations. *Optometry and Vision Science, 90*, 720–744.
49. Koene, S., Jansen, M., Verhaak, C. M., De Vrueh, R. L., De Groot, I. J., & Smeitink, J. A. (2013). Towards the harmonization of outcome measures in children with mitochondrial disorders. *Developmental Medicine and Child Neurology, 55*, 698–706.
50. Kroman, S. L., Roos, E. M., Bennell, K. L., Hinman, R. S., & Dobson, F. (2014). Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: A systematic review. *Osteoarthritis Cartilage, 22*, 26–39.
51. Lamers, I., Kelchtermans, S., Baert, I., & Feys, P. (2014). Upper limb assessment in multiple sclerosis: A systematic review of outcome measures and their psychometric properties. *Archives of Physical Medicine and Rehabilitation, 95*, 1184–1200.
52. Larsen, C. M., Juul-Kristensen, B., Lund, H., & Sogaard, K. (2014). Measurement properties of existing clinical assessment methods evaluating scapular positioning and function. A systematic review. *Physiotherapy Theory and Practice.*

53. Lee, E. H., Klassen, A. F., Nehal, K. S., Cano, S. J., Waters, J., & Pusic, A. L. (2013). A systematic review of patient-reported outcome instruments of nonmelanoma skin cancer in the dermatologic population. *Journal of the American Academy of Dermatology, 69*, e59–e67.
54. Lee, J., Kim, S. H., Moon, S. H., & Lee, E. H. (2014). Measurement properties of rheumatoid arthritis-specific quality-of-life questionnaires: Systematic review of the literature. *Quality of Life Research,*
55. Lee, K. S., & Moser, D. K. (2013). Heart failure symptom measures: Critical review. *European Journal of Cardiovascular Nursing, 12*, 418–428.
56. Makai, P., Brouwer, W. B., Koopmanschap, M. A., Stolk, E. A., & Nieboer, A. P. (2014). Quality of life instruments for economic evaluations in health and social care for older people: A systematic review. *Social Science and Medicine, 102*, 83–93.
57. Marks, M., Schoones, J. W., Kolling, C., Herren, D. B., Goldhahn, J., & Vliet Vlieland, T. P. (2013). Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis: A systematic literature review. *Journal of Hand Surgery (European Volume), 38*, 822–838.
58. Mollayeva, T., Kendzerska, T., & Colantonio, A. (2013). Self-report instruments for assessing sleep dysfunction in an adult traumatic brain injury population: A systematic review. *Sleep Medicine Reviews, 17*, 411–423.
59. Mosli, M. H., Feagan, B. G., Sandborn, W. J., D'haens, G., Behling, C., Kaplan, K., et al. (2014). Histologic evaluation of ulcerative colitis: A systematic review of disease activity indices. *Inflammatory Bowel Diseases, 20*, 564–575.
60. Murphy, D. R., & Lopez, M. (2013). Neck and back pain specific outcome assessment questionnaires in the Spanish language: A systematic literature review. *The Spine Journal, 13*, 1667–1674.
61. Muzzatti, B., & Annunziata, M. A. (2013). Assessing quality of life in long-term cancer survivors: A review of available tools. *Supportive Care in Cancer, 21*, 3143–3152.
62. Niu, H. Y., Niu, C. Y., Wang, J. H., Zhang, Y., & He, P. (2014). Health-related quality of life in women with breast cancer: A literature-based review of psychometric properties of breast cancer-specific measures. *Asian Pacific Journal of Cancer Prevention, 15*, 3533–3536.
63. Noben, C. Y., Evers, S. M., Nijhuis, F. J., & de Rijk, A. E. (2014). Quality appraisal of generic self-reported instruments measuring health-related productivity changes: A systematic review. *BMC Public Health, 14*, 115.
64. Oliveira, C. C., Lee, A., Granger, C. L., Miller, K. J., Irving, L. B., & Denehy, L. (2013). Postural control and fear of falling assessment in people with chronic obstructive pulmonary disease: A systematic review of instruments, international classification of functioning, disability and health linkage, and measurement properties. *Archives of Physical Medicine and Rehabilitation, 94*, 1784–1799.
65. Oppewal, A., Hilgenkamp, T. I., van, W. R., & Evenhuis, H. M. (2013). Cardiorespiratory fitness in individuals with intellectual disabilities—A review. *Research in Developmental Disabilities, 34*, 3301–3316.
66. Palese, A., Tameni, A., Ambrosi, E., Albanese, S., Barausse, M., Benazzi, B., et al. (2014). Clinical assessment instruments validated for nursing practice in the Italian context: A systematic review of the literature. *Annali dell Istituto Superiore di Sanita, 50*, 67–76.
67. Paltzer, J., Barker, E., & Witt, W. P. (2013). Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: A review of existing measures. *Quality of Life Research, 22*, 1177–1187.
68. Park, T., Reilly-Spong, M., & Gross, C. R. (2013). Mindfulness: A systematic review of instruments to measure an emergent patient-reported outcome (PRO). *Quality of Life Research, 22*, 2639–2659.
69. Peterson, D. A., Berque, P., Jabusch, H. C., Altenmuller, E., & Frucht, S. J. (2013). Rating scales for musician's dystonia: The state of the art. *Neurology, 81*, 589–598.
70. Phillips, R. L., Olds, T., Boshoff, K., & Lane, A. E. (2013). Measuring activity and participation in children and adolescents with disabilities: A literature review of available instruments. *Australian Occupational Therapy Journal, 60*, 288–300.
71. Pons, C., Remy-Neris, O., Medee, B., & Brochard, S. (2013). Validity and reliability of radiological methods to assess proximal hip geometry in children with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology, 55*, 1089–1102.
72. Pullmer, R., Linden, W., Rnic, K., & Vodermaier, A. (2014). Measuring symptoms in gastrointestinal cancer: A systematic review of assessment instruments. *Supportive Care in Cancer.*
73. Ray-Barruel, G., Polit, D. F., Murfield, J. E., & Rickard, C. M. (2014). Infusion phlebitis assessment measures: A systematic review. *Journal of Evaluation in Clinical Practice, 20*, 191–202.
74. Ritchie, L., Wright-St Clair, V. A., Keogh, J., & Gray, M. (2014). Community integration after traumatic brain injury: A systematic review of the clinical implications of measurement and service provision for older adults. *Archives of Physical Medicine and Rehabilitation, 95*, 163–174.
75. Ritmala-Castren, M., Lakanmaa, R. L., Virtanen, I., & Leino-Kilpi, H. (2013). Evaluating adult patients' sleep: An integrative literature review in critical care. *Scandinavian Journal of Caring Sciences.*
76. Robertson, S. J., Burnett, A. F., & Cochrane, J. (2014). Tests examining skill outcomes in sport: A systematic review of measurement properties and feasibility. *Sports Medicine (Auckland, N. Z.), 44*, 501–518.
77. Robinson, B. R., Berube, M., Barr, J., Riker, R., & Gelinas, C. (2013). Psychometric analysis of subjective sedation scales in critically ill adults. *Critical Care Medicine, 41*, S16–S29.
78. Saether, R., Helbostad, J. L., Riphagen, I. I., & Vik, T. (2013). Clinical tools to assess balance in children and adults with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology, 55*, 988–999.
79. Salvilla, S. A., Dubois, A. E., Flokstra-de Blok, B. M., Panesar, S. S., Worth, A., Patel, S., et al. (2014). Disease-specific health-related quality of life instruments for IgE-mediated food allergy. *Allergy, 69*, 834–844.
80. Samaan, M. A., Mosli, M. H., Sandborn, W. J., Feagan, B. G., D'Haens, G. R., & Dubcenco, E. et al. (2014). A systematic review of the measurement of endoscopic healing in ulcerative colitis clinical trials: Recommendations and implications for future research. *Inflammatory Bowel Diseases.*
81. Schmidt, S., Garin, O., Pardo, Y., Valderas, J. M., Alonso, J., & Rebollo, P. et al. (2014). Assessing quality of life in patients with prostate cancer: A systematic and standardized comparison of available instruments. *Quality of Life Research.*
82. Schmidt, S., Ferrer, M., Gonzalez, M., Gonzalez, N., Valderas, J. M., Alonso, J., et al. (2014). Evaluation of shoulder-specific patient-reported outcome measures: A systematic and standardized comparison of available evidence. *Journal of Shoulder and Elbow Surgery, 23*, 434–444.
83. Schmit, K. M., Coeytaux, R. R., Goode, A. P., McCrory, D. C., Yancy, W. S, Jr, Kemper, A. R., et al. (2013). Evaluating cough assessment tools: A systematic review. *Chest, 144*, 1819–1826.
84. Schmitt, J., Langan, S., Deckert, S., Svensson, A., von, K. L., Thomas, K., et al. (2013). Assessment of clinical signs of atopic dermatitis: A systematic review and recommendation. *J Allergy Clin Immunol, 132*, 1337–1347.
85. Sellers, D., Pennington, L., Mandy, A., & Morris, C. (2014). A systematic review of ordinal scales used to classify the eating

and drinking abilities of individuals with cerebral palsy. *Developmental Medicine and Child Neurology, 56*, 313–322.
86. Shanks, V., Williams, J., Leamy, M., Bird, V. J., Le, B. C., & Slade, M. (2013). Measures of personal recovery: A systematic review. *Psychiatric Services, 64*, 974–980.
87. Silva, P. F., Quintino, L. F., Franco, J., & Faria, C. D. (2014). Measurement properties and feasibility of clinical tests to assess sit-to-stand/stand-to-sit tasks in subjects with neurological disease: A systematic review. *Brazilian Journal of Physical Therapy, 18*, 99–110.
88. Singh, A., Tetreault, L., Casey, A., Laing, R., Statham, P., & Fehlings, M. G. (2013). A summary of assessment tools for patients suffering from cervical spondylotic myelopathy: A systematic review on validity, reliability and responsiveness. *European Spine Journal*.
89. Sklar, M., Groessl, E. J., O'Connell, M., Davidson, L., & Aarons, G. A. (2013). Instruments for measuring mental health recovery: A systematic review. *Clinical Psychology Review, 33*, 1082–1095.
90. Smith, T., Hameed, Y., Cross, J., Sahota, O., & Fox, C. (2013). Assessment of people with cognitive impairment and hip fracture: A systematic review and meta-analysis. *Archives of Gerontology and Geriatrics, 57*, 117–126.
91. Speyer, R., Cordier, R., Kertscher, B., & Heijnen, B. J. (2014). Psychometric properties of questionnaires on functional health status in Oropharyngeal Dysphagia: A systematic literature review. *BioMed Research International, 2014*, 458678.
92. Stephensen, D., Drechsler, W. I., & Scott, O. M. (2014). Outcome measures monitoring physical function in children with haemophilia: A systematic review. *Haemophilia, 20*, 306–321.
93. Stevelink, S. A., & van Brakel, W. H. (2013). The cross-cultural equivalence of participation instruments: A systematic review. *Disability and Rehabilitation, 35*, 1256–1268.
94. Stuart, S., Alcock, L., Galna, B., Lord, S., & Rochester, L. (2014). The measurement of visual sampling during real-world activity in Parkinson's disease and healthy controls: A structured literature review. *Journal of Neuroscience Methods, 222*, 175–188.
95. Tadic, V., Hogan, A., Sobti, N., Knowles, R. L., & Rahi, J. S. (2013). Patient-reported outcome measures (PROMs) in paediatric ophthalmology: A systematic review. *British Journal of Ophthalmology, 97*, 1369–1381.
96. The, B., Reininga, I. H., El, M. M., & Eygendaal, D. (2013). Elbow-specific clinical rating systems: Extent of established validity, reliability, and responsiveness. *Journal of Shoulder and Elbow Surgery, 22*, 1380–1394.
97. Timmerman, A. A., Speyer, R., Heijnen, B. J., & Klijn-Zwijnenberg, I. R. (2014). Psychometric characteristics of health-related quality-of-life questionnaires in oropharyngeal dysphagia. *Dysphagia, 29*, 183–198.
98. Tsai, A. C., Scott, J. A., Hung, K. J., Zhu, J. Q., Matthews, L. T., Psaros, C., et al. (2013). Reliability and validity of instruments for assessing perinatal depression in African settings: Systematic review and meta-analysis. *PLoS ONE, 8*, e82521.
99. Tsai, A. C. (2014). Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: Systematic review and meta-analysis. *Journal of Acquired Immune Deficiency Syndromes.*
100. Tyack, Z., Wasiak, J., Spinks, A., Kimble, R., & Simons, M. (2013). A guide to choosing a burn scar rating scale for clinical or research use. *Burns, 39*, 1341–1350.
101. Tyson, S. F., & Brown, P. (2013). How to measure pain in neurological conditions? A systematic review of psychometric properties and clinical utility of measurement tools. *Clinical Rehabilitation, 28*, 669–686.
102. Tyson, S. F. & Brown, P. (2014). How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far. *Clinical Rehabilitation.*
103. van der Linde, B. W., van Netten, J. J., Otten, E., Postema, K., Geuze, R. H., & Schoemaker, M. M. (2013). A systematic review of instruments for assessment of capacity in activities of daily living in children with developmental co-ordination disorder. *Child: Care, Health and Development.*
104. van der Meer, S., Trippolini, M. A., van der Palen, J., Verhoeven, J., & Reneman, M. F. (2013). Which instruments can detect submaximal physical and functional capacity in patients with chronic nonspecific back pain? A systematic review. *Spine (Phila Pa 1976), 38*, E1608–E1615.
105. van Tuyl, L. H., Lems, W. F., & Boers, M. (2014). Measurement of stiffness in patients with rheumatoid arthritis in low disease activity or remission: A systematic review. *BMC Musculoskeletal Disorders, 15*, 28.
106. Villaverde, V., Rosario, M. P., Loza, E., & Perez, F. (2014). Systematic review of the value of ultrasound and magnetic resonance musculoskeletal imaging in the evaluation of response to treatment of gout. *Reumatología Clínica, 10*, 160–163.
107. Wheelwright, S., Darlington, A. S., Hopkinson, J. B., Fitzsimmons, D., White, A., & Johnson, C. D. (2013). A systematic review of health-related quality of life instruments in patients with cancer cachexia. *Supportive Care in Cancer, 21*, 2625–2636.
108. Wigham, S., & McConachie, H. (2014). Systematic review of the properties of tools used to measure outcomes in anxiety intervention studies for children with autism spectrum disorders. *PLoS ONE, 9*, e85268.
109. Williamson, A., Andersen, M., Redman, S., Dadds, M., D'Este, C., Daniels, J., et al. (2014). Measuring mental health in Indigenous young people: A review of the literature from 1998–2008. *Clinical Child Psychology and Psychiatry, 19*, 260–272.
110. Winser, S. J., Smith, C. M., Hale, L. A., Claydon, L. S., Whitney, S. L., & Mehta, P. (2014). Systematic review of the psychometric properties of balance measures for cerebellar ataxia. *Clinical Rehabilitation.*
111. Xu, J., Evans, T. J., Coon, C., Copley-Merriman, K., & Su, Y. (2013). Measuring patient-reported outcomes in advanced gastric cancer. *Ecancermedicalscience, 7*, 351.
112. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34–42.
113. de Boer, M. R., Moll, A. C., de Vet, H. C., Terwee, C. B., Volker-Dieben, H. J., & van Rens, G. H. (2004). Psychometric properties of vision-related quality of life questionnaires: A systematic review. *Ophthalmic and Physiological Optics, 24*, 257–273.
114. Elbers, R. G., Rietberg, M. B., van Wegen, E. E., Verhoef, J., Kramer, S. F., Terwee, C. B., et al. (2012). Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: A systematic review of measurement properties. *Quality of Life Research, 21*, 925–944.
115. Terwee, C. B., Bouwmeester, W., van Elsland, S. L., de Vet, H. C., & Dekker, J. (2011). Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: A systematic review of measurement properties. *Osteoarthritis and Cartilage*.
116. Rethlefsen, M. L., Farrell, A. M., Osterhaus Trzasko, L. C., & Brigham, T. J. (2015). Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology, 68*, 617–626.
117. Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist

for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*, 22.

118. Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., & Kleijnen, J. (2003). The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology, 3*, 25.
119. Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., et al. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*, 529–536.
120. Prinsen, C. A., Vohra, S., Rose, M. R., King-Jones, S., Ishaque, S., Bhaloo, Z., et al. (2014). Core Outcome Measures in Effectiveness Trials (COMET) initiative: Protocol for an international Delphi study to achieve consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set'. *Trials, 15*, 247.
121. Brozek, J. L., Akl, E. A., Alonso-Coello, P., Lang, D., Jaeschke, R., Williams, J. W., et al. (2009). Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy, 64*, 669–677.
122. Brozek, J. L., Akl, E. A., Jaeschke, R., Lang, D. M., Bossuyt, P., Glasziou, P., et al. (2009). Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy, 64*, 1109–1116.
123. Brozek, J. L., Akl, E. A., Compalati, E., Kreis, J., Terracciano, L., Fiocchi, A., et al. (2011). Grading quality of evidence and strength of recommendations in clinical practice guidelines Part 3 of 3. The GRADE approach to developing recommendations. *Allergy, 66*, 588–595.
124. Lee, E., Dobbins, M., Decorby, K., McRae, L., Tirilis, D., & Husson, H. (2012). An optimal search filter for retrieving systematic reviews and meta-analyses. *BMC Medical Research Methodology, 12*, 51.