Impairment measures in rheumatic disorders for rehabilitation medicine and allied health care: a systematic review

Swinkels, R.A.H.M.; Bouter, L.M.; Oostendorp, R.A.B.; van den Ende, C.H.M.

published in
Rheumatology International
2005

DOI (link to publisher)
10.1007/s00296-005-0603-0

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

Rheumatol Int (2005) 25: 501–512
DOI 10.1007/s00296-005-0603-0

REVIEW ARTICLE

Raymond A. H. M. Swinkels · Lex M. Bouter
Rob A. B. Oostendorp · Cornelia H. M. van den Ende

Impairment measures in rheumatic disorders for rehabilitation medicine and allied health care: a systematic review

Received: 8 August 2004 / Accepted: 16 February 2005 / Published online: 24 June 2005

Abstract The objective of this study is to provide a critical overview of available instruments to assess impairments in patients with rheumatic disorders, and to recommend reliable and valid instruments for use in allied health care and rehabilitation medicine. A computer-aided literature search (1982–2004) in several databases was performed to identify studies focusing on the clinimetric properties of instruments designed to assess impairments in function in patients with rheumatic disorders. Data on intra-rater reliability, inter-rater reliability and construct validity were extracted in a standardized way. Explicit criteria were applied for reliability and validity. Results: The search identified a total of 49 instruments to assess impairments in functions in patients with rheumatic disorders; 19 met the criteria for reliability, 22 met the criteria for validity, and 11 out of the 49 appeared to meet the criteria for both reliability and validity. In summary, evidence of both reliability and validity was only found for 11 out of 49 instruments for the assessment of impairments in patients with rheumatic disorders. Only a limited number of the identified instruments for the assessment of impairments is both reliable and valid. Allied health care professionals should be cautious in the selection of measurement instruments to assess their patients.

Keywords Outcome assessment · Measurement instruments · Impairments · Rheumatic disorders · Assessment · Construct validity · Reliability

R. A. H. M. Swinkels (✉) · L. M. Bouter
Institute for Research in Extramural Medicine, VU University Medical Centre, Amsterdam, The Netherlands
E-mail: swinky@xs4all.nl

R. A. B. Oostendorp
Centre for Quality of Care Research, University Medical Centre Nijmegen, Nijmegen, The Netherlands

C. H. M. van den Ende
Netherlands Institute for Health Services Research, Utrecht, The Netherlands

Introduction

In rehabilitation medicine and allied health care for patients with rheumatic disorders, attention has shifted from disease severity to impairments, disabilities and problems in participation [1]. In particular, the development of the International Classification of Impairments, Disabilities and Handicaps (ICIDH) [2] —after its revision in the past decade now renamed the *International Classification of Functioning, Disabilities and Health Problems* (ICF) [3]—has encouraged this shift. The ICF distinguishes the disease itself from disease consequences and influencing contextual factors. In rehabilitation medicine and allied health care, the treatment goals cannot directly be derived from the severity of the disease. If the treatment goals are expressed in terms of impairment, disability or problems in participation (as in the ICF), these must be assessed and objectified [4, 5].

From this point of view it is essential to choose adequate measurement instruments, based on the evidence of their methodological quality [4, 6–11].

This study focuses on the psychometric properties of instruments to measure impairments in functions in the body and structures of patients with rheumatic disorders, a population which is frequently treated with physiotherapy and occupational therapy. "*Impairment*" is defined as any significant loss or abnormality of psychological, physiological or anatomical structure or function. Impairment is part of a health condition, but does not necessarily indicate that a disease is present or that the individual should be regarded as sick [12]. "*Disabilities*" are defined as limitations in the performance of a task or action by an individual [12]. We recently demonstrated that 57% of all available instruments to assess people with rheumatic disorders focus on impairments [13]. We started to focus on instruments to measure impairments because of this dominance. In the majority of clinical situations (60%), instruments are used to measure status praesens of the patient [13].

Inter-rater and intra-rater reliability are the first characteristic of an instrument that are required to be adequate; if a measurement is not reproducible it will not be useful. Furthermore, an important requirement of a measurement scale is the unidimensionality of subscales. If sufficient reliability has been demonstrated, it is also relevant to assess the validity.

In the literature, several aspects of validity are described, but construct validity is the most commonly assessed aspect. An instrument can be validated by studying its correlation with an optimally-comparable construct (e.g., the same impairment) or by correlation with an imperfect construct, such as gender, age, etc. In this study we distinguished various levels of construct validity, because we expected to find higher correlation values if measurement instruments were validated against an optimally-comparable construct than if validated against an imperfect construct [14]. If so, this could be relevant for interpretation of validity studies.

The aim of this study was threefold. The first objective was to make an inventory of available instruments and questionnaires for the assessment of impairments in patients with rheumatic disorders. The second aim was to investigate which of these instruments have acceptable methodological quality with regard to reliability and validity. The third aim was to investigate the assumption that construct validity results in higher correlation values when validated against a more suitable construct.

Method and materials

We performed a systematic review of the literature on the methodological characteristics of measurement instruments for all impairments that are relevant in allied health care for patients with rheumatic disorders. The term "measurement instrument" includes questionnaires (to be completed by the patient), observation lists (to be completed by the researcher) and anthropometric measurement instruments (technical measurement tools). Not all reported properties of instruments are appropriate for all kind of instruments. For example, inter-rater reliability (the degree to which two or more observers can obtain the same ratings) is inappropriate for questionnaires that are completed by the patient.

Inclusion procedure

The following criteria were applied:

- All studies had to focus on patients suffering from rheumatoid arthritis, seronegative polyarthritis (including psoriatic arthritis), osteoarthritis, ankylosing spondylitis, polymyositis or fibromyalgia.
- The studies had to contain information about the psychometric properties of instruments to assess impairments in mental functions, stiffness, pain, joint mobility (in terms of Range of Motion), muscle force and swelling (ICIDH-classification) [15].
- Many questionnaires focus on more than one domain of the ICIDH-classification, or measure more than one variable. Included were: (1) instruments which focus mainly (50% or more of the items) on the impairment to be measured; and (2) questionnaires with a sub-scale for the impairment in question that can be interpreted separately as a single entity.
- Different versions of an instrument were considered as separate measurement instruments.
- Only instruments for the measurement of adult patients were included.

Because of the method of data-reduction as described above, some frequently-used questionnaires, like the Health Assessment Questionnaire (HAQ), or the Western Ontario McMaster Osteoarthritis Index (WOMAC), the EuroQoL and Nottingham Health Profile are not included. The condition to be included was that 50% or more of the items should focus on the impairment to be measured. Questionnaires like HAQ, WOMAC and Nottingham Health Profile predominantly focus on disabilities.

Literature search

First the Medline database was searched for the period January 1982–May 2004, using search terms for the relevant rheumatic disorders and various search terms for psychometric properties[1]. The database of the Centre for Documentation of the Dutch Institute of Allied Health Care was also searched for the period January 1988–May 2004, using the same keywords. Furthermore, the search in both databases was repeated with the names of the identified measurement instruments. English, French, German and Dutch literature was included. The search was subsequently augmented with a manual search based on the references of the relevant publications, and therefore the search also yielded some publications from before 1982. In total, 156 measurement instruments were identified.

Data-extraction

All identified publications were assessed independently on the basis of their title, and abstract by two reviewers (RS and YK). In case of disagreement (3%) the article was also assessed by a third reviewer. The assessment was based on a standardized data-collection form (R.A.B. Oostendorp et al., unpublished data). This form consists of four sections: general description (name, first

[1] The detailed search strategy is available on request from the first author

Table 1 Applied cut-off points for intra-rater reliability, inter-rater reliability and construct validity

	Intra-rater reliability	Inter-rater reliability	Optimal convergent construct validity	Least convergent construct validity
Good	$x \geq 0.85$	$x \geq 0.80$	$x \geq 0.65$	$x \geq 0.50$
Moderate	$0.65 < x < 0.85$	$0.60 < x < 0.80$	$0.50 < x < 0.65$	$0.40 < x < 0.50$
Poor	$x < 0.65$	$x < 0.60$	$x < 0.50$	$x < 0.40$

x = Pearson's r, Spearman's ρ or intra-class correlation coefficient (ICC).

author, etc.), assessment domain (according to the ICIDH-classification), methodological aspects (concerning reliability, validity) and aspects of utilization.

Methodological criteria for psychometric properties

We investigated the following psychometric properties: intra-rater reliability, inter-rater reliability, construct validity and responsiveness. To interpret the data on reliability and validity we used criteria based on De Jong et al. [17], Eliasziw et al. [18] and Doeglas et al. [19] (Table 1).

The cut-off points for "good" reliability are supported by Weiner and Stewart, who also suggested 0.85 as a criterion [20, 21].

To investigate the influence of validating against different constructs (varying from optimally comparable to imperfectly comparable) the construct validity was divided into five clusters, in which the constructs against which a measurement instrument is validated are defined according to their anticipated degree of similarity to the instrument at issue (Table 2).

"Construct 1" is the most convergent construct, which means that the validity is measured against a variable which is very similar to the variable to be validated (e.g., a pain-intensity questionnaire [variable pain] that is validated against a Visual Analogue Scale for pain). Validity is defined as "construct 2" if the instrument is validated against instruments that measure the same construct as well as other impairments. "Construct 3" indicates that the construct relates to other impairments than the impairment to be validated. "Construct 4" relates to disabilities, and not to impairments. "Construct 5" is the least convergent validity, which means that the construct that is used to validate a variable relates to other domains than the variable that is to be validated (e.g., a pain-intensity questionnaire (variable pain) that is validated against disabilities in inter-personal relationships).

We clustered "construct 1" and "construct 2" as "optimally-convergent validity" and "construct 3", "construct 4" and "construct 5" as "imperfectly-convergent validity". This distinction is also reflected in the criteria for validity in the two last columns of Table 1. The argument for this distinction is the fact that optimally-convergent construct validity comes closest to the gold standard, and is therefore expected to result in higher values than imperfectly-convergent validity [14]. An explanation of all the abbreviations used in the "Result" section is given in the Appendix.

Data analysis

The data were analysed in the Statistical Package for Social Sciences (SPSS) 8.0. A classification of instruments was first made according to the type of impairments, based on the domain of assessment (impairments) [12].

The values of different studies of the same instrument were pooled for each psychometric property that was assessed (reliability and construct validity, related to sample size and correlation coefficients). By means of statistical pooling taking sample sizes into account, the results of separate studies on reliability and constructed validity, expressed in correlation coefficients, were combined into a single index. When the relevant information was available, mathematical pooling of the data was performed if the measurement instrument was validated against the same construct. The values were pooled per construct. The pooled index is composed of measurements, appropriately weighted:

$$X = \frac{\sum [n_i^* x_i]}{N}$$

where X = pooled index, n_i = number of persons included in the study, x_i = value of methodological aspect (Pearson's r, Spearman's ρ or intra-class correlation coefficient (ICC)) in the study, N = total number of persons in all studies included in the pooling [22]. The pooled index was computed for Pearson's/Spearman's correlation coefficients and the ICC values separately. Values for the construct validity of multidimensional instruments can be strongly influenced by values of one

Table 2 Categorization of comparators utilized in assessing construct validity

Construct level	Definition
Construct 1	Validation against sub-scales or instruments that measure the same impairment
Construct 2	Validation against instruments that measure the same impairment as well as other impairments
Construct 3	Validation against instruments that measure other impairments than the instrument to be validated
Construct 4	Validation against instruments that measure disabilities instead of the domain (impairments) to be validated
Construct 5	Validation against generic instruments that measure impairments as well as disabilities and participation problems

or more sub-scales. Therefore, whenever possible, the data were also pooled for the separate sub-scales.

Results

Table 3 presents the impairments in the ICIDH-classification that are considered to be relevant for people with rheumatic disorders in allied health care settings. The number of identified instruments is also shown for each group of relevant impairments.

In total the search identified 49 measurement instruments for the assessment of impairments in functions in patients with rheumatic disorders (Table 3). Several of these 49 instruments measure more than one impairment. Two adapted versions of the AIMS, a multidimensional questionnaire, are included: the AIMS2 and the AIMSS (for all abbreviations see the Appendix). As can be observed in Table 3, the categories of impairments that are most frequently assessed by measurement instruments are pain ($n = 20$; 41%) and impairments in musculoskeletal and movement related functions ($n = 15$; 33%).

With regard to stiffness, the Visual Analogue Scale-Stiffness and Morning Stiffness-Duration measure *only* stiffness. The BASDAI also measures other aspects of disease activity, such as (joint) pain, swelling and tiredness. Impairments in sensory functions are mainly assessed by instruments that focus on pain, with only one exception (the AIMS). Fifteen measurement instruments that were identified only measure impairments in musculoskeletal and movement-related functions. Six of these instruments have been specifically developed for assessing patients with ankylosing spondylitis.

Measurement of impairments in mental functions of patients with rheumatic disorders

Most instruments that measure impairments in mental functions of patients with rheumatic disorders (they all concern questionnaires) do not only measure these impairments, but also other impairments and/or disabilities. Four instruments that were identified focus only and specifically on impairments in mental functions (AHI, BDI, SSAI and STAI); the SF-36 and the AIMS and its modifications have only a sub-scale for measuring mental health, and focus mainly on disabilities.

The AIMS, in different modifications, is the only questionnaire that meets the criterion for reliability (Table 4). For the most convergent construct validity (validated against an optimally-comparable construct), two questionnaires meet the criterion: BDI (pooled $r = 0.67$), and STAI (pooled $r = 0.68$). For the imperfect construct validity, another two questionnaires can be qualified as "good": the AHI (pooled $r = 0.69$) and the AIMS-depri (pooled $r = 0.74$). None of the identified instruments had good reliability as well as good validity.

Table 3 Number of measurement instruments for each category of impairments

	Number of measurement instruments/sub-scales
Impairments in mental functions	11
Impairments in sensation (stiffness)	3
Impairments in sensory functions (pain)	20
Experience of pain	19
Pain behaviour	1
Impairments in neuro-musculoskeletal- and movement-related functions	15
In mobility of joints	11
In muscle force	3
Joint swelling	2

Measurement of impairments in sensation: stiffness

There are three available instruments to measure stiffness: BASDAI, MS-D and VAS-S (Table 4). In all three, the intra-rater reliability is good. There is a great variety in the validity of these instruments, with values that range from 0.26 to 0.70 if validated against instruments that measure the same impairment (Table 4). The highest value is found for the BASDAI ($r = 0.70$). Only the BASDAI and the VAS-S have good reliability as well as good validity (shown in italics in Table 4).

Measurement of impairments in sensory functions: pain

In total, 20 measurement instruments were identified for the assessment of pain in patients with rheumatic disorders (two modifications of the AIMS included). Eight out of these 20 only measure (modalities of) pain: AI, DoI, MPQ, OPB, P-NRS, RAPS, RPS and VAS-P. All the others are multidimensional instruments, or represent a separate sub-scale for pain in a multidimensional instrument. The results are presented in Table 4.

Reliability studies were performed for all the 20 measurement instruments, with the exception of the EI, the P-NRS and the RPS. Five of the instruments showed good reliability: AIMS (pooled $r = 0.86$), AIMS2-pain (ICC = 0.89), ASES (pooled $r = 0.88$), AUSCAN (ICC = 0.89), J-MAP (internal consistency = 0.90) and RAPS (internal consistency = 0.92).

Seven instruments showed good convergent construct validity, with the highest values for the AI, the DoI and the VAS-P. Seven instruments were found to have imperfect construct validity, all by correlation with composite measurement instruments, combining impairments, disabilities and problems in participation in one measure. If reliability and construct validity are both required to be "good", the AIMS and the AUSCAN-OHI are the only suitable instruments for pain assessment.

Table 4 Results for reliability and validity concerning measurement instruments per impairment. The best instruments/individual measurements (viz., showing evidence of both reliability and validity) are highlighted in bold

Measurement instrument/sub-scale[a]	Intra-rater reliability	Inter-rater reliability	Construct validity if validated against instruments measuring			
			Same impairment	Same impairment as well as other impairments	Disabilities	Impairments as well as disabilities and participation problems, or general aspects like, gender, age
Measurement of impairments in mental functions						
AHI [23, 24]	0.53		0.46[c]		0.69[c]	0.23 [23]
AIMS [25–46]	0.86[c]					0.33 [34]
AIMS-anxiety [25–46]			0.43[c]			0.37 [36]
AIMS-depri [25–46]			0.43[c]		0.28 [36]	0.74[c]
AIMS-Emof [25–46]			0.57[c]		0.19[c]	0.45[c]
AIMSS-anxiety [40]			0.41[c]			0.41[c]
AIMSS-depri [40]			0.47[c]		0.14[c]	0.46[c]
BDI [24, 47]			0.67[c]		0.27[c]	0.64[c]
SF36 [48–52]	0.66[b]		0.36[c]		0.26[c]	0.35[c]
SSAI [24]			0.57[c]		0.47 [24]	0.45[c]
STAI [24]			0.68[c]		0.57 [24]	0.51[c]
Measurement of impairments in sensation (stiffness)						
BASDAI [53–55]	***0.74*** [53–55]		***0.70*** [53–55]		0.57 [54]	***0.67*** [54]
MS-D [31, 43, 56, 57]			0.26 [58]		0.16[c]	0.19[c]
VAS-S [31, 43, 56–58]	***0.93***		0.51 [58]		***0.91***[c]	
Measurement of impairment in sensory functions (pain)						
ADLps [60]	0.31 [60]					
AI [61–72]	0.84[c] 0.83[b, c]		0.83[c]		0.24[c]	0.42 [72]
AIMS [25–46]	***0.86***[c]		0.58[c]			***0.75***[c]
AIMS-pain [25–46]	0.62		0.40[c]	0.54[c]	0.39[c]	0.60[c]
AIMS2-pain [32, 46, 73]	0.89[b]			0.49 [73]	0.38[c]	
AIMSS-pain [41]	0.57 [41]		0.41[c]		0.50[c]	0.61[c]
ASES [74]	0.88[c]					
AUSCAN-OHI [75]	***0.84***		***0.65***			
Dol [76–85]	0.81[c] 0.74[b, c]		0.79[c]			
EI [86]			0.67 [86]		0.36[c]	
FFI [87]	0.84[b, c]					
J-MAP [88]	0.90		0.62		0.41	
MPQ [37, 89–92]	0.61[c]		0.37 [90]			
OPB [93, 94]	0.7 [94]					
P-NRS [95, 96]					0.32[c]	
RAPS [97]	***0.92***		***0.52***, ***0.68***			
RPS [98]			0.53			−0.43
SAJ [70]	0.82[c]		0.55[c]			
Stest [94]	0.67[c]					
VAS-P [31, 43, 57, 76, 95, 96, 99, 100]	0.80[c] 0.73[b, c]		0.82[c]		0.26[c]	0.73 [100]
Measurement of impairments in joint mobility						
BASMI [101]			0.93[c]			
Chest [102]	***0.95*** [102]	0.53[b]			***0.60*** [102]	
EDI-abd [103]	0.63[b]					
EPM [104–106]	***0.87***[c]	***0.87***[c]			***0.54***	***0.54***[c]
Gonio [107]	***0.89*** [107]		***0.92*** [107]			
MKI [108]	0.90		−0.59			
MobSpine CCD [109, 110]	0.93[b]	0.72[b]			0.37 [110]	
MobSpine OWD [96, 109, 110]	0.93[b]	0.92[b]			0.49 [110]	
Shob [107, 109]	***0.95***[b]	***0.88***[b, c]			***0.66***[c]	
Spond [111, 112]			0.92 [111]			
Stest [112, 113]	0.97 [113]					
Measurement of impairments in muscle force						
Gripp [31, 114–116]	0.86[c]					
MSI [117]	0.84 [117]	0.78 [117]				
Sphy [118–123]		***0.93*** [121]	***0.87***[c]			
Measurement of joint swelling						
AI [61–72]	0.80[c] 0.83[b, c]	***0.82***[c]	***0.88***[c]			
SAJ [70]	0.77 [63]			−0.65 [70]		

[a]For explanation of abbreviations see Appendix
All values expressed in Pearson's *r* of Spearman's *r*
[b]Intra-class correlation coefficient value
[c]Pooled value
Values in italics are instruments that meet the criteria for reliability as well as for validity. The column "if validated against instruments that measure other impairments than the instrument to be validated" is lacking in this Table because those studies were not appropriate here (because no data are retrieved on this)

Measurement of impairments in joint mobility

For the assessment of joint mobility in patients with rheumatic disorders, 11 measurement instruments were identified (Table 4). On the basis of the available data, all instruments meet the criterion for intra-rater reliability, with the exception of the EDI. For the BASMI and the spondylometer there is no available information concerning reliability. The same was found for the inter-rater reliability, with the exception of Chest-expansion (ICC = 0.53), and MobSpine-CCD (ICC = 0.72). The BASMI (pooled r = 0.93), the goniometer (r = 0.92) and the spondylometer (r = 0.92) prove to be valid in patients with rheumatic disorders if validated against an optimally comparable construct. If the validity is based on correlation with an imperfect construct, the EPM and Shobert's test can also be qualified as valid. If both reliability and validity are required to be "good", four measurement instruments meet the criteria: the Chest-expansion, the EPM, the goniometer and Shobert's test.

Measurement of impairments in muscle force

Three instruments were identified to measure impairments in muscle force in patients with rheumatoid arthritis. The Grippit and the MSI are only intended to measure hand muscles, but the sphygmomanometer can also be used for other muscle groups (Table 4).

Only the sphygmomanometer meets the criteria for both reliability and validity (reliability r = 0.93; construct validity r = 0.87(pooled)).

Measurement of impairments in swelling

Two instruments are described for the assessment of (joint) swelling, both of which are sub-scales of an instrument to assess disease activity (Table 4).

The AI meets the criterion for inter-rater reliability (pooled r = 0.82). The validity of the AI (r = 0.88) and the SAJ (r = −0.65) was investigated. The AI was found to be the best instrument to assess joint swelling in patients with rheumatic disorders.

Consequences of distinguishing convergent construct validity (validated against an optimally-comparable construct) vs divergent construct validity (validated against an imperfect construct)

Summarizing the data on measurement instruments for relevant impairments, data on the validity of 40 instruments and/or sub-scales are available. For 21 out of these 40 measurement instruments, there are data on the optimally-comparable construct validity as well as the imperfect construct validity. In 14 of those 21 instruments, the correlation values for the optimally-comparable constructs proved to be higher than the values for validation against the imperfect constructs. In those studies reporting that the validity based on the imperfect constructs was better, in four out of six cases it concerned the AIMS, or modifications or sub-scales of the AIMS.

Discussion

We identified 49 instruments for the assessment of impairments in mental functions, stiffness, pain, joint swelling, mobility and muscle strength in patients with rheumatic disorders. Sixteen of these instruments were found to have good reliability, and 24 were found to have good validity. Only 9 out of the 49 measurement instruments had good reliability as well as good validity. None of the identified instruments for the assessment of impairments in mental functions were both reliable and valid. With regard to the other impairments, only one to three instruments per impairment met the criteria for reliability as well as validity. For assessment of stiffness, the best instruments are the BASDAI and the VAS-S. For assessment of pain the best instrument is the AIMS. For joint mobility, Chest, EPM and Gonio are the best available instruments, for muscle force the Sphy, and finally, for assessment of (joint) swelling the best instrument is the AI. The best available instruments are given in italics in Table 4.

Clinical consequences of this study

Many measurement instruments are developed for the assessment of impairments in people with rheumatic disorders. For clinical use, unidimensionality, good reliability and good validity are required for a useful application. The question is not what kind of new instruments should be developed, but rather which of the existing measurement instruments should be rejected because of insufficient quality or lack of information about their psychometric properties. For at least seven instruments there is a lack of information concerning their reliability. This might be a consequence of the fact that this review was restricted to studies focusing on populations with rheumatic disorders. The reliability and validity of the instruments might well have been investigated in other patient populations. The results of this review show that six instruments had insufficient reliability as well as insufficient validity, so there is evidence to justify rejection of those six instruments for populations with rheumatic disorders: MS-D, ADLps, MPQ, P-NRS, RPS and EDI-abd.

Many questionnaires that are intended to be used by people with chronic diseases are multidimensional in order to get as complete an overview of the impact of the disease as possible. However, this multidimensionality makes validation of the questionnaire more complex. It may be easier to find an optimally-comparable construct to validate against for unidimensional questionnaires than for a multidimensional questionnaire. For that reason we analysed data for the separate sub-scales if possible, in order to study the influence of the validity of subscales on the validity of the total questionnaire. The assumption is that each sub-scale is unidimensional, which is the first requirement of a measurement scale and also requires a good internal consistency.

It is remarkable that unsatisfactory validity was found for several sub-scales of the AIMS; namely, the AIMS-pain, AIMSS-pain, AIMS-anxiety, AIMS-depri and AIMS-emof. Those (sub-)scales were especially developed to assess a specific impairment, so one would expect them to have high correlation values if validated against an optimally-comparable construct. The reliability of sub-scales of the AIMS has not been investigated. Research on this topic is restricted to the AIMS as a whole. On the basis of our results it could be concluded that the AIMS as a whole is reliable and valid, but that there is insufficient evidence for "good" psychometric properties of separate sub-scales of the AIMS. Furthermore, restrictions to this conclusion must be made because of the fact that the conclusion regarding the validity of the AIMS as a whole is based on comparison with an imperfect construct. In fact, in general, construct validity is a way of hypothesis testing, where the hypothesis is hidden in the hypothetical (theoretical) constructs. The hypothetical constructs contain proposed underlying factors (which we tried to classify in Table 2). Possibly the proposed underlying factors (which are also incorporated in the different items of a questionnaire) are not fully correct, which might also explain the differences found in correlations. In fact, the results of construct validity must be interpreted in a broader context than only the strength of correlation with other measures. It is often advocated in the literature that validity is also related to reliability and internal consistency, as well as to the proposed underlying factors in the hypothetical constructs. The results of our study should be interpreted in the light of this.

Despite the classification of constructs into "convergent constructs" vs "imperfectly-convergent constructs" and our expectation to find stronger correlations in validation against "convergent constructs", our results did not confirm this in 100% of cases. In particular, some impairment-measures showed strong correlations when validated against disabilities. This could possibly be explained by the fact that some impairments are good predictors for certain disabilities. For example "depression" (AIMS-depri) will have a relative strong impact on disability and participation problems (correlation $r = 0.74$; Table 4).

The majority of instruments to measure joint mobility were found to have good intra- and/or inter-rater reliability. Some of the instruments are designed for specific groups of patients: for measuring mobility in patients with ankylosing spondylitis, the most reliable and valid available instruments are the Chest-expansion measurement and Shobert's test. For patients with rheumatoid arthritis the best option is the EPM-ROM. For all rheumatic disorders, in general, the best options are the goniometer, the spondylometer and the BASMI.

The sphygmomanometer is the most reliable and valid instrument to assess muscle strength. It is mainly used for the assessment of grip strength, but can also be applied to assess the strength of other muscles [119]. When assessing grip strength in particular, the sphygmomanometer can measure isometric muscle strength reliably and quickly.

In this study we did not report the responsiveness of the identified instruments. So far, research into the responsiveness of measurement instruments is hampered by the lack of consensus about the preferred method. Further research is needed to investigate the responsiveness of the identified instruments in the assessment of impairments.

In this study, several levels of construct validity were distinguished. We only accepted a value of ≥0.50 as "good" if validated against an imperfect construct, and a value of ≥0.65 as "good" if validated against an optimally-comparable construct. Our cut-off points are based on the literature, but in fact many authors seem to deviate from literature when interpreting the results of their own study [106, 124–126]. To our knowledge, different cut-off points for optimally-convergent validity vs imperfectly-convergent validity have never before been applied in validation studies. The results of our study demonstrate that the use of different cut-off points is justified; most measurement instruments showed better correlations with optimally-comparable constructs than with imperfect constructs, and this information is relevant for the interpretation of validity studies in general. However, the levels of the cut-off points applied in this study remain arbitrary. Regarding correct interpretation of the cut-off points for "good" reliability, it must be emphasized that these also depend on the sample size, since a sample size of 1,000 can tolerate a much less reliable instrument than a sample of 10 [21]. The aim of our study was to give an overview of reliability of instruments, intended for individual use in clinical situations. As a consequence of pooling data, the total number of patients generating our results is large; therefore, a high cut-off point is required to justify extrapolation to level of individuals.

The workgroup Outcome Measures in Rheumatoid Arthritis Clinical Trials

The workgroup Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) also emphasizes the importance of the methodological quality of measurement instruments. OMERACT established a core set of eight end-points that should be evaluated in clinical trials (disease activity, disability, pain, patient global assessment, physician global assessment, swollen joint, tender joint, and joint imaging) [127, 128]. The instruments discussed in this article focus on two of these domains (pain and the number of tender joints). The results of our study might also be useful in selecting the most appropriate measurement instrument(s) to assess joint swelling and pain.

In summary, evidence of both reliability and validity was only found for 11 out of 49 instruments for the assessment of impairments in patients with rheumatic disorders. Those 11 best instruments are reflected in bold in Table 4. In six instruments neither the reliability nor the validity was found to be adequate. Evidence concerning the reliability and validity of sub-scales of the AIMS is lacking.

Acknowledgements We wish to thank Mrs. Y. Kappe (YK) for reviewing the majority of the articles in conjunction with the first author. This study is supported by a grant from the Health Insurance Executive Board (CvZ), no. CURE/24101/97.

Appendix

Table a

Table a List of abbreviations of measurement instruments and sub-scales

Abbreviation	Name of measurement instrument	Abbreviations of sub-scales
AHI	Arthritis Helplessness Index	Anx, anxiety Depri, depression
AI	Articular Index	Pain, pain
AIM2D	Arthritis Impact Measurement Scale2 Dutch	Anx, anxiety Depr, depression EmoF, emotional function MenH, mental health Mob, mobility Pain, pain
AIMS	Arthritis Impact Measurement Scale	Anx, anxiety Depri, depression EmoF, emotional function MenH, mental health Mob, mobility Pain, pain
AIMS2	Arthritis Impact Measurement Scale 2	Anx, anxiety Depri, depression EmoF, emotional function MenH, mental health Mob, mobility Pain, pain
AIMSD	Arthritis Impact Measurement Scale–Dutch	Anx, anxiety Depri, depression EmoF, emotional function MenH, mental health Mob, mobility Pain, pain
AIMSS	AIMS short version	Anx, anxiety Depri, depression EmoF, emotional function MenH, mental health Mob, mobility Pain, pain
ASES	Arthritis Self Efficacy Scale	
AUSCAN-OHI	Australian/Canadian Osteoarthritis Hand Index	Pain, pain Stiff, stiffness PhysF, physical function
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index	
BASMI	Bath Ankylosing Spondylitis Metrology Index	Stiff-D, stiffness duration Stiff-S, stiffness severity
BDI	Beck Depression Inventory	
Chest	Chest Expansion	
Dol	Dolorimeter	
EDI	Electric Digital Inclinometer-320	
EI	Enthesis Index	
EPM	Escola Paulista de Medicina Range of Motion Scale	
FFD	Finger Floor Distance	
FFI	Functional Foot Index	
Gonio	Goniometer	
Gripp	Grippit	
J-MAP	Joint-Specific Multidimensional Assessment of Pain	
MKI	Modified Kapandji Index	

Table a (Contd.)

Abbreviation	Name of measurement instrument	Abbreviations of sub-scales
MobSpine	Mobility assessment spine in ankylosing spondylitis	CCD, chin to chest distance ChExp, chest expansion OWD, occiput to wall distance VitCp, vital capacity
MPQ	McGill Pain Questionnaire	
MS-D	Morning Stiffness, duration	Stiff-D, stiffness durationStiff-S, stiffness severity
MS–S	Morning Stiffness, severity	
P-NRS	Pain Numeric Rating Scale	
RAPS	Rheumatoid Arthritis Pain Scale	
RPS	Regional Pain Scale	
SAJ	Self Assessment Joint Count	
Shob	Shobert's test	
Sphy	Sphygmomanometer (handheld dynamometer)	
Spond	Spondylometer (Dunham)	
SSAI	Spielberger State-Anxiety Inventory	
STAI	Spielberger Trait-Anxiety Inventory	
Stest	Stiffness test	
VAS-P	Visual Analogue Scale Pain	

References

1. Rimmer JH (1999) Health promotion for people with disabilities: the emerging paradigm shift from disability prevention to prevention of secondary conditions. Phys Ther 79:495–502
2. World Health Organization (1980) International classification of impairments, disabilities and handicaps: a manual of classification relating to the consequences of disease. World Health Organization, Geneva
3. World Health Organization (2001) International classification of functioning, disabilities and health problems. World Health Organization, Geneva
4. Feinstein AR (1987) Clinimetrics. Yale University Press, New Haven
5. Hendriks HJM, Brandsma JW, Heerkens YF, Oostendorp RAB, Nelson RM (1997) Intraobserver and interobserver reliability of assessments of impairments and disabilities. Phys Ther 77:1097–1106
6. Kirschner B, Guyatt GH (1985) A methodological framework for assessing health indices. J Chronic Dis 38:27–36
7. Bombardier C, Tugwell P (1987) Methodological considerations in functional assessment. J Rheumatol 14(Suppl 15):6–10
8. Tugwell P, Bombardier C (1982) A methodologic framework for developing and selecting endpoints in clinical trials. J Rheumatol 9:758–762
9. Beckerman H, Bouter LM (1993) Efficacy of physiotherapy for musculoskeletal disorders: what can we learn from research?. Br J Gen Pract 43:483–491
10. McDowell I, Newell C (1987) Measuring health: a guide to rating scales and questionnaires. Oxford University Press, New York
11. Streiner DL, Norman GR (eds) (1999) Health measurement scales. Oxford University Press, Oxford, pp 1–68
12. World Health Organization (1999) International classification of functioning and disability. Beta-2 draft, short version. World Health Organization, Geneva
13. Swinkels RAHM (2004) The ICF-classification as a system for structuring outcome assessment. Physiotherapy Singapore 7(3):7–13
14. Swinkels RAHM, Vet HCWde, Oostendorp RAB, Bouter LM (2004) Construct validity of impairment measures in rheumatic disorders: which constructs are selected for validation? (submitted)
15. World Health Organization (1995) Adaptation of the classification of impairments for five Dutch allied health professions. World Health Organization, Geneva
17. DeJong A, Molenaar IW (1987) An application of Mokken's model for stochastic cummulative scaling in psychiatric research. J Psychiatr Res 21:137–149
18. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K (1994) Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Phys Ther 74:777–788
19. Doeglas D, Krol B et al (1995) The assessment of functional status in rheumatoid arthritis: a cross-cultural longitudinal comparison of the Health Assessment Questionnaire and the Groningen Activity Restriction Scale. J Rheumatol 22(10):1834–1843
20. Weiner E, Stewart B (1984) Assessing individuals. Little Brown, Boston
21. Streiner DL, Norman GR (2003) Health status measurement scales. A practical guide to their development and use. Oxford University Press, Oxford
22. Neter J, Wasserman W, Whitmore GA (1982) Applied statistics. Allyn and Bacon Inc, Boston
23. Nicassio PM, Wallston KA, Callahan LF, Herbert M, Pincus T (1985) The measurement of helplessness in rheumatoid arthritis. The development of the arthritis helplessness index. J Rheumatol 12:462–467
24. Hagglund KJ, Roth DL, Haley WE, Alarcon GS (1989) Discriminant and convergent validity of self-report measures of affective distress in patients with rheumatoid arthritis. J Rheumatol 16:1428–1432
25. Meenan RF, Anderson JJ, Kazis LE (1984) Outcome assessment in clinical trials. Evidence for the sensitivity of a health status measure. Arthritis Rheum 27:1344–1352
26. Meenan RF, Gertman PM, Mason JH (1980) Measuring health status in arthritis. The arthritis impact measurement scales. Arthritis Rheum 23:146–152
27. Kazis LE, Anderson JJ, Meenan RF (1989) Effect sizes for interpreting changes in health status. Med Care 27:S178–S189
28. Anderson JJ, Firschein HE, Meenan RF (1989) Sensitivity of a health status measure to short-term clinical changes in arthritis. Arthritis Rheum 32:844–850
29. Hidding A, Santen Mv, Klerk Ed, Gielen X, Boers M (1993) Discordance between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. Group Phys Ther in

ankylosing spondylitis (thesis). Meppel, Krips Repro, pp 94–104
30. Buchbinder R, Bombardier C, Yeung M, Tugwell P (1995) Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 38:1568–1580
31. Anderson JJ, Chernoff MC (1993) Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. J Rheumatol 20:535–537
32. Potts MK, Brandt KD (1987) Evidence of the validity of the Arthritis Impact Measurement Scales. Arthritis Rheum 30:93–96
33. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE (1992) AIMS2 The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 35:1–10
34. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A (1993) A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. J Rheumatol 20:429–436
35. Liang MH, Larson MG, Cullen KE, Schwartz JA (1985) Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 28:542–547
36. Meenan RF, Gertman PM, Mason JH, Dunaif R (1982) The arthritis impact measurement scales. Further investigations of a health status measure. Arthritis Rheum 25:1048–1053
37. Keefe FJ, Caldwell DS, Queen KT (1987) Pain coping strategies in osteoarthritis patients. J Consult Clin Psychol 55:208–212
38. Liang MH, Fossel AH, Larson MG (1990) Comparisons of five health status instruments for orthopedic evaluation. Med Care 28:632–642
39. Goeppinger J, Doyle MA, Charlton SL, Lorig K (1988) A nursing perspective on the assessment of function in persons with arthritis. Res Nurs Health 11:321–331
40. Jacobs JW, Oosterveld FG, Deuxbouts N (1992) Opinions of patients with rheumatoid arthritis about their own functional capacity: how valid is it? Ann Rheum Dis 51:765–768
41. Wallston KA, Brown GK, Stein MJ, Dobbins CJ (1989) Comparing the short and long versions of the arthritis impact measurement scales. J Rheumatol 16:1105–1109
42. Duffy CM, Watanabe Duffy KN, Gladman DD (1992) The utility of the arthritis impact measurement scales for patients with psoriatic arthritis. J Rheumatol 19:1727–1732
43. Hawley DJ, Wolfe F (1992) Sensitivity to change of the health assessment questionnaire (HAQ) and other clinical and health status measures in rheumatoid arthritis: results of short-term clinical trials and observational studies versus long-term observational studies. Arthritis Care Res 5:130–136
44. Vliet Vlieland TP, van der Wijk TP, Jolie IM, Zwinderman AH, Hazes JM (1996) Determinants of hand function in patients with rheumatoid arthritis. J Rheumatol 23:835–840
45. Weinberger M, Samsa GP, Tierney WM, Belyea MJ, Hiner SL (1992) Generic versus disease specific health status measures: comparing the sickness impact profile and the arthritis impact measurement scales. J Rheumatol 19:543–546
46. Husted J, Gladman DD, Farewell VT, Long JA (1996) Validation of the revised and expanded version of the arthritis impact measurement scales for patients with psoriatic arthritis. J Rheumatol 23:1015–1019
47. Callahan LF, Kaplan MR, Pincus T (1991) The Beck Depression Inventory, Center for Epidemiological Studies Depression Scale (CES-D), and General Well-Being Schedule depression subscale in rheumatoid arthritis. Criterion contamination of responses. Arthritis Care Res 4:3–11
48. Martin DP, Engelberg R, Agel J, Swiontkowski MF (1997) Comparison of the Musculoskeletal Function Assessment Quenstionnaire with the Short Form-36, the Western Ontario and McMaster University Osteoarthritis Index, and the Sickness Impact Profile Health-status measures. J Bone Joint Surg Am 79(9):1323–1335
49. Wright JG, Young NL (1997) A comparison of different indices of responsiveness. J Clin Epidemiol 50:239–246
50. Husted JA, Gladman DD, Farewell VT, Long JA, Cook RJ (1997) Validating the SF-36 Health Survey Questionnaire in patients with psoriatic arthritis. J Rheumatol 24:511–17
51. Kosinski M, Keller SD, Hatoum HT, Kong SX, Ware JE Jr (1999) The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: tests of data quality, scaling assumptions and score reliability. Med Care 37:MS10–MS22
52. Soderman P, Malchau H (2000) Validity and reliability of Swedish WOMAC osteoarthritis index: a self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). Acta Orthop Scand 1:39–46
53. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A (1994) A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 21:2286–2291
54. Van Tubergen A, Coenen J, Landewe R (2002) Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. Arthritis Rheum 47:8–16
55. Van Tubergen A, Debats I, Ryser L (2002) Use of a numerical rating scale as an answer modality in ankylosing spondylitis-specific questionnaires. Arthritis Rheum 47:242–248
56. Rhind VM, Bird HA, Wright V (1980) A comparison of clinical assessments of disease activity in rheumatoid arthritis. Ann Rheum Dis 39:135–137
57. Mallya RK, Mace BE (1981) The assessment of disease activity in rheumatoid arthritis using a multivariate analysis. Rheumatol Rehabil 20:14–17
58. Vliet-Vlieland TPM, Zwinderman AH, Breedveld FC, Hazes JMW (1997) Measurement of morning stiffness in rheumatoid arthritis clinical trials. J Clin Epidemiol 50:6
59. Oosterhof J (1992) Research to the Reproducibility of the VAS for stiffness in patients with rheumatoid arthritis. Unpublished
60. Callahan LF, Brooks RH, Summey JA, Pincus T (1987) Quantitative pain assessment for routine care of rheumatoid arthritis patients, using a pain scale based on activities of daily living and a visual analog pain scale. Arthritis Rheum 30:630–636
61. Ritchie DM, Boyle JA, McInnes JM (1968) Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. Q J Med 37:393–406
62. Camp AV (1971) An articular index for the assessment of rheumatoid arthritis. Orthopaedics 4:39–45
63. Dougados M, Gueguen A, Nakache JP, Nguyen M, Mery C, Amor B (1988) Evaluation of a functional index and an articular index in ankylosing spondylitis. J Rheumatol15:302–307
64. Prevoo ML, Riel PLCMv, Hof MAv' (1995) Validity and reliability of joint counts. A longitudinal study in patients with recent onset rheumatoid arthritis. In: Prevoo ML (ed) Clinical assessment in rheumatoid arthritis: aspects of validity and reproducibility. Catholic University Nijmegen, Nijmegen, pp 19–31
65. Hart LE, Tugwell P, Buchanan WW, Norman GR, Grace EM, Southwell D (1985) Grading of tenderness as a source of interrater error in the Ritchie articular index. J Rheumatol 12:716–717
66. Doyle DV, Dieppe PA, Scott J, Huskisson EC (1981) An articular index for the assessment of osteoarthritis. Ann Rheum Dis 40:75–78
67. Gaston-Johansson F, Gustafsson M (1990) Rheumatoid arthritis: determination of pain characteristics and comparison of RAI and VAS in its measurement. Pain 41:35–40
68. Kalla AA, Kotze TJ, Meyers OL, Parkyn ND (1988) Clinical assessment of disease activity in rheumatoid arthritis: evaluation of a functional test. Ann Rheum Dis 47:773–779

69. Fuchs HA, Brooks RH, Callahan LF, Pincus T (1989) A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. Arthritis Rheum 32:531–537
70. Prevoo ML, Kuper IH, van't Hof MA, van Leeuwen MA, van de Putte LB, van Riel PL (1996) Validity and reproducibility of self-administered joint counts. A prospective longitudinal follow-up study in patients with rheumatoid arthritis. J Rheumatol 23:841–845
71. Smolen JS, Breedveld FC, Eberl G (1995) Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. Arthritis Rheum 38:38–43
72. Orces CH, Del RI, Abel MP, Escalante A (2002) The number of deformed joints as a surrogate measure of damage in rheumatoid arthritis. Arthritis Rheum 47:67–72
73. Riemsma RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O (1996) Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. Br J Rheumatol 35:755–760
74. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR (1989) Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 32:37–44
75. Bellamy N, Campbell J, Haraoui B (2002) Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. Osteoarthritis Cartilage 10:863–869
76. Wigers SH, Skrondal A, Finset A, Götestam KG (1997) Measuring change in fibromyalgic pain: the relevance of pain distribution. J Musculoskeletal Pain 5:29–41
77. McCarthy DJ Jr, Gatter RA, Phelps P (1965) A dolorimeter for quantification of articular tenderness. Arthritis Rheum 44:551–559
78. Wessel J (1995) The reliability and validity of pain threshold measurements in osteoarthritis of the knee. Scand J Rheumatol 24:238–242
79. Puttick M, Schulzer M, Klinkhoff A, Koehler B, Rangno K, Chalmers A (1995) Reliability and reproducibility of fibromyalgic tenderness, measurement by electronic and mechanical dolorimeters. J Musculoskeletal Pain 3:3–14
80. Atkins CJ, Zielinski A, Klinkhoff AV (1992) An electronic method for measuring joint tenderness in rheumatoid arthritis. Arthritis Rheum 35:407–410
81. Guthrie NG, Edworthy SM, Martin L, Fritzler MJ, Kinsella TD, Ryan JP (1994) Comparing three techniques for the evaluation of fibromyalgia. Physiotherapy Can 46:94–97
82. Daltroy LH, Larson MG, Roberts NW, Liang MH (1990) A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol 17:946–950
83. Smythe HA, Buskila D, Urowitz S, Langevitz P (1992) Control and "fibrositic" tenderness: comparison of two dolorimeters. J Rheumatol 19:768–771
84. Tunks E, Crook J, Norman G, Kalaher S (1988) Tender points in fibromyalgia. Pain 34:11–19
85. Langley GB, Fowles M, Sheppeard H, Wigley RD (1983) A simple pressure dolorimeter for the quantification of joint tenderness in inflammatory arthritis. Rheumatol Int 3:109–112
86. Mander M, Simpson JM, McLellan A, Walker D, Goodacre JA, Dick WC (1987) Studies with an enthesis index as a method of clinical assessment in ankylosing spondylitis. Ann Rheum Dis 46:197–202
87. Saag KG, Saltzman CL, Brown CK, Budiman-Mak E (1996) The Foot Function Index for measuring rheumatoid arthritis pain: evaluating side-to-side reliability. Foot Ankle Int 17:506–510
88. O'Malley KJ, Suarez-Almazor M, Aniol J (2003) Joint-specific multidimensional assessment of pain (J-MAP): factor structure, reliability, validity, and responsiveness in patients with knee osteoarthritis. J Rheumatol 30:534–543
89. Lankveld Wv, Pad-Bosch Pv', Putte LBAvd, Staak Cvd, Näring G (1992) Pijn bij reumatoïde artritis gemeten met de visuele analoge schaal en de Nederlandse versie van de "McGill Pain Questionnaire". Ned Tijdschr Geneeskd 136:1166–1170
90. Papageorgiou AC, Badley EM (1989) The quality of pain in arthritis: the words patients use to describe overall pain and pain in individual joints at rest and on movement. J Rheumatol 16:106–112
91. Burckhardt CS, Bjelle A (1994) A Swedish version of the short-form McGill Pain Questionnaire. Scand J Rheumatol 23:77–81
92. Melzack R (1975) The McGill Pain Questionnaire: major properties and scoring methods. Pain 1:277–299
93. Anderson KO, Bradley LA, McDaniel LK (1987) The assessment of pain in rheumatoid arthritis: disease differentiation and temporal stability of a behavioral observation method. J Rheumatol 14:700–704
94. Molendijk M, Colard I, Jacobs JWG, Oosterveld FGJ (1991) Drukpunten bij het fibromyalgie syndroom; de betrouwbaarheid in de dagelijkse praktijk. Ned Tijdschr Fysioth 101:31–36
95. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA (1978) Studies with pain rating scales. Ann Rheum Dis 37:378–381
96. Langley GB, Sheppaerd H (1984) Problems associated with pain measurement in arthritis: comparison of the visual analogue and verbal rating scales. Clin Exp Rheumatol 2:231–234
97. Anderson DL (2001) Development of an instrument to measure pain in rheumatoid arthritis: Rheumatoid Arthritis Pain Scale (RAPS). Arthritis Rheum 45:317–323
98. Wolfe F (2003) Pain extent and diagnosis: development and validation of the regional pain scale in 12,799 patients with rheumatic disease. J Rheumatol 30:369–378
99. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A (1997) Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 36:551–559
100. Tamiya N, Araki S, Ohi G (2002) Assessment of pain, depression, and anxiety by visual analogue scale in Japanese women with rheumatoid arthritis. Scand J Caring Sci 16:137–141
101. Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, Calin A (1994) Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. J Rheumatol 21:1694–1698
102. Moll JM, Wright V (1973) The pattern of chest and spinal mobility in ankylosing spondylitis. An objective clinical study of 106 patients. Rheumatol Rehabil 12:115–134
103. Heemskerk MAMB, Aarst Mv, Windt DAWM (1997) De reproduceerbaarheid van het meten van de passieve beweeglijkheid van de schouder met de EDI-320 digitale hoekmeter. Ned Tijdschr Fysioth 107:146–149
104. Ende CHMvd, Vliet-Vlieland TPM, Breedveld FC, Hazes JM (1994) De EPM-ROM schaal: een waardevol meetinstrument voor de beweeglijkheid van de gewrichten van patiënten met reumatoïde artritis. Ned Tijdschr Fysioth 104:86–91
105. Vlieland TP, van Den Ende CH, Breedveld FC, Hazes JM (1993) Evaluation of joint mobility in rheumatoid arthritis trials: the value of the EPM-range of motion scale. J Rheumatol 20:2010–2014
106. Ferraz MB, Oliveira LM, Araujo PM, Atra E, Walter SD (1990) EPM-ROM Scale: an evaluative instrument to be used in rheumatoid arthritis trials. Clin Exp Rheumatol 8:491–494
107. Reynolds PM (1975) Measurement of spinal mobility: a comparison of three methods. Rheumatol Rehabil 14:180–185
108. Lefevre-Colau MM, Poiraudeau S, Oberlin C (2003) Reliability, validity, and responsiveness of the modified Kapandji index for assessment of functional mobility of the rheumatoid hand. Arch Phys Med Rehabil 84:1032–1038
109. Viitanen JV, Kautiainen H, Suni J, Kokko ML, Lehtinen K (1995) The relative value of spinal and thoracic mobility measurements in ankylosing spondylitis. Scand J Rheumatol 24:94–97
110. Viitanen JV (1993) Thoracolumbar rotation in ankylosing spondylitis. A new noninvasive measurement method. Spine 18:880–883

111. Sturrock RD, Wojtulewski JA, Hart FD (1973) Spondylometry in a normal population and in ankylosing spondylitis. Rheumatol Rehabil 12:135–142
112. Roberts WN, Liang MH, Pallozzi LM, Daltroy LH (1988) Effects of warming up on reliability of anthropometric techniques in ankylosing spondylitis. Arthritis Rheum 31:549–552
113. Newcomer KL, Krug HE, Mahowald ML (1993) Validity and reliability of the timed-stands test for patients with rheumatoid arthritis and other chronic diseases. J Rheumatol 20:21–27
114. Nordenskiold UM, Grimby G (1993) Grip force in patients with rheumatoid arthritis and fibromyalgia and in healthy subjects. A study with the Grippit instrument. Scand J Rheumatol 22:14–19
115. Dellhag B, Bjelle A (1995) A Grip Ability Test for use in rheumatology practice. J Rheumatol 22:1559–1565
116. Dellhag B, Hosseini N, Bremell T, Ingvarsson PE (2001) Disturbed grip function in women with rheumatoid arthritis. J Rheumatol 28:2624–2633
117. Stucki G, Schonbachler J, Bruhlmann P, Mariacher S, Stoll T, Michel BA (1994) Does a muscle strength index provide complementary information to traditional disease activity variables in patients with rheumatoid arthritis?. J Rheumatol 21:2200–2205
118. Hamilton GF, McDonald C, Chenier TC (1992) Measurement of grip strength: validity and reliability of the sphygmomanometer and Jamar Grip Dynamometer. J Orthop Sports Phys Ther 16:215–219
119. Helewa A, Goldsmith CH, Smythe HA (1981) The modified sphygmomanometer—an instrument to measure muscle strength: a validation study. J Chronic Dis 34:353–361
120. Helewa A, Goldsmith CH, Smythe HA (1986) Patient, observer and instrument variation in the measurement of strength of shoulder abductor muscles in patients with rheumatoid arthritis using a modified sphygmomanometer. J Rheumatol 13:1044–1049
121. Spiegel JS, Paulus HE, Ward NB, Spiegel TM, Leake B, Kane RL (1987) What are we measuring? An examination of walk time and grip strength. J Rheumatol 14:80–86
122. Lee P, Baxter A, Dick WC, Webb J (1974) An assessment of grip strength measurement in rheumatoid arthritis. Scand J Rheumatol 3:17–23
123. Pincus T, Brooks RH, Callahan LF (1991) Reliability of grip strength, walking time and button test performed according to a standard protocol. J Rheumatol 18:997–1000
124. Ghiselle EE (1966) The measurement of occupational aptitude. Berkely, Los Angeles
125. Huiskes CJAE, Kraaimaat FW, Bijlsma JWJ (1990) Development of a self-report questionnaire to assess the impact of rheumatic diseases on health and lifestyle. J Rehabil Sci 3:65–70
126. Stewart AL, Ware JE (1992) Measuring functioning and well-being: the Medical Outcomes Study Approach. Duke University Press, Durham, North Carolina
127. Tugwell P, Boers M (1993) OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: conclusion. J Rheumatol 20:590
128. Wolfe F, Lassere M, van der HD (1999) Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. J Rheumatol 26:484–489