Evaluation of the methodological quality of systematic reviews of health status measurement instruments
Mokkink, L.B.; Terwee, C.B.; Stratford, P.; Alonso, J.R.; Patrick, D.L.; Riphagen, I.; Knol, D.L.; Bouter, L.M.; de Vet, H.C.W.

published in
Quality of Life Research
2009

DOI (link to publisher)
10.1007/s11136-009-9451-9

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Mokkink, L. B., Terwee, C. B., Stratford, P., Alonso, J. R., Patrick, D. L., Riphagen, I., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Quality of Life Research, 18(3), 313-333. https://doi.org/10.1007/s11136-009-9451-9

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 27. Apr. 2022
Evaluation of the methodological quality of systematic reviews of health status measurement instruments

Lidwine B. Mokkink · Caroline B. Terwee · Paul W. Stratford · Jordi Alonso · Donald L. Patrick · Ingrid Riphagen · Dirk L. Knol · Lex M. Bouter · Henrica C. W. de Vet

Accepted: 29 January 2009 / Published online: 24 February 2009 © The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract A systematic review of measurement properties of health-status instruments is a tool for evaluating the quality of instruments. Our aim was to appraise the quality of the review process, to describe how authors assess the methodological quality of primary studies of measurement properties, and to describe how authors evaluate results of the studies. Literature searches were performed in three databases. One hundred and forty-eight reviews were included. The purpose of included reviews was to identify health status instruments used in an evaluative application and to report on the measurement properties of these instruments. Two independent reviewers selected the articles and extracted the data. Reviews were often of low quality: 22% of the reviews used one database, the search strategy was often poorly described, and in many cases it was not reported whether article selection (75%) and data extraction (71%) was done by two independent reviewers. In 11 reviews the methodological quality of the primary studies was evaluated for all measurement properties, and of these 11 reviews only 7 evaluated the results. Methods to evaluate the quality of the primary studies and the results differed widely. The poor quality of reviews hampers evidence-based selection of instruments. Guidelines for conducting and reporting systematic reviews of measurement properties should be developed.

Keywords Systematic review · Measurement properties · Methodological quality · Health status

L. B. Mokkink (✉) · C. B. Terwee · D. L. Knol · L. M. Bouter · H. C. W. de Vet
Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands
E-mail: w.mokkink@vumc.nl
URL: www.emgo.nl; www.cosmin.nl

P. W. Stratford
School of Rehabilitation Science and Department of Clinical Epidemiology and Biostatistics, McMaster University, 1400 Main St. West (4th Floor, Rm 403), Hamilton, ON L8S 1C7, Canada

J. Alonso
Health Services Research Unit, Institut Municipal d’Investigacio Medica (IMIM-Hospital del Mar), Carrer del Doctor Aiguader, 88, Edifici PRBB, 08003 Barcelona, Spain

J. Alonso
CIBER en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

D. L. Patrick
Department of Health Services, University of Washington, Box 357660 H 689, 1934 NE Pacific, Seattle, WA 98195-7660, USA

I. Riphagen
Unit of Applied Clinical Research, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

I. Riphagen
University Library, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

L. M. Bouter
Executive Board of VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
Introduction

Thousands of health status measurement instruments are used in research and clinical practice, and there are often many instruments for one single concept. Researchers, doctors, and policy-makers use the results obtained by instruments for further research, evidence-based patient care, guideline development, and evidence-based policy making.

The choice of an instrument depends on several factors, one of the most important being the measurement properties. The decision in favor of an instrument may have important consequences. Marshall et al. [1] showed that in schizophrenia trials authors were more likely to report that treatment was superior to control when an unpublished instrument was used in the comparison, rather than a published instrument. Furthermore, the selection of instruments with good measurement properties will lead to the detection of smaller treatment effects, or more power to draw stronger conclusions, and therefore to better interpretation of study results. In other words, if the measurement error of an instrument is small in relation to its minimal important change (MIC), one will be able to conduct clinical trials with relatively small sample sizes [2].

A systematic review of measurement properties critically appraises and compares the content and measurement properties of all instruments measuring a certain construct. High-quality systematic reviews of measurement properties provide evidence for the selection of the best instruments. The methodological quality of such a review should be thoroughly appraised in order to be confident that the design, conduct, analysis, and interpretation of the review was adequate, and to reveal any possible bias that might influence its conclusions. In general the critical appraisal of a systematic review consists of five steps: (1) reporting of relevant descriptive information, e.g., the target population, concept of interest, and the number of studies or instruments included, (2) appraisal of the quality of the review process, (3) appraisal of the methods used by the authors of reviews to assess the methodological quality of the primary studies included in the review, (4) appraisal of the results of the primary studies, and (5) a synthesis of the above-mentioned data (steps 3 and 4) to come to an overall conclusion for each instrument.

Existing guidelines for the appraisal of systematic reviews of clinical trials (e.g., Cochrane Collaboration [3] or AMSTAR [4]) or diagnostic studies [5, 6] can be used to appraise the quality of the systematic review process (step 2). These guidelines contain items on the quality of the search strategy [4], article selection and data extraction [3, 7, 8], and inclusion and exclusion criteria [6]. The methodological quality of systematic reviews of measurement properties has not been systematically assessed yet.

Authors of reviews should appraise the methodological quality and results of the primary studies [3] (steps 3 and 4). Accepted guidelines are available to appraise the methodological quality of clinical trials (e.g., Delphi List [9]) or diagnostic studies (QUADAS [10]). Several guidelines have been developed to appraise the methodological quality of studies on measurement properties [e.g., 11–13]. It is unknown which of these guidelines are used most often in systematic reviews of measurement properties.

It was our aim (1) to find all existing systematic reviews of measurement properties, (2) to appraise the quality of the review process of these reviews, (3) to describe if and how the authors of reviews assessed the methodological quality of the primary studies included in these reviews, (4) to describe if and how the authors of reviews evaluated the results of the primary studies, and (5) to describe if authors of reviews synthesized the above-mentioned data (steps 3 and 4) to come to an overall conclusion regarding the quality of each instrument.

Methods

Identification of reviews

To identify systematic reviews of measurement properties, we searched PubMed (up to March 2007), EMBASE (up to March 2007), and PsycINFO (up to June 2005). The full search strategies can be found in Appendix 1. Additional articles were identified by manually searching references from the retrieved articles and the authors’ own literature.

We included articles that

- Claimed to be “systematic reviews”
- Aimed to identify all available health status measurement instruments in a particular population, as stated by the author
- Concern health status measurement instruments that have been applied in an evaluative situation, i.e., instruments aimed to measure changes in health status over time in a longitudinal study
- Aimed to report on or evaluate the measurement properties of the measurement instruments

Based on guidelines for systematic reviews of back and neck pain trials [8], we considered a review to be systematic if at least one search in an electronic database was performed. We considered the following concepts to represent “health status” based on the model of Wilson and Cleary [14]: biological and physiological processes, symptoms, functional status (i.e., both physical functioning and psychosocial functioning), or general health perceptions. We consider health-related quality of life (HR-QoL)
as general health perception, and we excluded overall QoL. We excluded reviews that focused only on instruments applied in a discriminative situation, because these reviews are likely to have missed instruments that were used only in evaluative applications. We also excluded reviews that focused on instruments with a diagnostic or screening, or prognostic purpose.

Our aim was to find reviews that intended to find all available instruments for measuring a particular construct. We therefore excluded reviews of only one, or only the most commonly used instruments, or reviews that only included randomized clinical trials (RCTs). Reviews of RCTs very likely do not include all instruments that measure the construct of interest. Reviews that only described the instruments (e.g., format) were excluded. Only reviews written in English were included.

To determine the eligibility of the articles, two authors (L.M. and C.T.) independently reviewed title and abstract of every record retrieved from the searches. Full articles were retrieved for further assessment when the abstract suggested that the study might meet the inclusion criteria. Disagreements were resolved through consensus. A third reviewer (H.V.) was consulted in case of persisting disagreement.

Data extraction

Two authors (L.M. and C.T.) independently extracted data on (1) descriptive information, (2) the quality of the review process, (3) if and how the authors of reviews assessed the methodological quality of the primary studies included in the review, (4) if and how the results of the primary studies were evaluated and compared, and (5) if authors of reviews synthesized data to come to an overall conclusion on the quality of each instrument. Note that we only critically appraise the review process, and we simply describe if and how authors of reviews evaluate primary studies. A standard data extraction form was used (Appendix 2).

Descriptive information on reviews

Descriptive information that we extracted included year of publication, description of the health status concept of interest, study population of interest, number of health status instruments included, and type of health status instruments, i.e., patient-reported outcomes (PROs), proxy-reported outcomes or non-PROs. PRO was defined as a measurement of any aspect of a patient’s health status that comes directly from the patient, i.e., without the interpretation of the patient’s responses by a physician or anyone else [15]. Modes of data collection in PRO instruments include interviewer-administered instruments, self-administered instruments, computer-administered instruments or interactively administered instruments [16]. Proxy-reported outcomes include any endpoint obtained from a proxy, such as parent-assessed ratings measuring health-related quality of life in childhood acute lymphoblastic leukaemia (ALL) [17], or reports of a caregiver measuring pain in nonverbal older adults with advanced dementia [18]. Non-PROs are instruments that are based on other sources than patient or proxy reports, such as performance-based instruments [19], or clinical ratings, for example, to measure the severity of asthma in preschool children [20]. Finally, we extracted which measurement properties were reported in each review, and how they were reported, i.e., whether the exact results were reported or only the references to the publications.

Appraisal of the review process

To appraise the quality of the review process, we recorded whether the search strategy was described, which databases were searched, whether article selection and data extraction were performed by at least two persons, and whether inclusion and exclusion criteria for primary studies were described.

Description of the assessment of the methodological quality of primary studies

To describe if and how the methodological quality of the primary studies was assessed by the authors of the reviews, we recorded whether the methodological quality of each primary study was evaluated, i.e., if standards were applied to the primary studies. Standards refer to the study design and statistical analyses. An example of a standard for reliability is “rating ‘+’,” when an intraclass correlation coefficient (ICC) was used.” If one or more standards were applied, we recorded for which measurement properties standards were applied, which standards were applied, and whether they were described completely, i.e., were reproducible.

Description of the evaluation of the results of primary studies

To describe if and how the results of the primary studies were assessed by the authors of the reviews, we recorded whether they applied criteria of adequacy for what constitutes good measurement properties. An example is “ICC should be at least 0.70.” We recorded whether the results were evaluated and, if so, for which measurement properties, which criteria were applied, and whether they were completely described, i.e., were reproducible.
Description of synthesizing the methodological quality and the results

We furthermore documented two characteristics regarding whether or not authors of reviews formulated an overall conclusion for each instrument: we recorded whether authors gave a total score for the quality of each health status instrument, and we recorded whether some order of importance of the measurement properties was taken into account when giving a total score (see also Appendix 2).

Results

Identification of reviews

The searches yielded 7,779 records. We included 148 systematic reviews of measurement properties (Fig. 1). Most of the excluded articles did not meet the inclusion criteria of being a systematic review of measurement properties of all available health status instruments; for example, we excluded reviews of only a selection of existing instruments, reviews of health status instruments used only in randomized clinical trials (RCTs), and reviews in which measurement properties were not reported or evaluated.

Publication of systematic reviews of measurement properties has increased from less than one review per year in the 1990s up to 31 in 2005 (Fig. 2). The decrease in the number of reviews published in 2006 is possibly due to a delay in the recording of articles in PubMed and EMBASE. The concepts of interest in the included systematic reviews were general health perceptions (43%), functional status (21%), symptoms (17%), and biological and physiological processes (5%). The other reviews (14%) focused on a combination of these concepts. The reviews focused on a variety of populations, such as children, general population or patient populations with specific diseases, such as cerebral palsy or multiple sclerosis, or disease groups, such as cancer, neurological diseases or rheumatic disorders. Information about the study population and the number and type of instruments included in each review is presented in Table 1.

Appraisal of the review process

Table 2 shows the results of the quality assessment of the review process of the systematic reviews with regard to the description of the search strategy, the databases used, the article selection and data extraction, and the description of inclusion and exclusion criteria. In 84% of the reviews the authors described the search strategy in some way. This varied from describing only the most important keywords to reporting the full search strategy, including MeSH terms and text words for each database. The search strategies were often limited. For example, only MeSH headings were used, and no free text words [21, 22]; or only a few synonyms were used, for example, only “measur* or assess*”; words such as “question*”, “self-report”, “test”, “scale”, “outcome” or “interview” were not used [23]. In some reviews only the text words “psychometrics” [24] or “clinimetrics” [25] were used. Furthermore, the use of truncation was poorly described in most reviews. Finally, in quite a few reviews (14%) the time period during which the databases were searched, and some reviews (7%) searched a period of only 10 years or less was not specified.
Table 1  Descriptive information of the included systematic reviews of measurement properties

| Reference  | Population                        | Health status concept                  | Year of publication | PRO<sup>a</sup> | Proxy<sup>b</sup> | Non-PRO<sup>c</sup> | Number of instr. <sup>d</sup> |
|------------|-----------------------------------|----------------------------------------|---------------------|-----------------|------------------|---------------------|--------------------------|
| Pickard [17]| Childhood acute lymphoblastic     | HR-QoL (health-related quality of life)| 2004                | Yes             | Yes              | Yes                 | 20                       |
| Eiser [51] | Children                          | QoL (quality of life)                  | 2001                | Yes             | Yes              | No                  | 43                       |
| Pal [52]   | Children                          | Health status                          | 1996                | Yes             | Yes              | Yes                 | 9                        |
| Schmidt [53]| Children and adolescents         | HR-QoL                                 | 2002                | Yes             | Yes              | No                  | 16                       |
| Davis [54] | Children (0–12 years)             | HR-QoL and QoL                         | 2006                | Yes             | Yes              | Yes                 | 38                       |
| Hunter [55]| Children and adolescents         | Mental health                          | 1996                | Yes             | Yes              | Yes                 | 19                       |
| Brouwer [56]| Children with otitis media (0–18 years) | HR-QoL                              | 2005                | No              | Yes              | Yes                 | 15                       |
| Haywood [57]| People aged 60 years and over    | HR-QoL                                 | 2005                | Yes             | Yes              | No                  | 18                       |
| Haywood [39]| Older people aged 60 years and over | HR-QoL                              | 2005                | Yes             | Yes              | No                  | 15                       |
| Haywood [58]| Older people                     | HR-QoL                                 | 2006                | Yes             | Yes              | No                  | 45                       |
| Hollifield [59]| Refugees                        | Health status (mental and physical), trauma, quality of care, and diagnostic | 2002                | Yes             | No              | No                  | 12                       |
| Haywood [60]| Ankylosing spondylitis (AS)      | Health or HR-QoL                       | 2005                | Yes             | No              | No                  | 15                       |
| Namjoshi [61]| Bipolar disorder                 | HR-QoL                                 | 2001                | Yes             | No              | No                  | 14                       |
| Michalak [62]| Bipolar disorder                 | HR-QoL                                 | 2005                | Yes             | No              | No                  | 8                        |
| Okamoto [63]| Breast cancer                    | QoL                                    | 2003                | Yes             | No              | No                  | 11                       |
| Edwards [64]| Caregivers of patients with cancer | QoL                                  | 2002                | Yes             | No              | No                  | 4                        |
| Ringash [65]| Head and neck cancer            | Disease-specific HR-QoL                | 2001                | Yes             | No              | No                  | 11                       |
| Van Korlaar [66]| Chronic venous disease       | QoL                                    | 2003                | Yes             | No              | No                  | 16                       |
| Neelakantan [35]| Women with chronic pelvic pain | HR-QoL                                 | 2004                | Yes             | No              | No                  | 30                       |
| Riemsma [67]| Cognitive impairment due to acquired brain injury | General health status                  | 2001                | Yes             | Yes              | No                  | 34                       |
| Jones [68]| Common chronic, benign gynecologic conditions | HR-QoL                                 | 2002                | Yes             | No              | No                  | 14                       |
| Ettema [69]| Dementia                         | QoL                                    | 2005                | Yes             | Yes              | Yes                 | 17                       |
| Salek [31]| Dementia/Alzheimer’s            | QoL                                    | 1998                | Yes             | Yes              | No                  | 9                        |
| Walker [32]| Dementia/Alzheimer’s            | QoL                                    | 1998                | Yes             | Yes              | Yes                 | 19                       |
| De Tiedra [70]| Dermatology                    | HR-QoL                                 | 1998                | Yes             | No              | No                  | 23                       |
| Garratt [71]| Diabetes mellitus (type 1 and 2) | Disease-specific HR-QoL                | 2002                | Yes             | No              | No                  | 9                        |
| Luscombe [72]| Diabetes mellitus type 2       | HR-QoL                                 | 2000                | Yes             | No              | No                  | 31                       |
| Cagney [73]| End-stage renal disease (ESRD)   | QoL                                    | 2000                | Yes             | No              | No                  | 53                       |
| Edgell [74]| End-stage renal disease (ESRD) patients | HR-QoL                              | 1996                | Yes             | No              | Yes                 | 16                       |
| Kline [75]| Epilepsy and antiepileptic drug (AED) treatment | Condition specific HR-QoL               | 1998                | Yes             | No              | No                  | 4                        |
| Leone [76]| Epilepsy (adults)               | HR-QoL                                 | 2005                | Yes             | No              | No                  | 45                       |
| Szende [77]| Hemophilia                       | HR-QoL and health status               | 2003                | Yes             | Yes              | No                  | 19                       |
| De Kleijn [78]| Hemophilia (age >16 years)     | Health status; body structure, body function, activities | 2002                | Yes             | No              | Yes                 | 34                       |
| De Boer [27]| HIV infected                    | HR-QoL                                 | 1995                | Yes             | No              | No                  | 12                       |
| Clayson [79]| HIV/AIDS                        | HR-QoL                                 | 2006                | Yes             | No              | No                  | 34                       |
| Bonomi [80]| Acute, chronic, and cancer pain | QoL, utility instruments               | 2000                | Yes             | No              | No                  | 18                       |
| Symonds [81]| Incontinency                    | HR-QoL                                 | 2003                | Yes             | No              | No                  | 10                       |
| Pallis [82]| Inflammatory bowel disease (IBD) | HR-QoL                                 | 2000                | Yes             | No              | ?                   | 12                       |
| Cummins [83]| Intellectual disability        | QoL                                    | 1997                | Yes             | No              | No                  | 13                       |
| Garratt [84]| Knee problems                   | Health and QoL                         | 2004                | Yes             | No              | No                  | 16                       |
| Reference       | Population                                                  | Health status concept                                      | Year of publication | PROa | Proxyb | Non-PROc | Number of instr.d |
|-----------------|-------------------------------------------------------------|------------------------------------------------------------|---------------------|------|--------|----------|------------------|
| Zanoli [85]     | Lumbar disorders                                            | HR-QoL                                                     | 2000                | Yes  | No     | No       | 92               |
| Clark [86]      | Menorrhagia                                                 | QoL                                                        | 2002                | Yes  | No     | No       | 30               |
| Van Nieuwenhuizen [87] | Mental illness (severe)                                      | QoL                                                        | 1997                | Yes  | Yes    | Yes      | 11               |
| Lehman [88]     | Mental illnesses (severe and persistent)                   | QoL                                                        | 1996                | Yes  | No     | No       | 10               |
| Gruenewald [24] | Multiple sclerosis (severe)                                 | HR-QoL                                                     | 2004                | Yes  | Yes    | No       | 23               |
| Marinus [89]    | Parkinson’s disease                                         | QoL                                                        | 2002                | Yes  | No     | No       | 4                |
| Heffernan [90]  | Three degenerative neurological conditions: multiple sclerosis (MS), Parkinson’s disease, and motor neuron disease (MND) | Disease-specific health status                             | 2005                | Yes  | No     | No       | 16               |
| Jørstad [91]    | Population over 50 years who had not suffered a stroke or Parkinson’s disease, or had undergone a lower limb amputation | Fall-related psychological outcome measures               | 2005                | Yes  | No     | ?        | 26               |
| Rannard [92]    | Primary biliary cirrhosis (PBC)                            | HR-QoL                                                     | 2004                | Yes  | No     | Yes      | 20               |
| De Korte [93]   | Psoriasis                                                   | QoL                                                        | 2002                | Yes  | No     | No       | 6                |
| Lewis [94]      | Psoriasis                                                   | HR-QoL                                                     | 2005                | Yes  | Yes    | No       | 14               |
| Hallin [95]     | Spinal cord injury (SCI)                                    | QoL                                                        | 2000                | Yes  | No     | No       | 14               |
| Matza [96]      | Stress urinary incontinence or overactive bladder (OAB)     | Condition-specific HR-QoL                                  | 2004                | Yes  | No     | ?        | 16               |
| Golomb [46]     | Stroke                                                      | HR-QoL including functioning and well-being               | 2001                | Yes  | No     | No       | 32               |
| Buck [97]       | Stroke                                                      | QoL                                                        | 2000                | Yes  | No     | No       | 25               |
| Drake [26]      | Total knee arthroplasty (TKA)                               | Global patient rating scales                              | 1994                | No   | No     | Yes      | 34               |
| Prasad [98]     | Working adults                                              | Health-related work outcomes                              | 2004                | Yes  | No     | No       | 12               |
| Lofland [99]    | Various                                                     | Health-related loss in work productivity                  | 2004                | Yes  | No     | No       | 11               |
| De Boer [34]    | Vision impairments                                          | Vision-related QoL                                         | 2004                | Yes  | No     | No       | 31               |
| Lundström [100] | Sight-threatening eye disease                               | HR-QoL/vision-related QoL                                  | 2006                | Yes  | No     | No       | 16               |
| Tripop [101]    | Glaucoma                                                    | HR-QoL                                                     | 2005                | Yes  | No     | No       | 10               |
| Franie [102]    | Voice disorders                                             | QoL                                                        | 2005                | Yes  | No     | No       | 9                |
| Morley [103]    | Chronic rhinosinusitis (patients undergoing endoscopic sinus surgery for) | HR-QoL                                                     | 2006                | Yes  | No     | No       | 20               |
| Watt [104]      | Benign thyroid disorders                                    | HR-QoL                                                     | 2006                | Yes  | No     | No       | 6                |
| Functional status (physical and psychosocial)             |                                                             |                                                            |                     |      |        |          |                  |
| Ketelaar [105]  | Children with cerebral palsy                               | Disability                                                 | 1998                | No   | No     | Yes      | 17               |
| Boyce [106]     | Children with cerebral palsy                               | Motor performance or quality of movement                   | 1991                | No   | No     | Yes      | 10               |
| Buffart [107]   | Children with congenital (unilateral) transverse or longitudinal reduction deficiencies of the upper limb | Arm/hand functioning                                      | 2006                | Yes  | Yes    | Yes      | 23               |
| Pakulis [108]   | Adolescent sarcoma patients (bone tumor)                   | Physical functioning                                      | 2005                | Yes  | No     | Yes      | 7                |
| Moore [109]     | English-speaking adult population                           | Functional living skills                                  | 2006                | No   | No     | Yes      | 31               |
| Arrington [21]  | Chronic medical or general populations                     | Sexual function                                           | 2004                | Yes  | No     | No       | 57               |
| MacKnight [110] | Community-dwelling elderly                                 | Performance-based mobility                                | 1995                | No   | No     | Yes      | 41               |
| Wind [111]      | Healthy and disabled subjects                              | Functional capacity                                       | 2005                | Yes  | No     | Yes      | 27               |
| Ramaker [25]    | Parkinson’s disease                                         | Impairment and disability                                | 2002                | No   | No     | Yes      | 11               |
| Mannerkorpi [112] | Fibromyalgia syndrome (FMS)                               | Functional limitations and disability                     | 1997                | Yes  | No     | Yes      | 15               |
| Reference | Population | Health status concept | Year of publication | PRO\(^a\) | Proxy\(^b\) | Non-PRO\(^b\) | Number of instr.\(^d\) |
|-----------|------------|-----------------------|---------------------|----------|-----------|-------------|------------------|
| Grotle [28] | Low back pain | Functional status and disability | 2004 | Yes | No | No | 36 |
| Millard [113] | Chronic pain | Pain-related disability | 1997 | Yes | No | No | 35 |
| Dowrick [114] | Musculoskeletal disorders of the upper extremity/orthopaedic trauma population (e.g., fracture or dislocation) | Functional outcomes | 2005 | Yes | No | Yes | 7 |
| Law [29] | Occupational therapy | Functional ability in activities of daily living (ADL) | 1989 | Yes | No | Yes | 13 |
| Terwee [19] | Osteoarthritis of the hip or knee | Physical function | 2006 | No | No | Yes | 26 |
| Dziedzic [115] | Osteoarthritis of the hand | Hand disability/functional disability | 2005 | Yes | No | No | 5 |
| Swinkels [116] | Rheumatic disorders | Personal care disabilities | 2005 | Yes | No | No | 19 |
| Swinkels [117] | Patients with rheumatic disorders | Impairments in functions | 2005 | Yes | No | Yes | 49 |
| Swinkels [118] | Rheumatic disorders | Impairment | 2005 | Yes | No | Yes | 42 |
| Swinkels [119] | Rheumatoid arthritis | Disabilities in gait and gait-related activities | 2004 | Yes | No | No | 61 |
| McKibbin [120] | Seriously mentally ill, schizophrenia | Functioning | 2004 | No | No | Yes | 8 |
| Keskula [121] | Shoulder conditions (athletes) | Functional limitations and disability | 2001 | Yes | No | Yes | 9 |
| Michener [122] | Shoulder dysfunction | Functional limitations and disability | 2001 | Yes | No | No | 11 |
| Bot [41] | Shoulder or shoulder-upper limb problems | Shoulder disability | 2004 | Yes | No | No | 16 |
| Salerno [123] | Disorders of the neck and upper extremity (mild to moderate) | Functional status | 2002 | Yes | No | No | 13 |
| Chong [124] | Stroke | Instrumental activities of daily living (IADL) | 1995 | Yes | No | Yes | 4 |
| Croarkin [125] | Stroke | Upper extremity motor function tests | 2004 | No | No | Yes | 9 |
| McGee [126] | Cardiac rehabilitation | Psychological outcome: depression, anxiety, and other negative affective states | 1999 | Yes | No | ? | 13 |
| Sakzewski [127] | Children with cerebral palsy (5–13 years) | Participation | 2007 | Yes | Yes | Yes | 7 |
| Morris [128] | Children with cerebral palsy (5–15 years) | Activity performance and participation as defined by ICF | 2005 | Yes | Yes | No | 7 |
| Eadie [129] | Speech-language pathology | Communicative (functioning) participation | 2006 | Yes | No | No | 6 |

**Symptoms**

| Reference | Population | Health status concept | Year of publication | PRO\(^a\) | Proxy\(^b\) | Non-PRO\(^b\) | Number of instr.\(^d\) |
|-----------|------------|-----------------------|---------------------|----------|-----------|-------------|------------------|
| Brooks [130] | Adolescents | (Diagnose or measure) anxiety symptoms | 2003 | Yes | Yes | Yes | 9 |
| Duhn [131] | Infants | Pain assessment | 2004 | No | Yes | Yes | 35 |
| Ramelet [132] | Children (0–3 years) | Pain | 2004 | No | ? | Yes | 28 |
| Stinson [133] | Children and adolescents | Pain | 2006 | Yes | No | No | 7 |
| Eccleston [134] | Adolescents (11–18 years) | (Impact of) pain | 2005 | Yes | Yes | No | 43 |
| Birken [135] | Preschool children (0–6 years) | Clinical asthma severity | 2004 | No | No | Yes | 10 |
| Linder [136] | Children with cancer (0–18 years) | Physical symptoms | 2005 | Yes | Yes | No | 23 |
| Stover [137] | Children less than 6 years old | PTSD symptoms and diagnostic measures | 2005 | Yes | Yes | Yes | 7 |
| Devine [138] | Adults | Sleep dysfunction | 2005 | Yes | No | Yes | 22 |
| Kirkova [139] | Adult cancer patients | Cancer symptoms | 2006 | Yes | Yes | Yes | 21 |
| Reference       | Population                                                                 | Health status concept                                      | Year of publication | PRO\(^a\) | Proxy\(^b\) | Non-PRO\(^c\) | Number of instr.\(^d\) |
|-----------------|-----------------------------------------------------------------------------|------------------------------------------------------------|----------------------|------------|-------------|----------------|------------------------|
| Vadaparampil [140] | Adults with hereditary breast, ovarian, and colon cancer                   | Psychological factors (depression, anxiety or distress)    | 2005                 | Yes        | No          | No             | 11                     |
| Van Herk [141]   | Older adults with severe cognitive impairments or communication difficulties | Pain                                                        | 2007                 | No         | No          | Yes            | 13                     |
| Stolee [142]     | Cognitively impaired older persons                                          | Pain                                                        | 2005                 | Yes        | No          | Yes            | 30                     |
| Zwakhalen [143]  | Elderly people with dementia                                               | Pain                                                        | 2006                 | No         | No          | Yes            | 12                     |
| Herr [144]       | Nonverbal older adults with dementia                                        | Pain                                                        | 2006                 | No         | No          | Yes            | 10                     |
| Smith [18]       | Nonverbal older adults with advanced dementia                               | Pain                                                        | 2005                 | No         | Yes         | Yes            | 7                      |
| Schofield [145]  | Adults with cognitive impairment                                           | Pain                                                        | 2005                 | Yes        | Yes         | Yes            | 9                      |
| Schuurmans [146] | Delirium                                                                    | Delirium (symptom severity)                                 | 2003                 | Yes        | No          | Yes            | 8                      |
| Stanghellini [22] | Gastro-oesophageal reflux disease (GERD)                                    | Symptom scales                                              | 2004                 | Yes        | No          | Yes            | 20                     |
| Fraser [147]     | Gastro-oesophageal reflux disease (GERD) or dyspepsia                       | Frequency or severity of GERD or dyspepsia symptoms         | 2005                 | Yes        | No          | No             | 26                     |
| Bouchard [148]   | Panic, panic disorders, agoraphobia                                         | Aspects of panic attacks or panic disorder                  | 1997                 | Yes        | ?           | ?              | 14                     |
| Dorman [36]      | Patients in palliative care                                                | Breathlessness                                              | 2007                 | Yes        | ?           | ?              | 29                     |
| Bausewein [149]  | Chronic conditions such as OPD, cancer, chronic heart failure, and motor neuron disease | Breathlessness                                              | 2007                 | Yes        | No          | Yes            | 33                     |
| Dittner [150]    | Various                                                                     | Fatigue                                                     | 2004                 | Yes        | No          | ?              | 30                     |
| Mota [151]       | Adults                                                                      | Fatigue                                                     | 2006                 | Yes        | No          | No             | 18                     |
| Biological and physiological processes | Preschool children (0–5 years)                                             | Clinical scores for acute asthma                            | 1994                 | No         | No          | Yes            | 8                      |
| Moreau [152]     | Low back pain                                                               | Isometric back extension endurance                          | 2001                 | No         | No          | Yes            | 6                      |
| Charman [153]    | Atopic eczema                                                               | Disease-specific objective skin examination scales (severity)| 2000                 | No         | No          | Yes            | 13                     |
| Sun [154]        | Osteoarthritis of hip and knee joints                                       | Clinical rating systems                                     | 1997                 | Yes        | No          | Yes            | 45                     |
| Innes [155]      | General population/occupational therapy                                     | Grip strength                                               | 1999                 | No         | No          | Yes            | 13                     |
| Kettler [156]    | People with cervical and lumbar disc and facet joint degeneration          | Grading systems                                             | 2006                 | No         | Yes         | No             | 42                     |
| Hudson [157]     | Systemic sclerosis                                                          | Disease activity in systemic sclerosis                      | 2007                 | No         | No          | Yes            | 3                      |
| Combination      | General population                                                          | Sexual function, satisfaction or quality of life            | 2002                 | Yes        | No          | No             | 23                     |
| Cremeens [158]   | Children (3–8 years)                                                        | QoL, self-esteem, self-concept, and mental health measures  | 2006                 | Yes        | No          | No             | 53                     |
| Hayes [159]      | Critical care survivors                                                     | Impairment, functional status, and HR-QoL outcome measures | 2000                 | Yes        | No          | Yes            | 36                     |
| Pietronbon [160] | Various                                                                     | Neck pain or dysfunction                                    | 2002                 | Yes        | No          | No             | 5                      |
| Linder [161]     | Acute sinusitis                                                             | HR-QoL and symptom scores                                   | 2003                 | Yes        | No          | No             | 21                     |
| Hearn [162]      | Cancer (advanced)                                                           | Outcome measures                                            | 1997                 | Yes        | Yes         | Yes            | 12                     |
| Eechaute [42]    | Chronic ankle instability                                                   | Patient-assessed instruments                                | 2007                 | Yes        | No          | No             | 3                      |
Description of the assessment of the methodological quality of the primary studies and evaluation of the results

In 44% \((n = 65/148)\) of the reviews the methodological quality of the included studies was not assessed and the results were not appraised, but only reported, i.e., steps 3 and 4 were omitted.

Of these reviews, 32% \((n = 21/65)\) only reported references of the primary studies and not the results; 38% \((n = 25/65)\) reported the results, 28% \((n = 18/65)\) reported partly results and partly references, and 2% \((n = 1/65)\) stated that no studies of measurement properties were found for any of the included instruments \([26]\). References were mainly reported for validity, and results for reliability.

In 56% \((n = 83/148)\) of the reviews the methodological quality of the included studies was (partly) assessed by the authors of the reviews and (some of) the results were evaluated, i.e., standards and/or criteria of adequacy were applied to one or more measurement properties (steps 3 and 4). In 53% \((n = 44/83)\) of these reviews (some) standards as well as criteria of adequacy were applied. In 46% \((n = 38/83)\) of these reviews only (some) criteria of adequacy were applied, and in one review only standards were applied.

Often a limited number of standards and/or criteria of adequacy were applied; for example, in some cases only a standard and a criterion for internal consistency were used \([27]\). Eleven reviews described and applied a complete set of standards, i.e., fully described and reproducible standards of reliability, validity, and responsiveness.

---

**Table 1 continued**

| Reference | Population                                      | Health status concept                        | Year of publication | PRO\(^a\) | Proxy\(^b\) | Non-PRO\(^c\) | Number of instr.\(^d\) |
|-----------|-------------------------------------------------|----------------------------------------------|---------------------|------------|-------------|---------------|----------------------|
| Dorey [38] | Erectile dysfunction                           | Outcome measure                              | 2002                | Yes        | Yes         | Yes           | 26                   |
| Veenhof [30] | Hip and/or knee OA           | Pain, physical function, and patient global assessment | 2006                | Yes        | No          | No            | 32                   |
| Razvi [163] | Hypothyroidism (adult)                  | Symptoms, health status, and QoL            | 2005                | Yes        | No          | Yes           | 9                    |
| Bijkerk [164] | Inflammatory bowel syndrome (IBS)   | HR-QoL or symptoms                           | 2003                | Yes        | No          | ?             | 10                   |
| Haywood [165] | Lateral ligament injury of the ankle | Multi-item measures of health outcome         | 2004                | Yes        | No          | Yes           | 9                    |
| Costa [43] | Low back pain                                 | Outcome measures                             | 2007                | Yes        | No          | No            | 15                   |
| Poolsup [166] | Mania                        | Global rating scales and symptom rating scales | 1999                | Yes        | No          | Yes           | 13                   |
| Platz [23] | Spasticity                                   | Clinical phenomena, function                 | 2005                | Yes        | No          | Yes           | 37                   |
| D’Olhaberria-gue [167] | Stroke                  | Neurological examination; deficit or handicap and disability | 1996                | No         | No          | Yes           | 14                   |
| Van Tuijl [40] | Tetraplegics                  | Upper extremity tests: strength tests, functional tests, and ADL tests | 2002                | Yes        | No          | Yes           | 24                   |
| Margolis [168] | Visually impairments       | Vision-specific HR-QoL or functioning or impact | 2002                | Yes        | No          | No            | 22                   |
| Ashcroft [169] | Psoriasis                        | Clinical outcome measures to evaluate severity of psoriasis and its response to treatment | 1999                | Yes        | No          | Yes           | 7                    |
| Avery [37] | Urinary and anal incontinence and vaginal and pelvic problems | QoL and symptoms                              | 2007                | Yes        | No          | No            | 23                   |
| Bialocerkowski [170] | Wrist complaints            | Wrist outcome instruments, performance or function | 2000                | Yes        | No          | Yes           | 32                   |

\(^a\) Patient-reported outcomes

\(^b\) Proxy-reported outcomes

\(^c\) Non-patient-reported outcomes, such as clinical ratings and performance-based outcomes

\(^d\) Number of instruments included in the systematic review

*ICF* International Classification of Functioning, Disability and Health, *PTSD* Post-traumatic stress disorder
Twelve reviews described and applied a complete set of criteria of adequacy, i.e., fully described and reproducible criteria of adequacy of reliability, validity, and responsiveness. In seven reviews both a complete set of standards and a complete set of criteria of adequacy were described and applied.

In Table 3 we summarize the standards and criteria of adequacy used by the authors of the reviews. Standards were most often applied for reliability (use of an ICC), internal consistency (use of Cronbach’s alpha), and construct validity (confirming hypotheses). Criteria of adequacy were most often applied for reliability (e.g., ICC >0.70) and for internal consistency (Cronbach’s alpha >0.70). Standards and criteria of adequacy for measurement error and interpretability were rarely used. Few authors of reviews mentioned that the use of Pearson’s correlation coefficients was not adequate to measure reliability [19, 28, 29]. Only two reviews gave an exact number as a minimum of the sample size (i.e., at least 50) for reliability [19, 30] and two reviews required that the sample size for reliability must be “reasonably large” [31, 32]. Criteria for construct validity varied from qualitative criteria such as “hypotheses confirmed” to quantitative criteria such as “r ≥ 0.40.” Standards given for responsiveness included confirming hypotheses, effect sizes or standardized response mean or other methods.

Description of synthesizing methodological quality and results

In 7% (n = 10/148) of the systematic reviews a total score was given for the quality of each instrument, and in 5% (n = 8/148) of the systematic reviews an order of importance of measurement properties was taken into account when making the quality assessment. There was no agreement among the reviews regarding which property was most important. Some considered content validity as most important [33–35], while others considered construct validity [36], responsiveness [29, 36] or validity and reliability [37] as the most important measurement properties.

The reviews frequently used rating systems to indicate whether a standard or a criterion of adequacy was met. Different rating systems were used. An example of a nonspecified rating system is “0 = no numerical results reported; + = weak evidence; ++ = adequate evidence; +++ = good evidence” [38–40]. An example of a rating system in which the standard and the criterion are combined is “+ adequate design & method (i.e. factor analysis and Cronbach’s alpha), and alpha is between 0.70 and 0.90; ± doubtful method used (no factor analysis); − inadequate internal consistency (alpha <0.70); ? no information found on internal consistency” [30, 41, 42].

Discussion

It was our aim to identify all systematic reviews of measurement properties, to appraise the quality of the review process, and to describe whether the authors of the reviews appraised the methodological quality and results of the primary studies. We observed an increase in published systematic reviews of measurement properties in the last few years. Information required to assess the quality of the review process is often poorly described. More than half of the authors of the reviews evaluated neither the methodological quality of the primary studies nor the results of these studies. The reviews that did evaluate methodological quality and results used different standards and criteria of adequacy.

We attempted to use transparent and reproducible methods. However, because of the considerable variation in design, performance, and data presentation of the included reviews, some degree of judgement in appraising the quality of the systematic reviews and describing the standards and criteria was unavoidable.
### Internal consistency

**Standards**
- Cronbach’s alpha (18×)
- KR-20 (2×), kappa (1×)
- Cronbach’s alpha is calculated for either the whole scale or for subscales depending on the outcome of the factor analysis (5×)
- Rasch analysis (2×)
- Rating system not specified (2×)

**Criteria of adequacy**
- Alpha > 0.70 (26×)
- Alpha < 0.90 (9×), or not too high (1×)
- Alpha > 0.80 (3×)
- Alpha > 0.95 (2×)
- Range (e.g., 0.00–0.39 low; 0.40–0.59 moderate; 0.60–0.79 moderately high; 0.80–1.0 high, or alpha < 0.70 questionable; 0.71–0.80 moderate; >0.80 good) (10×)
- Distinction between cut-off score for group level and clinical use (2×)
- Rating system not specified (2×)

### Reliability

**Standards**
- ICC: (18×)
- Kappa (10×)
- Correlation coefficient (e.g., Pearson’s or Spearman) (11×)
- Correlation not adequate (3×)
- Time interval mentioned (3×)
- Other measures, e.g., MDC, CV, Kendall’s tau, t-test, Goodman-Kruskall gamma, odds ratio, percentage agreement (7×)
- Rating system not specified (7×)

**Criteria of adequacy**
- ICC > 0.70 (19×)
- ICC between 0.70 and 0.90 (7×)
- ICC > 0.50 (1×), >0.60 (2×), >0.75 (2×), >0.80 (3×), >0.90 (7×)
- Lower limit ICC > 0.60 (1×)
- Range ICC, kappa or r (18×)
- Distinction between, e.g., test-retest reliability and interrater reliability or discriminative versus evaluative use (3×)
- Minimum sample size (3×)
- Rating system not specified (13×)
- Example: Test-retest reliability: ICC < 0.6; ±ICC 0.6–0.8; +ICC > 0.8; Interobserver reliability: ICC < 0.5; ±ICC 0.5–0.7; +ICC > 0.7.

### Measurement error

**Standards**
- Bland & Altman 95% LoA (5×)
- SEM (5×)
- Kappa (3×)
- MDC (1×)
- SDD/SDC (2×)

**Criteria of adequacy**
- Rating system not specified (3×)
- LoA or SDC < M(C)IC (1×)

### Validity

**Standards**
- Rating system not specified (3×)

**Criteria of adequacy**
- Rating system not specified (12×)
- Correlation between 0.4 and 0.8 (1×)

### Content validity

**Standards and/or criteria of adequacy**
- Involvement of patients (7×)
- Judgement by reviewer (3×)
- Involvement of experts (4×)
- Examining the literature (2×)
- Statistical procedure (e.g., impact method, principal component analysis) (4×)
- Rating system not specified (3×)

### Construct validity

**Standards**
- Rating system not specified (3×)

**Criteria of adequacy**
- Rating system not specified (12×)
- Correlation between 0.4 and 0.8 (1×)
We identified three major aspects: a lack of methodological quality of systematic reviews of measurement properties, i.e., low quality of search strategy, a lack of good reporting of the methods used to perform the systematic review, and a lack of use of standards and criteria of adequacy to assess the methodological quality of the primary studies.

Appraisal of the review process

Firstly, the quality and reporting of the search strategy was often poor. It was obvious that search strategies were often too narrow and that many systematic reviews were likely to be incomplete; for example, Costa et al. [43] found 17 primary studies on the Roland Morris Disability Questionnaire (RDQ) by using a search strategy consisting of several terms for low back pain with the terms “questionnaire(s) OR outcome measure(s) OR index OR scale”. However, a simple PubMed search “Roland AND (responsive* OR sensitiv*)” resulted in 11 additional responsiveness studies of the RDQ that were not included in the review. Furthermore, the review of Costa was limited to a time period from January 2001 to July 2007. With our simple PubMed search described above, we found another 12 responsiveness studies of the RDQ before 2001.

We recommend that the search strategy consist of terms describing the concept to be measured, terms describing the population of interest, and terms describing the type of

Table 3 continued

| Confirming hypotheses (11×) | Calculation of correlation (8×) | Distinction between different forms of validity (e.g., convergent validity, divergent validity, known group validity) (6×) | Rating system not specified (3×) |
|---------------------------|-------------------------------|---------------------------------------------------------------------------------------------------------------------|-------------------------------|

**Criterion validity**

Standards (4×)

- Correlation of percentage agreement between instrument and “gold standard” (4×)
- Magnitude of the coefficients is hypothesis dependent (1×)

**Responsiveness**

Standards (17×)

- “Adequate measure” used, e.g., ES, SRM (7×)
- Confirming hypotheses (6×)
- Calculating change scores (3×)
- Other measures, e.g., ROC curves (1×), Guyatt index of responsiveness (1×), relative efficacy (1×), Student’s t-test/Wilcoxon’s test (1×)
- Rating system not specified (5×)

**Interpretability**

Standards and/or criteria of adequacy (7×)

- Presenting MIC/MCIC (4×)
- Presenting mean and SDs (e.g., for different subgroups, or before and after treatment) (4×)
- Rating system not specified (1×)

MDC minimal detectable change, CV coefficient of variation, LoA limits of agreement, SEM standard error of measurement, SDD/SDC smallest detectable difference/change, M(C)IC minimal (clinically) important change, ES effect size, SRM standardized response mean

a Standards refer to the study design and statistical analyses

b Number of reviews in which the standard/criterion is mentioned

c Criteria of adequacy refer to what constitutes good measurement properties
instruments of interest, such as questionnaire, performance-based measure, etc. For each of these parts a comprehensive list of possible synonyms should be used, preferably drawn up in cooperation with a clinical librarian. Platz et al. [23] published a systematic review that aimed to characterize clinical assessment methods for spasticity and/or functional consequences in clinical patient populations at risk to suffer from spasticity. Their search strategy was adequate. They started with search terms for the construct (i.e., spas*, hyperton* or reflex*), secondly they used terms for the type of instrument (i.e., measure* or assess*) and thirdly terms for the population of interest (i.e., stroke or CVA or multiple sclerosis or MS or spinal cord injury or SCI or cerebral palsy or CP). Additionally, we recommend not to limit the search to a specific time period.

In many search strategies the focus is on finding all health status instruments, without focusing on finding all studies of measurement properties of these instruments. An additional search strategy, including the names of the instruments, is often needed to find all these studies. In our experience these studies of measurement properties do not always contain terms of measurement properties such as “reliability,” “validity,” and “responsiveness” in the title, abstract or keywords. Furthermore, the large variety in terms of measurement properties used in the literature makes it difficult to design a sensitive search strategy. The use of a methodological search filter with terms for measurement properties will inevitably result in missing studies and should therefore be discouraged. This is in line with what is known about the performance of other methodological search filters, e.g., for finding diagnostic studies [44]. In 21% of the reviews only one database was used. In guidelines for systematic reviews of clinical trials [3, 8] and observational studies [45] it is suggested that limiting a search to a single database will not provide a thorough summary of the existing literature.

Secondly, there is a lack of adequate reporting of the methods used in the systematic reviews of measurement properties. Because of this, it is difficult to assess the methodological quality of the reviews. It was often unclear if things were not done (e.g., data extraction performed by at least two independent reviewers) or if they were not reported. For example, Law and Letts clearly described that the data extraction was performed by two people, but they did not describe if the article selection was also performed by two people [29]. As we only used information from the published reviews and did not contact authors to ask for additional information, it is possible that we may have slightly underrated the quality of the reviews. However, we believe that our article clearly shows the need for guidelines for assessing the quality of systematic reviews of measurement properties and guidelines for reporting on these reviews.

Description of the assessment of the methodological quality of primary studies and the evaluation of the results of primary studies

Thirdly, more than half of the reviews did not evaluate either the methodological quality of the primary studies (step 3), or the results of these studies (step 4), i.e., standards for the appropriateness of the study design and statistical analyses, and criteria for what constitutes good measurement properties were often not applied; for example, Golomb et al. [46] published a review on health-related quality-of-life measures in stroke. They provided definitions of the measurement properties and adequately described the results of the measurement properties for each of the available measurement instruments, but they did not apply a priori determined standards to the methods used to assess the measurement properties, or criteria of adequacy to the results of those studies.

In our opinion it is important to assess the methodological quality of included primary studies in order to decrease the risk of bias in the review. Considering the large variety of methods used to evaluate the methodological quality of the individual studies, there is a need for guidance. Within this guidance more attention should be paid to techniques based on item response theory (IRT). IRT has many advantages over classical test theory; for example, shorter questionnaires with equal or even better reliability can be developed [47]. Furthermore, the ability scores are test independent [48], and scores obtained on different instruments measuring the same construct can be linked, so that they are comparable [49]. We think that standards and criteria of adequacy are most likely to be widely used when consensus is reached among international experts about the preferred standards and criteria of adequacy. We therefore started the Consensus-based Standards for the selection of Health Measurement INstruments (COSMIN) initiative with the aim to draw up a consensus-based checklist for the evaluation of the methodological quality of studies on measurement properties [50].

Conclusion

A systematic review of measurement properties is a useful tool for evaluating the quality of an instrument, or for interpreting results based on an instrument. In the last few years the number of such systematic reviews published has increased enormously every year. However, the methodological quality of these reviews leaves much to be desired and should be improved. We feel it is essential to develop guidelines for the assessment of the methodological quality of systematic reviews of measurement properties. This includes guidelines for the review process, guidelines to assess the methodological quality of the studies that
evaluate measurement properties, and guidelines for criteria of adequacy for good measurement properties.

Acknowledgements This study is financially supported by the EMGO Institute, VU University Medical Center, Amsterdam, and the Anna Foundation, Leiden, The Netherlands. These funding organizations did not play any role in the study design, data collection, data analysis, data interpretation or publication.

Conflict of interest The authors of this review, except IR, are all members of the Steering Committee of the COSMIN study.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix 1: Search strategies

PubMed

(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])
AND
(systematic[sb] OR (literature AND search*) OR (medline AND search*) OR review[ti])
AND
(reproducibility of results[MeSH] OR Psychometrics[MeSH] OR Observer variation[MeSH] OR quality[ti] OR assess*[ti] OR validation studies[pt] OR evaluation studies[pt] OR reproduc*[tiab] OR relia*b*[tiab] OR intra-class correlation[tiab] OR internal consistency[tiab] OR valid*[tiab] OR responsive*[tiab] OR agreement[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR factor structure[tiab] OR discriminant analysis[tiab] OR ((clinimetric[tiab] OR psychometric[tiab]) AND (propert*[tiab] OR analys*[tiab])) OR (measurement[tiab] AND propert*[tiab]) OR ((minimal*[tiab] OR smallest[tiab]) AND (important[tiab] OR detectable[tiab] OR real[tiab]) AND (change[tiab] OR difference[tiab])))
NOT
(meta-analysis[pt] OR meta-analysis[ti] OR metaanalysis[ti] OR case reports[pt] OR ‘delphi-technique’[ti] OR cross-sectional[ti])

EMBASE (through Embase.com)

Bloc 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 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)

Bloc 2:

review:ti OR (literature AND search*) OR (medline AND search*) OR ‘systematic review’/exp

Bloc 3:

quality:ti OR assess*:ti OR reproduce*:ti,ab OR relia*b*:ti,ab OR intra-class-correlation:ti,ab OR internal-consistency:ti,ab OR valid*:ti,ab OR valid*:ti,ab OR method:ti,ab OR reproduc*:ti,ab OR observer:ti,ab OR factor-analysis:ti,ab OR factor-analyses:ti,ab OR factor:ti,ab OR structure:ti,ab OR discriminant:ti,ab OR ((clinimetric:ti,ab OR psychometric:ti,ab) AND (propert*:ti,ab OR analyst*:ti,ab)) OR (measurement:ti,ab AND propert*:ti,ab) OR (minimal*:ti,ab OR smallest:ti,ab) AND (important:ti,ab OR detectable:ti,ab OR real:ti,ab) AND (change:ti,ab OR difference:ti,ab))
OR ‘psychometry’/exp OR ‘clinimetry’/exp OR ‘observer variation’/exp OR ‘reliability’/exp OR ‘reproducibility’/exp OR ‘variance’/exp OR ‘correlation coefficient’/exp OR ‘validation process’/exp

Bloc 4:

meta-analysis:ti OR meta-analyses:ti OR ‘Delphi technique’:ti OR Cross-sectional:ti OR ‘diagnosis’/exp OR ‘case report’/de OR ‘meta-analysis’:ti OR ‘screening’/exp OR letter:it OR animal/exp OR ‘animal model’/exp OR ‘animal experiment’/exp

PsycINFO (through WebSPIRS)

 Bloc 1:

(instruments in ti,ab) or (scales in ti,ab) or (Questionnaires in ti,ab) or (measures in ti) or (methods in ti) or (outcome measurements in ti,ab) or ((tests in ti,ab) and (review in ti,ab)) or (explode “Attitude-Measures” in MJ,MN) or (explode “Questionnaires-” in MJ,MN) or (explode “Psychotherapeutic-Outcomes” in MJ,MN) or (explode “Treatment-Outcomes” in MJ,MN) or (explode “Psychological-Assessment” in MJ,MN) or (explode “Measurement-” in MJ,MN) or (explode “Pain-Measurement” in MJ,MN) or (explode “Interviewing-” in MJ,MN)

Bloc 2:

(literature and search*) or (Medline and search*) or (Psycinfo and search*) or (Psycinfo and search*) or (review in
Bloc 3:
(explode “Psychometrics-” in MJ,MN) or (explode “Statistically-Validity” in MJ,MN) or (explode “Test-Validity” in MJ,MN) or (explode “Statistical-Reliability” in MJ,MN) or (explode “Test-Reliability” in MJ,MN) or (explode “Test-Scores” in MJ,MN) or (explode “Test-Interpretation” in MJ,MN) or (explode “Test-Items” in MJ,MN) or (explode “Response-Variability” in MJ,MN) or (explode “Variability-Measurement” in MJ,MN) or (explode “Response-Variability” in MJ,MN) or (explode “Variability-Measurement” in MJ,MN) or (explode “Evaluation-” in MJ,MN) or (explode “Error-of-Measurement” in MJ,MN) or (explode “Consistency-Measurement” in MJ,MN) or (explode “Statistical-Correlation” in MJ,MN) or (explode “Statistical-Measurement” in MJ,MN) or (quality in ti) or (assess* in ti,ab) or (reproduc* in ti,ab) or (reliab* in ti,ab) or (intraclass correlation in ti,ab) or (internal consistency in ti,ab) or (valid* in ti,ab) or (responsive* in ti,ab) or (agreement in ti,ab) or (factor analysis in ti,ab) or (factor analyses in ti,ab) or (factor structure in ti,ab) or (discriminant analysis in ti,ab) or (coefficient in ti,ab) or (psychometric in ti,ab) and ((quality in ti,ab) or (propert* in ti,ab) or (analys* in ti,ab)) or (measurement in ti,ab) and (proport* in ti,ab)) or (measurement in ti,ab) and (propert* in ti,ab) or (minimal* in ti,ab) or (smallest in ti,ab) and (important in ti,ab) or (detectable in ti,ab) or (real in ti,ab) and (change in ti,ab) or (difference in ti,ab))

Bloc 4:
(explode “Meta-Analysis” in MJ,MN) or (meta analysis in ti) or (metaanalysis in ti) or (delphi technique in ti) or (cross sectional in ti) or (explode “Diagnosis-” in MJ,MN) or (explode “Case-Report” in MJ,MN) or (explode “Screening-Tests” in MJ,MN)
(#1 and #2 and #3) NOT #4

Appendix 2

Data extraction form COSMIN review

| Appendix 2 continued |
|-----------------------|
| 3. Health status concept—according to authors—that the reviewed measurement instruments are supposed to measure: **multiple answers possible** |
| □ Biological and physiological process |
| □ Symptoms |
| □ Physical functioning |
| □ Social psychological functioning |
| □ General health perception (including health-related quality of life) |
| Other: …………………………………|
| 4. Type of measurement instruments that are being reviewed: **multiple answers possible** |
| □ PRO (e.g. self-administered, interview, telephone administered) |
| □ Proxy |
| □ Non-PRO (e.g. performance based test, observation or rating by professional, clinical value (e.g. lab value)) |
| □ Other: …………………………………|
| 5. Target population(s) in with the reviewed measurement instrument were validated |
| ……………………………………………|
| 6. Number of measurement instruments included in the review: |
| …………………………………|
| 7. Is the search strategy used and described? Described—not descr |
| 8. Which databases are searched? …………………………………|
| 9. Is the selection of articles performed by at least two reviewers? Yes/no? |
| 10. Is the data extraction performed by at least two reviewers? Yes/no/?/n.a. |
| 11. Did the authors search for all validation studies per measurement instrument? Yes/No/Don’t know |
| 12. Are the in- and exclusion criteria for articles described? Yes/no |
| 13. Gave the authors a total assessment of the quality of each measurement instrument (inclusion of all measurement prop)? Yes/no |
| 14. Is some order if importance of the properties taken into account Yes/no |
| 15. Which properties are reported: …………………………………|
| 16. How are they reported? (references or values?): …………………………………|

Methodological quality of individual studies
17. Are one or more standards applied?
18. Which standards (per property) and are they fully described (i.e. reproducible)?
19. Is per measurement instrument described if it fulfils the standard?
Appendix 2 continued

Results of individual studies
20. Are the results evaluated, i.e. are one or more criteria applied?
21. Which criteria (per property) and are they fully described (i.e. reproducible)?
22. Is per measurement instrument described if it fulfills the criterion?

References

1. Marshall, M., Lockwood, A., Bradley, C., Adams, C., Joy, C., & Fenton, M. (2000). Unpublished rating scales: A major source of bias in randomised controlled trials of treatments for schizophrenia. The British Journal of Psychiatry, 176, 249–252. doi: 10.1192/bjp.176.3.249.
2. Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. Journal of Chronic Diseases, 40, 171–178. doi:10.1016/0149-2918(87)90069-5.
3. Higgins, J. P. T., & Green, S. (Eds.). (2006). Cochrane handbook for systematic reviews of interventions. Version 5.0.1 (update September 2008). The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.
4. Shea, B. J., Grimshaw, J. M., Wells, G. A., et al. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology, 7, 10. doi:10.1186/1471-2288-7-10.
5. Irwig, L. M., Tosteson, A. N. N., Gatsonis, C., et al. (1994). Guidelines for meta-analyses evaluating diagnostic tests. Annals of Internal Medicine, 120, 667–676.
6. Khan, K. S. (2005). Systematic reviews of diagnostic tests: A guide to methods and application. Best Practice & Research. Clinical Obstetrics & Gynaecology, 19, 37–46.
7. Edwards, P., Clarke, M., DiGuiseppi, C., Pratap, S., Roberts, I., & Wentz, R. (2002). Identification of randomized controlled trials in systematic reviews: Accuracy and reliability of screening records. Statistics in Medicine, 21, 1635–1640. doi:10.1002/sim.1190.
8. Van Tulder, M., Furlan, A., Bombardier, C., & Bouter, L. (2003). Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. Spine, 28, 1290–1299. doi:10.1097/00007632-200306150-00014.
9. Verhagen, A. P., De Vet, H. C., De Bie, R. A., et al. (1998). The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology, 51, 1235–1241. doi:10.1016/S0895-4356(98)00131-0.
10. 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. doi:10.1186/1471-2288-3-25.
11. Lohr, K. N., Aaronson, N. K., Alonso, J., et al. (1996). Evaluating quality-of-life and health status instruments: Development of scientific review criteria. Clinical Therapeutics, 18, 979–992. doi:10.1016/S0149-2918(96)80054-3.
12. Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. Quality of Life Research, 11, 193–205. doi:10.1023/A:1015291021312.
13. Terwee, C. B., Bot, S. D., De Boer, M. R., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60, 34–42. doi:10.1016/j.jclinepi.2006.03.012.
14. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. Journal of the American Medical Association, 273, 59–65. doi:10.1001/jama.273.1.59.
15. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research. U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. (2006). Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. Health and Quality of Life Outcomes, 4, 79. doi:10.1186/1477-7525-4-79.
16. Patrick, D. L., Guyatt, G. H., & Acquaro, C. (2008). Patient-reported outcomes. In J. Higgins & S. Green (Eds.), The Cochrane library (Chap. 17, issue 7 ed). Chichester: Wiley.
17. Pickard, A. S., Topfer, L. A., & Feeny, D. H. (2004). A structured review of studies on health-related quality of life and economic evaluation in pediatric acute lymphoblastic leukemia. Journal of the National Cancer Institute. Monographs, 33, 102–125. doi:10.1093/jncimonographs/lgh002.
18. Smith, M. (2005). Pain assessment in nonverbal older adults with advanced dementia. Perspectives in Psychiatric Care, 41, 99–113. doi:10.1111/j.1744-6163.2005.00021.x.
19. Terwee, C. B., Mokkink, L. B., Steultjens, M. P., & Dekker, J. (2006). Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: A systematic review of measurement properties. Rheumatology (Oxford, England), 45(7), 890–902. doi:10.1093/rheumatology/ke267.
20. Van der Windt, D. A., Nagelkerke, A. F., Bouter, L. M., Dankert-Roelse, J. E., & Veerman, A. J. (1994). Clinical scores for acute asthma in pre-school children. A review of the literature. Journal of Clinical Epidemiology, 47, 635–646. doi:10.1016/0895-4356(94)00211-9.
21. Arrington, R., Cofrancesco, J., & Wu, A. W. (2004). Questionnaires to measure sexual quality of life. Quality of Life Research, 13, 1643–1658. doi:10.1007/s11136-004-7625-z.
22. Stanghellini, V., Armstrong, D., Monnikes, H., & Bardhan, K. D. (2004). Systematic review: Do we need a new gastro-oesophageal reflux disease questionnaire? Alimentary Pharmacology & Therapeutics, 19, 463–479. doi:10.1046/j.1365-2036.2004.01861.x.
23. Platz, T., Eickhof, C., Nuyens, G., & Vuadens, P. (2005). Clinical scales for the assessment of spasticity, associated phenomena, and function: A systematic review of the literature. Disability and Rehabilitation, 27, 7–18. doi:10.1080/09638280400014634.
24. Gruenewald, D. A., Higginsson, I. J., Vivat, B., Edmonds, P., & Burman, R. E. (2004). Quality of life measures for the palliative care of people severely affected by multiple sclerosis: A systematic review. Multiple Sclerosis, 10, 690–704. doi:10.11191/1352458504ms1161r.
25. Ramaker, C., Marinus, J., Stiggelbout, A. M., & Van Hilten, B. J. (2002). Systematic evaluation of rating scales for impairment and disability in Parkinson’s disease. Movement Disorders, 17, 867–876. doi:10.1002/mds.10248.
26. Drake, B. G., Callahan, C. M., Dittus, R. S., & Wright, J. G. (1994). Global rating systems used in assessing knee arthroplasty outcomes. The Journal of Arthroplasty, 9, 409–417. doi:10.1001/jama.273.1.59.
27. De Boer, J. B., Van Dam, F. S., & Sprangers, M. A. (1995). Health-related quality-of-life evaluation in HIV-infected patients. A review of the literature. Pharmacoeconomics, 8, 291–304.
28. Grotle, M., Brox, J. I., & Vollestad, N. K. (2005). Functional status and disability questionnaires: What do they assess? A systematic review of back-specific outcome questionnaires. *Spine*, 30, 130–140.

29. Law, M., & Letts, L. (1989). A critical review of scales of activities of daily living. *The American Journal of Occupational Therapy*, 43, 522–528.

30. Veenhof, C., Bijlsma, J. W., Van den Ende, C. H., Van Dijk, G. M., Pisters, M. F., & Dekker, J. (2006). Psychometric evaluation of osteoarthritis questionnaires: A systematic review of the literature. *Arthritis and Rheumatism*, 53, 480–492. doi: 10.1002/art.22001.

31. Salek, S. S., Walker, M. D., & Bayer, A. J. (1998). A review of quality of life in Alzheimer’s disease. Part 2: Issues in assessing drug effects. *PharmacoEconomics*, 14, 613–627. doi:10.2165/00019053-199814050-00003.

32. Walker, M. D., Salek, S. S., & Bayer, A. J. (1998). A review of quality of life in Alzheimer’s disease. Part 1: Issues in assessing disease impact. *PharmacoEconomics*, 14, 499–530. doi:10.2165/00019053-199814050-00004.

33. Daker-White, G. (2002). Reliable and valid self-report outcome measures in sexual (dys)function: A systematic review. *Archives of Sexual Behavior*, 31, 197–209. doi:10.1023/A:1014743304566.

34. 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 & Physiological Optics*, 24, 257–273. doi:10.1111/j.1475-1313.2004.00187.x.

35. Neebakantan, D., Omosefolu, F., Clark, T. J., Gupta, J. K., & Khan, K. S. (2004). Quality of life instruments in studies of chronic pelvic pain: A systematic review. *Journal of Obstetrics & Gynaecology*, 24, 851–858. doi:10.1080/0144361040019138.x.

36. Dorman, S., Byrne, A., & Edwards, A. (2007). Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. *Palliative Medicine*, 21, 177–191. doi:10.1177/0269216307076398.

37. Avery, K. N., Bosch, J. L., Gotoh, M., et al. (2007). Questionnaires to assess urinary and anal incontinence: Review and recommendations. *The Journal of Urology*, 177, 39–49. doi:10.1016/j.juro.2006.08.075.

38. Dorey, G. (2002). Outcome measures for erectile dysfunction. 1: Literature review. *British Journal of Nursing (Mark Allen Publishing)*, 11, 54–64.

39. Haywood, K. L., Garratt, A. M., & Fitzpatrick, R. (2005). Quality of life in older people: A structured review of generic self-assessed health instruments. *Quality of Life Research*, 14, 1651–1668. doi:10.1007/s11136-005-1743-0.

40. Van Tuijl, J. H., Janssen-Potten, Y. J., & Seelen, H. A. (2002). Outcome measures for erectile dysfunction. 1: Dimensional health status measures. *BMC Musculoskeletal Disorders*, 3, 335–341. doi:10.1186/1471-4448-3-6.

41. Bot, S. D., Terwee, C. B., Van der Windt, D. A., Bouter, L. M., Dekker, J., & De Vet, H. C. (2004). Clinimetric evaluation of shoulder disability questionnaires: A systematic review of the literature. *Annales of the Rheumatic Diseases*, 63, 335–341. doi:10.1136/ard.2003.007724.

42. Echer, C., Vaes, P., Van Aerschot, L., Asman, S., & Duquet, W. (2007). The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: A systematic review. *BMC Musculoskeletal Disorders*, 8, 6. doi:10.1186/1471-2474-8-6.

43. Costa, L. O. P., Maher, C. G., & Latimer, J. (2007). Self-report outcome measures for low back pain—Searching for international cross-cultural adaptations. *Spine*, 32, 1028–1037. doi:10.1097/BRS.0b00002612427926f.

44. Leeflang, M. M., Scholten, R. J., Rutjes, A. W., Reitsma, J. B., & Bossuyt, P. M. (2006). Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *Journal of Clinical Epidemiology*, 59, 234–240. doi:10.1016/j.jclinepi.2005.07.014.

45. Lemeshow, A. R., Blum, R. E., Berlin, J. A., Stoto, M. A., & Colditz, G. A. (2005). Searching one or two databases was insufficient for meta-analysis of observational studies. *Journal of Clinical Epidemiology*, 58, 867–873. doi:10.1016/j.jclinepi.2005.03.004.

46. Golomb, B. A., Vickrey, B. G., & Hays, R. D. (2001). A review of health-related quality-of-life measures in stroke. *PharmacoEconomics*, 19, 155–185. doi:10.2165/00019053-200119020-00004.

47. Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

48. Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12, 38–47. doi:10.1111/j.1745-3992.1993.tb00543.x.

49. Dorans, N. J. (2007). Linking scores from multiple health outcome instruments. *Quality of Life Research*, 16(S1), 85–94. doi:10.1007/s11136-006-9155-3.

50. Mokkink, L. B., Terwee, C. B., Knol, D. L., et al. (2006). Protocol of the COSMIN study: COmponent-based Standards for the selection of health Measurement Instruments. *BMC Medical Research Methodology*, 6, 2. doi:10.1186/1471-2288-6-2.

51. Eiser, C., & Morse, R. (2001). Quality-of-life measures in chronic diseases of childhood. *Health Technology Assessment*, 5, 1–157.

52. Pal, D. K. (1996). Quality of life assessment in children: A review of conceptual and methodological issues in multi-dimensional health status measures. *Journal of Epidemiology and Community Health*, 50, 391–396. doi:10.1136/jech.50.4.391.

53. Schmidt, L. J., Garratt, A. M., & Fitzpatrick, R. (2002). Child/parent-assessed population health outcome measures: A structured review. *Child: Care, Health and Development*, 28, 227–237. doi:10.1046/j.1651-2227.2002.00266.x.

54. Davis, E., Waters, E., Mackinnon, A., et al. (2006). Paediatric quality of life instruments: A review of the impact of the conceptual framework on outcomes. *Developmental Medicine and Child Neurology*, 48, 311–318. doi:10.1017/S0001212506000173.

55. Hunter, J., Higginson, I., & Garralda, E. (1996). Systematic literature review: Outcome measures for child and adolescent mental health services. *Journal of Public Health Medicine*, 18, 197–206.

56. Brouwer, C. N., Maille, A. R., Rovers, M. M., Grobbée, D. E., Sanders, E. A., & Schilder, A. G. (2005). Health-related quality of life in children with otitis media. *International Journal of Pediatric Otorhinolaryngology*, 69, 1031–1041. doi:10.1016/j.ijporl.2005.03.013.

57. Haywood, K. L., Garratt, A. M., & Fitzpatrick, R. (2005). Older people specific health status and quality of life: A structured review of self-assessed instruments. *Journal of Evaluation in Clinical Practice*, 11, 315–327. doi:10.1111/j.1365-2753.2005.00538.x.

58. Haywood, K. L., Garratt, A. M., & Fitzpatrick, R. (2006). Quality of life in older people: A structured review of self-assessed health instruments. *Expert Review of Pharmacoconomics & Outcomes Research*, 6, 181–194. doi:10.1586/1471-7167.6.2.181.

59. Hollifield, M., Warner, T. D., Lian, N., et al. (2002). Measuring trauma and health status in refugees: A critical review. *Journal of the American Medical Association*, 288, 611–621. doi:10.1001/jama.288.5.611.
Clinical Gastroenterology and Hepatology, 2, 164–174. doi: 10.1016/S1542-3565(03)00323-9.
93. De Korte, J., Mombres, F. M., Sprangers, M. A., & Bos, J. D. (2002). The suitability of quality-of-life questionnaires for psoriasis research: A systematic literature review. Archives of Dermatology, 138, 1221–1227. doi:10.1001/archderm.138.9.1221.
94. Lewis, V. J., & Finlay, A. Y. (2005). A critical review of Quality-of-Life Scales for Psoriasis. Dermatologic Clinics, 23, 707–716. doi:10.1016/j.det.2005.05.016.
95. Hallin, P., Sullivan, M., & Kreuter, M. (2000). Spinal cord injury and quality of life measures: A review of instrument psychometric quality. Spinal Cord, 38, 509–523. doi:10.1038/sj.sc.3101054.
96. Matza, L. S., Zyczynski, T. M., & Bavendam, T. (2004). A review of quality-of-life questionnaires for urinary incontinence and overactive bladder: Which ones to use and why? Current Urology Reports, 5, 336–342. doi:10.1007/s11934-004-0079-6.
97. Buck, D., Jacoby, A., Massey, A., & Ford, G. (2000). Evaluation of measures used to assess quality of life after stroke. Stroke, 31, 2004–2010.
98. Prasad, M., Wahlqvist, P., Shikhar, R., & Shih, Y. C. (2004). A review of self-report instruments measuring health-related work productivity: A patient-reported outcomes perspective. Pharmacoeconomics, 22, 225–244. doi:10.2165/00019053-20042204-00002.
99. Lolland, J. H., Pizzi, L., & Frick, K. D. (2004). A review of health-related workplace productivity loss instruments. Pharmacoeconomics, 22, 165–184. doi:10.2165/00019053-20042200-00003.
100. Lundstrom, M., & Wendel, E. (2006). Assessment of vision-related quality of life measures in ophthalmic conditions. Expert Review of Pharmacoeconomics & Outcomes Research, 6, 691–724. doi:10.1586/14737167.6.6.691.
101. Trippop, S., Pratheepawanit, S., Prathepawanit, N., Asawaphureekorn, S., Anuthangkoon, W., & Inthayung, S. (2005). Health related quality of life instruments for glaucoma: A comprehensive review. Journal of the Medical Association of Thailand, 88(S9), S155–S162.
102. Franic, D. M., Bramlett, R. E., & Bothe, A. C. (2005). Psychometric evaluation of disease specific quality of life instruments in voice disorders. Journal of Voice, 19, 300–315. doi:10.1016/j.jvoice.2004.03.003.
103. Morley, A. D., & Sharp, H. R. (2006). A review of sinonasal outcome scoring systems—Which is best? Clinical Otolaryngology, 31, 103–109. doi:10.1111/j.1749-4486.2006.01155.x.
104. Watt, T., Groenvold, M., Rasmussen, A. K., et al. (2006). Quality of life in patients with benign thyroid disorders. A review. European Journal of Endocrinology, 154, 501–510. doi:10.1530/eje.1.02124.
105. Ketelaar, M., Vermeer, A., & Holders, P. J. (1998). Functional motor abilities of children with cerebral palsy: A systematic literature review of assessment measures. Clinical Rehabilitation, 12, 369–380. doi:10.1097/011921598673571117.
106. Boyce, W. F., Gowland, C., Rosenbaum, P. L., et al. (1991). Measuring quality of movement in cerebral palsy: A review of instruments. Physical Therapy, 71, 813–819.
107. Buffart, L. M., Roebroeck, M. E., Pesch-Batenburg, J. M., Janssen, W. G., & Stam, H. J. (2006). Assessment of arm/hand functioning in children with a congenital transverse or longitudinal reduction deficiency of the upper limb. Disability and Rehabilitation, 28, 85–95. doi:10.1080/09638280500158406.
108. Pakulis, P. J., Young, N. L., & Davis, A. M. (2005). Evaluating physical function in an adolescent bone tumor population. Pediatric Blood & Cancer, 45, 635–643. doi:10.1002/pbc.20383.
109. Moore, D. J., Palmer, B., Patterson, T. L., & Jeste, D. V. (2007). A review of performance-based measures of functional living skills. Journal of Psychiatric Research, 41, 97–118. doi:10.1016/j.jpsychires.2005.10.008.
110. MacKnight, C., & Rockwood, K. (1995). Assessing mobility in elderly people. A review of performance-based measures of balance, gait and mobility for bedside use. Reviews in Clinical Gerontology, 5, 464–486. doi:10.1097/0001925980004895.
111. Wind, H., Gouttebarge, V., Kuijer, P. P. F. M., & Frings-Dresen, M. H. W. (2005). Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: A systematic review. Journal of Occupational Rehabilitation, 15, 253–272. doi:10.1023/b:joer.00000926-005-1223-y.
112. Mannerkorpi, K., & Ekdahl, C. (1997). Assessment of functional limitation and disability in patients with fibromyalgia. Scandinavian Journal of Rheumatology, 26, 4–13.
113. Millard, R. W., Beattie, P. F., & Jones, R. H. (1997). A comprehensive review of questionnaires to evaluate chronic pain-related disability. Critical Reviews in Physical and Rehabilitation Medicine, 9, 35–52.
114. Dowrick, A. S., Gabbe, B. J., Williamson, O. D., & Cameron, P. A. (2005). Outcome instruments for the assessment of the upper extremity following trauma: A review. Injury, 36, 468–476. doi:10.1016/j.injury.2004.06.014.
115. Dziedzic, K. S., Thomas, E., & Hay, E. M. (2005). A systematic search and critical review of measures of disability for use in a population survey of hand osteoarthritis (OA). Osteoarthritis and Cartilage, 13, 1–12. doi:10.1016/j.joca.2004.09.010.
116. Swinkels, R. A., Dijkstra, P. U., & Bouter, L. M. (2005). Reliability, validity and responsiveness of instruments to assess disabilities in personal care in patients with rheumatic disorders. A systematic review. Clinical and Experimental Rheumatology, 23, 71–79.
117. Swinkels, R. A., Bouter, L. M., Oostendorp, R. A., & Van den Ende, C. H. (2005). Impairment measures in rheumatic disorders for rehabilitation medicine and allied health care: A systematic review. Rheumatology International, 25, 501–512. doi:10.1007/s00296-005-0603-0.
118. Swinkels, R. A., Bouter, L. M., Oostendorp, R. A., Swinkels-Meewisse, I. J., Dijkstra, P. U., & De Vet, H. C. (2006). Construct validity of instruments measuring impairments in body structures and function in rheumatic disorders: Which constructs are selected for validation? A systematic review. Clinical and Experimental Rheumatology, 24, 93–102.
119. Swinkels, R. A., Oostendorp, R. A., & Bouter, L. M. (2004). Which are the best instruments for measuring disabilities in gait and gait-related activities in patients with rheumatic disorders. Clinical and Experimental Rheumatology, 22, 25–33.
120. McKibbin, C. L., Brekke, J. S., Sires, D., Jeste, D. V., & Patterson, T. L. (2004). Direct assessment of functional abilities: Relevance to persons with schizophrenia. Schizophrenia Research, 72, 53–67. doi:10.1016/j.schres.2004.09.011.
121. Keskula, D. R., & Lott, J. (2001). Defining and measuring functional limitations and disability in the athletic shoulder. Journal of Sport Rehabilitation, 10, 221–231.
122. Michener, L. A., & Leggin, B. G. (2001). A review of self-report scales for the assessment of functional limitation and disability of the shoulder. Journal of Hand Therapy, 14, 68–76.
123. Salerno, D. F., Copley-Merriman, C., Taylor, T. N., Shingole, J., & Schulz, R. M. (2002). A review of functional status measures for workers with upper extremity disorders. Occupational and Environmental Medicine, 59, 664–670. doi:10.1136/oem.59.10.664.
124. Chong, D. K. (1995). Measurement of instrumental activities of daily living in stroke. Stroke, 26, 1119–1122.
125. Croarkin, E., Danoff, J., & Barnes, C. (2004). Evidence-based rating of upper-extremity motor function tests used for people following a stroke. Physical Therapy, 84, 62–74.
158. Cremeens, J., Eiser, C., & Blades, M. (2006). Characteristics of Health-related Self-report Measures for Children Aged Three to Eight Years: A Review of the Literature. *Quality of Life Research, 15*, 739–754. doi: 10.1007/s11136-005-4184-x.

159. Hayes, J. A., Black, N. A., Jenkinson, C., et al. (2000). Outcome measures for adult critical care: A systematic review. *Health Technology Assessment, 4*, 1–111.

160. Pietrobon, R., Coeytaux, R. R., Carey, T. S., Richardson, W. J., & DeVellis, R. F. (2002). Standard scales for measurement of functional outcome for cervical pain or dysfunction: A systematic review. *Spine, 27*, 515–522. doi: 10.1097/00007632-200203010-00012.

161. Linder, J. A., Singer, D. E., Ancker, M., & Atlas, S. J. (2003). Measures of health-related quality of life for adults with acute sinusitis. A systematic review. *Journal of General Internal Medicine, 18*, 390–401. doi: 10.1046/j.1525-1497.2003.02744.x.

162. Hearn, J., & Higginson, I. J. (1997). Outcome measures in palliative care for advanced cancer patients: A review. *Journal of Public Health Medicine, 19*, 193–199.

163. Razvi, S., McMillan, C. V., & Weaver, J. U. (2005). Instruments used in measuring symptoms, health status and quality of life in hypothyroidism: A systematic qualitative review. *Clinical Endocrinology, 63*, 617–624. doi: 10.1111/j.1365-2265.2005.02381.x.

164. Bijkerk, C. J., De Wit, N. J., Muris, J. W., Jones, R. H., Knottnerus, J. A., & Hoes, A. W. (2003). Outcome measures in irritable bowel syndrome: Comparison of psychometric and methodological characteristics. *The American Journal of Gastroenterology, 98*, 122–127. doi: 10.1111/j.1572-0241.2003.07158.x.

165. Haywood, K. L., Hargreaves, J., & Lamb, S. E. (2004). Multi-item outcome measures for lateral ligament injury of the ankle: A structured review. *Journal of Evaluation in Clinical Practice, 10*, 339–352. doi: 10.1111/j.1365-2753.2003.00435.x.

166. Poolsup, N., Li Wan, P. A., & Oyebode, F. (1999). Measuring mania and critical appraisal of rating scales. *Journal of Clinical Pharmacy and Therapeutics, 24*, 433–443. doi: 10.1046/j.1365-2710.1999.00250.x.

167. D’Olhaberriague, L., Litvan, I., Mitsias, P., & Mansbach, H. H. (1996). A reappraisal of reliability and validity studies in stroke. *Stroke, 27*, 2331–2336.

168. Margolis, M. K., Coyne, K., Kennedy-Martin, T., Baker, T., Schein, O., & Revicki, D. A. (2002). Vision-specific instruments for the assessment of health-related quality of life and visual functioning: A literature review. *PharmacoEconomics, 20*, 791–812. doi: 10.2165/00019053-200220120-00001.

169. Ashcroft, D. M., Wan Po, A. L., Williams, H. C., & Griffiths, C. E. (1999). Clinical measures of disease severity and outcome in psoriasis: A critical appraisal of their quality. *The British Journal of Dermatology, 141*, 185–191. doi: 10.1046/j.1365-2133.1999.02963.x.

170. Bialocerkowski, A. E., Grimmer, K. A., & Bain, G. I. (2000). A systematic review of the content and quality of wrist outcome instruments. *International Journal for Quality in Health Care, 12*, 149–157. doi: 10.1093/intqhc/12.2.149.