Review

Construct validity of instruments measuring impairments in body structures and function in rheumatic disorders: Which constructs are selected for validation? A systematic review

R.A.H.M. Swinkels[1], L.M. Bouter[1], R.A.B. Oostendorp[2,3], I.J.C.M. Swinkels-Meewisse[4], P.U. Dijkstra[5], H.C.W. de Vet[1]

[1]VU University Medical Center, Institute for Research in Extramural Medicine, Amsterdam; [2]Research Centre for Allied Health Care, Dep. Quality of Care Research, Nijmegen; [3]Dutch Institute of Allied Health Care, Amersfoort; [4]Department of Medical, Clinical, and Experimental Psychology, University of Maastricht; [5]Center for Rehabilitation, University Medical Center Groningen, Groningen, The Netherlands.

Raymond A.H.M. Swinkels, PhD, PT, MT; Prof. Lex M. Bouter, PhD; Prof. Rob A.B. Oostendorp, PhD, PT, MT; Ilse J.C.M. Swinkels-Meewisse, MSc PT, MT; Pieter U. Dijkstra, PhD, PT, MT; Prof. Henrica C.W. de Vet, PhD.

Please address correspondence to: Raymond A.H.M. Swinkels, Ulenpas 80, 5655 JD Eindhoven, The Netherlands. E-mail: swinky@xs4all.nl

Received on May 16, 2005; accepted in revised form on September 22, 2005.

Clin Exp Rheumatol 2006; 24: 93-102.

Key words: Systematic review, validity, outcome assessment, clinimetrics, methodological quality.

ABSTRACT

Purpose. *This paper focuses on the construct validity of instruments measuring impairments in body structures and function in rheumatic disorders. The objective is: 1) to make an inventory of constructs, based on the domains of the International Classification of Functioning, Disabilities and Health problems (ICF), against which instruments measuring impairments in body structures and function were validated; 2) to analyse whether validation against a similar construct resulted in higher correlation coefficients than validation against a dissimilar construct.*

Methods. *In a systematic review papers were identified in which instruments measuring impairments in body structures and function for patients with rheumatic disorders were validated. The instruments identified were assessed on their methodological properties and the constructs against which they were validated. Subsequently, pooled (interclass) correlations of similar constructs and dissimilar constructs against which was validated were compared. An instrument was decided to have good construct validity, if the correlation coefficient was 0.50 or higher, and the measurement instrument in question is validated against similar constructs.*

Results. *In total 216 papers were identified analysing the validity of 42 different instruments. Only 16% of these instruments were validated against instruments that represent the most similar construct. In general, estimates of construct validity were lower when validated against dissimilar constructs, except for instruments measuring impairments in mental functions.*

Conclusion. *There is a trend that validation against a similar construct yields higher correlation coefficients than validation against a dissimilar construct. If an instrument measuring impairments is validated against the most similar construct, and a criterion of r > 0.50 is applied, only 10 out of the 42 identified instruments turned out to be valid.*

Introduction

This paper focuses on the construct validity of instruments measuring impairments in body structures and function in patients with rheumatic disorders. In the past decades the content and methodological quality of clinical outcome measures have increasingly become the focus of research. Methodological quality includes standardisation of measurement, reliability, responsiveness and validity.

Several kinds of validity are described in literature, e.g. face validity, criterion validity, and construct validity. Face validity is the validity based on its appearance to the observer. Criterion validity is validity based on comparison with a gold standard. However, in the majority of measurement instruments there is no gold standard available (1). In these instances the focus is on the construct validity in which the correlation between the instrument under study with other instruments measuring this construct is assessed. The process of construct validation presents a considerable challenge to the researcher, because many constructs are multidimensional, (for example personal care, or quality of life) and therefore it is not easy to determine whether an instrument is actually measuring all aspects of the construct of interest (1). Surprisingly, this dilemma is seldom

discussed in validity studies. In the majority of validation studies there is a lack of motivation why a certain construct is chosen to validate against and there is lack of reflection whether the comparative instrument actually measures the construct it is intended to measure.

The kind of construct used for validation could cause considerable differences in the correlation with scores on the measure to be validated. For instance in patients with rheumatoid arthritis, the Visual Analogue Scale for pain has been validated against the Articular Index, but also against the Keitel Function Index and against the Hemoglobine-proportion. The resulting correlations were 0.86, 0.20, and 0.21, respectively, indicating that the choice of the construct may influence the strength of the relationship found (2-4).

Consequences of rheumatic disorders can be considered at different domains: at the level of tissues/organisms (impairments in body structures and function), at the level of limitations in daily activities (disabilities), and at the level of participation in social life, including work and hobby's. However, in the majority of validation studies concerning the consequences of rheumatic disorders the validated measures were compared with measures focussing on various constructs, ignoring attention for the degree of similarity or the content of the domains. As shown in the example above, validating against a dissimilar construct may result in lower correlations than validating against a similar construct.

The International Classification of Functioning, Disabilities and Health problems (5) (ICF), which is often used in the field of physiotherapy and rehabilitation medicine, differentiates between the domains mentioned above. Therefore the ICF is the basis for our attempt to differentiate between constructs by classifying the constructs in the domains of the ICF[#].

By using the ICF-classification as a system to qualify constructs one might observe that, for example, a measure for impairments in body structure and function is validated against a measure for limitations in activities.

Besides the choice of the construct to validate against, there is a lack of consensus about which magnitude of the correlation coefficients is acceptable for validity and which is not. Correlation coefficients between 0.35 and 0.45 have been considered acceptable (6). But a correlation of 0.40 has also been defined as 'reasonably high' (7). Whereas others state that a rho > 0.40 and p < 0.001 is cut-off point for acceptable validity (8). While it is clear what the maximum correlation is, a perfect correlation in validity studies is not desirable, because it would point at redundancy(9).

Our hypothesis is that validating against a dissimilar construct results in lower correlation coefficients than validating against a similar, optimally comparable, construct. For more similar constructs it can be argued that the cut-off point for 'good' validity should be higher than that for dissimilar constructs. This should be taken into account if one wants to define a priori cut off points for acceptable construct validity.

The aims of our systematic review were: 1) to investigate the constructs against which measures of impairments in body structures or functions for rheumatic disorders are validated; 2) to examine whether validation against the same or a similar construct results in higher correlation values than validation against a dissimilar construct.

Methods and materials

A systematic literature review of the methodological aspects of measurement instruments was performed for all relevant impairments in body structures and functions related to rheumatic disorders. The 'rheumatic disorders' included were: rheumatoid arthritis, seronegative polyarthritis (including psoriatic arthritis), osteoarthritis, ankylosing spondylitis, polymyositis and fibromyalgia. An impairment in body structure and function was defined as 'relevant' if there was consensus among an expert panel (two experienced rheumatologists, clinically as well as in research, and two collaborators of the WHO-working group ICF) that the impairment is present in the majority of patients with rheumatic disorders.

In a parallel paper, similar methodological aspects of instruments measuring limitations in activities and participation were analysed (10).

Literature search

The Medline database was searched for the period January 1982 - April 2001, using specific search terms for the relevant rheumatic disorders and various search terms for clinimetric properties. The database of the Centre for Documentation of the Dutch National Institute of Allied Health Professions was also searched for the period January 1988 – April 2001. In this search the keywords 'clinimetrics', 'assessment and methods', and all search terms for 'measurement instruments', were used. Additionally the names of the relevant impairments, and of the rheumatic disorders were used as free text words. Finally the names of measurement instruments identified in the first searches were used as free text words in additional searches in the databases. Papers written in English, French, German and Dutch were included. The search was subsequently augmented with a manual search based on the reference list of the identified papers. This final search yielded a small number of papers written before 1982.

Inclusion

Papers were included in the current review if they were performed in populations with the above mentioned rheumatic disorders and if they generated information about the clinimetric properties of the measurement instruments. No restrictions were applied with respect to study design. Further eligibility criteria were: 1) instruments should focus mainly (50% or more of the questions or sub-scales) on impairments in body structures and function; 2) instru-

[#] The ICF was still under construction during the research period of this project; in fact we used the precursor of the ICF: The International Classification of Impairments, Disabilities and Handicaps (ICIDH). For this paper we used the ICF terminology: 'impairment' is an ICIDH-term which is substituted by the ICF-term 'impairment in body structures and function'; 'disability' (ICIDH) is substituted by 'limitations in activities' (ICF) and 'handicaps' (ICIDH) is substituted by 'participation' (ICF).

ments should contain a sub-scale for the impairment that might be interpreted separately as a single entity, independently from the other parts of the questionnaire or other sub-scales.

Data extraction

All papers selected were assessed by two reviewers independently according to a standardised scoring form (11), which was modified for rheumatic disorders. The assessment of the described measurement instruments included description of the construct against which was validated (based on the domains of the ICF: see Table I), and recording of the correlation measures between the used constructs and the measurement instrument to be validated. Correlation coefficients were scored for entire measurement instruments, as well as for sub-scales, whenever separate information about the construct validity of the sub-scales was available. In case of disagreement between the observers (3% of cases), the paper was also assessed by a third reviewer.

All names of instruments and subscales used in this review are abbreviated (Appendix). References for each instrument are included in the tables to identify the relevant literature per measurement instrument.

Construct validity was classified into six levels of constructs, corresponding to the degree of (dis)similarity of the constructs (based of the ICF) against which a measurement instrument was validated (Table I). A measurement instrument was assigned a numerical rating ranging from1 to 6, based on this classification system.

Level 1 is the most similar construct (for example: a pain intensity questionnaire is validated against a Visual Analogue Scale for pain). Levels 5 and 6 constructs are the most dissimilar. Level 5 construct is an imperfect construct to validate against, because the construct includes all domains of the ICF instead of the domain the instrument intends to measure (for example: a pain intensity questionnaire is validated against disabilities in inter-personal relationships). In level 6, measures for assessment of impairments in body structures and function were validated with constructs concerning general characteristics such as age, gender, duration of complaints, etc. Therefore, validation with a level 6 construct means that it is validated against the most dissimilar construct. In this context, the words 'similar' and 'dissimilar' give no indication of the strength of the correlation coefficient (quantification); they qualify whether a measurement instrument is validated against a similar construct, or against a dissimilar construct. It is hypothesized that validation against a similar construct will result in a high correlation, and validation against a dissimilar construct will result in a low correlation.

The strengths of the correlations in the validations studies were qualified as follows, a correlation coefficient < 0.50 was considered 'poor', a correlation coefficient between 0.50 and 0.65 was considered 'moderate', and a correlation coefficient ≥ 0.65 was considered 'good' (12).

A measurement instrument was labelled to have good construct validity, if two conditions were met: 1) the measurement instrument in question is validated against similar constructs, (level 1 and level 2), and 2) the correlation coefficient was 0.50 or higher.

Data analysis

Pooling of the data was performed within the construct level for each measurement instrument or subscale. The pooled index was:

$$X = (\Sigma[n_i * x_i])/N$$

where X = pooled index, n_i = number of persons included in the study, x_i = value of correlation coefficient (Pearson's r, Spearman's rho) in the study, N = total number of persons in all studies focusing on a specific measurement instrument in one of the construct levels. The pooled index was computed for the values of the correlation coefficients (Pearson's r, Spearman's rho). ICC-values are pooled separately in the same way. In measurement instruments with various subscales values for the validity of instruments may be strongly influenced by the values of one or more sub-scales: if one of the subscales of a questionnaire is poorly correlated with the measure to be validated, it decreases the validity of the total questionnaire, whereas the other subscales might posses good validity. Therefore, the data were pooled for the separate sub-scales as far as possible and whenever they were available.

Table I. Definition of construct level.

Level of construct	Definition
Level 1	Validation against sub-scales or instruments that measure the most similar construct
Level 2	Validation against instruments that measure the same construct (impairment in body structures and function) as well as other impairments or disabilities
Level 3	Validation against instruments that measure other impairments in body structures and function than the ones to be validated
Level 4	Validation against instruments that measure limitations in activities instead of the domain to be validated (impairments in body structures and function)
Level 5	Validation against general measurement instruments that measure all domains of the ICF
Level 6	Validation against general characteristics like age, sex etc.

Results

In total 319 papers were identified. The references mentioned in these papers were checked for potential eligibility and relevance by reading title and abstract. This procedure generated an additional 405 papers resulting in 724 papers which were assessed. Of these 724 papers 216 fulfilled the inclusion criteria for detailed reviewing. The number (percentage) of validation studies of the different impairment constructs, together with the levels of the constructs against which they were validated, are shown in Table II.

The most frequently validated instruments are those intended to assess impairments in mental functions (158

Table II. Frequency of validation of measurement instruments for the assessment of impairments in body structures and function in patients with rheumatic disorders against different levels of construct.

Impairments	Instruments and subscales (no.)	No. of times an instrument was validated (no.)	Level 1 no.(%)	Level 2 no.(%)	Level 3 no.(%)	Level 4 no.(%)	Level 5 no.(%)	Level 6 no.(%)
Mental functions	9	158	24 (15%)	0	0	75 (47%)	24 (15%)	35 (22%)
Stiffness	4	45	3 (7%)	2 (4%)	0	28 (63%)	7 (14%)	5 (12%)
Sensory functions*	13	210	43 (20%)	19 (9%)	0	78 (37%)	21 (10%)	49 (23%)
Mobility	11	64	9 (14%)	0	0	36 (57%)	7 (11%)	12 (18%)
Muscle force	3	61	13 (21%)	0	0	36 (59%)	1 (1%)	11 (17%)
Swelling	2	62	2 (3%)	2 (3%)	0	10 (16%)	0	48 (77%)
Total	42	600#	94 (15.7%)	23 (3.8%)	0	263 (44%	60 (10%)	160 (27%

Levels indicate level of construct as defined in Table I.
The percentages are row percentages based on the numbers presented in the third column.
* Sensory functions including pain
# In total 216 articles were included, which generated 600 validity calculations, because in the majority of articles the measurement instrument was validated against more than one construct.

times) and those assessing sensory functions including pain (210 times). The majority (44%; n = 263) of the instruments and/or subscales was validated against level 4 construct. Furthermore, 27% (n = 160) of all investigated instruments was validated against level 6 construct. None of the instruments was validated against level 3 construct, and only a small number of instruments (16%; n = 94) were validated against the most similar construct (level 1).

Correlation coefficients per measurement instrument and subscales, when available, are presented for each group of impairments per level of construct in Table III. In the table the correlations for the measurement instruments and/-or sub-scales are only included if data concerning at least level 1 and level 2 construct, or level 1 construct as well as level 4, 5 or 6 construct were available. This presentation enables comparison of the validity of the most similar construct with the more dissimilar construct. Only for the measurement instruments of mobility (Table III) all correlations are presented, as no comparison was possible between the most similar and the most dissimilar construct validity because of lack of studies providing these data. If an instrument measuring impairments is validated against the most similar construct, and a criterion of $r>0.50$ is applied, only 10 out of 42 instruments turn out to be valid.

Comparison between level 1 or 2 construct and level 4, 5 and/or 6 was possible for 22 instruments. In 10 out of these 22 instruments (45%) correlations (Pearson's r or Spearman's rho) were stronger when validated against level 1 or 2 construct than when validated against level 4-5-6 construct (Table III, indicated in dark grey). For 7 (32%) instruments correlation coefficients were similar or weaker (Pearson r or Spearman's rho) when validated against level 1 or 2 construct compared to validation against level 4-5-6 construct (Table III, indicated in light grey). For 5 instruments the results were conflicting (Table III, indicated in middle grey).

Discussion

The aim of the current study was to investigate whether the level of constructs against which an instrument for assessment of impairments in body structures or function was validated, influences the value of the correlation coefficients. Until now this is the first study that focuses specifically on the level of the constructs that are used for validation of measurement instruments.

The results show that instruments measuring impairments in body structures and functions were most often (44%) validated against a level 4 construct (instruments that measure other impairments in body structures and functions than the one to be validated), while 27% was validated against a level 6 construct. On the average only 16% of the instruments was validated against sub-scales or instruments that measure the most optimal comparable (level 1) construct. The validation against a dissimilar construct did not always result in lower correlation coefficients than validation against a similar construct. In validation of instruments for the assessment of impairments in body structures and function we found earlier that this hypothesis is correct (100). However, in validation of instruments for the assessment of disabilities in personal care and instruments for the assessment of disabilities in gait and gait-related activities this relationship was not convincingly present (10; 101).

We wanted to compare correlations of the instruments when validated against different levels of construct. It is therefore possible that some well known and valid instruments (i.e. the HAQ and the WOMAC) were not included in this review, i.c. measurement instruments that are not validated against construct 1-2 as well as against construct 4-5-6, are not included in our data, because in that situations it was not possible to compare the influence of the (dis)similarity of the constructs. Furthermore, some frequently used questionnaires, like the Health Assessment Questionnaire (HAQ), or the Western Ontario McMaster Osteoarthritis Index (WO-

Table III. Correlation coefficients for construct validity of instruments for the assessment of mental functions, stiffness, pain, mobility, muscle force and swelling.

Instrument and sub-scale [@]	Level 1 Construct	Level 2 Construct	Level 4 Construct	Level 5 Construct	Level 6 Construct	Ref.
Mental functions						
AHI	0.46*		0.69*	0.23	0.35	(13; 14)
AIMS-anxiety	0.43*		0.19*	0.45*	0.13*	(15-37)
AIMS-depri	0.43*		0.24*	0.45*	0.23*	see AIMS
AIMS-EmoF	0.57*		0.25*			see AIMS
AIMSS-anxiety	0.41*		0.47	0.45*		(32)
AIMSS-depri	0.47*		0.57	0.51*		(32)
BDI	0.67*				0.82	(14; 38)
SSAI	0.57*					(14)
STAI	0.68*					(14)
Stiffness						
BASDA	0.69*					(39-41)
MS-D	0.26		0.16*	0.19*	-0.27	(4; 21; 34; 42; 43)
MS-S	0.51		0.37*	0.28*		(4; 21; 34; 42; 43)
VAS-S			0.91*			(44)
Pain						
AI	0.83*		0.24*	0.42	0.38*	(2; 4; 21; 34 ;45-53)
AIMS	0.58*			0.75*	0.54*	(14-20 ;22-27 ;29-37; 54)
AIMSD			0.65	0.65	0.43	(19; 28; 31; 35; 54-57)
AIMS-pain	0.40*	0.54*	0.39*	0.60*	0.44*	see AIMS
AIMSD-pain		0.63*		0.81	0.49*	see AIMSD
AIMS2-pain		0.49	0.38*		0.56	(23; 58; 59)
AIMS2Dpain	0.66	0.21	0.33*	0.40	0.31	(59)
AIMSSpain	0.41*		0.50*	0.61*	0.46	(32)
Dol	0.79*					(60-69)
EI	0.67		0.36*			(70)
MPQ	0.37					(27; 71-74)
SAJ	0.55*		0.63			(48; 53)
VAS-P	0.83*		0.26*	0.73		(4; 34; 43; 60; 75-77)
Mobility						
BASMI	0.93*					(40; 41; 78)
Chest			0.60			(79-81)
EPM			0.54	0.54*		(56; 82; 83)
Gonio	0.92					(84)
MobSpine-CCD			0.37			(79-81; 85)
MobSpine-OWD			0.49			(79-81; 85)
MobSpine-VitCp			0.23			(79-81; 85)
Shob			0.66*		0.19	(80; 81; 85; 86)
Spond	0.92					(84; 87)
Stest						(58; 88)
FFD			0.18			(80; 81; 85; 87)
Muscle force						
Gripp			0.31*		-0.33	(21; 43; 89-91)
MSI			0.24			(92)
Sphy	0.87*		0.35*	0.87*	0.21*	(34; 93-98)
Swelling						
AI	0.88*		0.35*	0.43*		(2-4; 21; 34; 45-52; 99)
SAJ		0.64*	0.63			(48; 53)

@ for explanation of abbreviations see Appendix 1
Level 1 construct: level of the construct (see Table 1) against which the instrument is validated. All values expressed in Pearson's r or Spearmen's rho
*: pooled value
Ref: reference numbers
Dark grey = stronger correlations if validated against level 1 or 2 construct than when validated against level 4-5-6 construct.
Light grey = similar or weaker correlations if validated against level 1 or 2 construct compared to validation against level 4-5-6 construct.
Middle grey = conflicting results in correlations between validation against level 1-2 construct versus level 4-5-6 construct.

MAC) were not included, because the main focus of these instruments is not on impairments in body structures and functions. Beside that, instruments that did not include or present separate data for a subscale impairment were excluded from analysis.

Recently, an ICF Core Set was developed for rheumatoid arthritis (102), which is part of a series of Core Sets (osteoarthritis, osteoporosis, low back pain and widespread pain) (103-106). The ICF Core Set for rheumatoid arthritis describes the domains of the ICF classification that are considered to be most relevant, but does not focus on specific measurement instruments. The ICF Core Set can serve as a framework for comparison of the results of the current study with that Core Set: both are based on the ICF classification. New in our study is the fact that we examined in which way instruments for the assessment of impairments in body structures and function are validated and whether the level of constructs used for validation influences the outcome of the study. We would like to emphasize that the aim of the current study was not to identify the most valid instruments.

In about half of the studies, the correlations for all measurement instruments were stronger if validated against the most similar construct. Sometimes the correlations of instruments that were validated against constructs 5-6 level (general; multidimensional) are stronger. However, this phenomenon mainly occurred in the multidimensional AIMS, in particular AIMS-anxiety, AIMS-depri and AIMS-EmoF. On the basis of our systematic review validity of the AIMS subscales 'depression' and 'anxiety' might be reconsidered. These subscales might not exactly cover the construct they are intended to measure. 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 I). Possibly the proposed underlying factors (which are also incorporated in the different items of a questionnaire) are not fully correct, which might explain the found differences in correlations.

From the results of the current study, it occurred that the correlations for construct validity are, in general, weaker when validated against the most dissimilar construct (level 6 construct), compared to the correlations if validated against level 1 construct or level 2 construct. On the other hand, the correlations for the construct validity of instruments to measure impairments in body structures and function were in seven out of 22 cases similar or stronger, when validated against construct 4-5-6 (in general multidimensional instruments), compared to validation against constructs 1 or 2. An explanation for this phenomenon might be that weak correlations of certain sub-scales are compensated by strong correlations of other sub-scales that are more similar with the general construct against which they are validated. Based on this assumption it can be concluded that multidimensional questionnaires need to be validated against multi-dimensional constructs, or separately for each sub-scale against similar constructs.

Another explanation for the discrepancy found might be the correlation between impairments in body structures and function on one hand, and limitations in activities on the other hand. For example, the correlation between pain and disabilities in daily living is r = 0.39 (107). Probably more than one impairment is present in patients with a rheumatic disorder. Beside pain high scores on measures of disease activity, stiffness and swollen joints might result in considerable disability in activities of daily living. Finally it is possible that the construct of the measurement instrument to be validated actually covers another (or broader) construct than it is supposed to do and therefore might result in a weaker correlation than expected if validated against a dissimilar construct.

At the moment of our study the ICIDH-2 was the most recent official version accorded by the World Health Organisation (the ICF was still under construction). Fortunately, the content of the constructs was not changed by the ICF-terminology. As the aim of this study was to compare the different constructs that were used for validation of measurement instruments the changed terminology (from ICIDH to ICF)has not affected our results. As seen in this systematic review the reported validity of a measurement instrument depends on the level of construct against which it is validated. If the differentiation in levels of construct as suggested in this paper is applied, the scores for validation against construct 1 are highest. For that reason, only correlation coefficients > 0.50 were considered acceptable, between 0.50 and 0.65 were considered moderate, and if they were greater than 0.65 they were considered to be good (if validated against level 1 or level 2 construct). If this criterion is applied to instruments of this systematic review, the number of valid instruments decreased considerably. For the measurement of impairments in mental functions, only four instruments meet this validity criterion. For the measurement of stiffness two out of four instruments met this validity criterion. For the assessment of pain all instruments except the MPQ and AIMS-pain met this criterion. For the assessment of mobility of joints only BASMI, Gonio and Spond met the validity criterion, and for the assessment of muscle force only the Sphy met the validity criterion (Table III).

Validation against a dissimilar (imperfect) construct resulted in lower validity values, thus different 'cut-off points' should be used to decide whether to accept or to reject a certain validity. Currently, for assessing construct validity it is advocated to formulate hypotheses about the correlation of a scale under study with other instruments (108). 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 I).

In these hypotheses one should take into account that high correlations can only be expected with instruments measuring a similar construct in the same domain (level 1 or 2). Useful hypotheses are, for example: 'the correla-

tion of an impairment scale is higher with another impairment scale than with an activity scale (a more dissimilar construct)', or 'the correlation of a specific subscale of an instrument with a similar subscale of another instrument is higher than with the other subscale of that instrument' (109).

There is clearly a need for further research, in which hypotheses are formulated about the correlation to be expected, taking into account the extent of similarity of the constructs to be compared.

Conclusions

1) A majority of 80% of the investigated instruments are validated against dissimilar constructs that measure another impairment and/or an activity, or other aspects than the domain to be validated.
2) An average of 16% of instruments measuring impairments is validated against sub-scales or instruments that measure the most similar construct. The available data show that in about 50% the values for construct validity are much lower if validated against a dissimilar construct, compared to validation against the most similar construct or the same impairment combined with (an)other impairment(s) in body structures and function.
3) Multidimensional questionnaires need to be validated multi-dimensionally, or separately for each subscale against other constructs.

Acknowledgements

We kindly thank Jan Pool for his critical comments on previous versions of this manuscript.

References

1. PORTNEY LG, WATKINS MP: *Foundations of Clinical Research. Applications to Practice*. Appleton and Lange, 2000: 1-722.
2. GASTON-JOHANSSON F, GUSTAFSSON M: Rheumatoid arthritis: determination of pain characteristics and comparison of RAI and VAS in its measurement. *Pain* 1990; 41: 35-40.
3. KALLA AA, KOTZE TJ, MEYERS OL, PARKYN ND: Clinical assessment of disease activity in rheumatoid arthritis: evaluation of a functional test. *Ann Rheum Dis* 1988; 47: 773-9.
4. MALLAYA RK, MACE BE: The assessment of disease activity in rheumatoid arthritis using a multivariate analysis. *Rheumatol Rehabil* 1981; 20: 14-7.
5. WORLD HEALTH ORGANIZATION: *International Classification of Functioning, Disabilities and Health Problems*. Geneva, World Health Organization, 2001.
6. GHISELLE EE: *The Measurement of Occupational Aptitude*. 1st ed., Los Angeles, Berkeley, 1966.
7. HUISKES CJAE, KRAAIMAAT FW, BIJLSMA JW: Development of a self-report questionnaire to assess the impact of rheumatic diseases on health and lifestyle. *J Rehab Sciences* 1990; 3: 65-70.
8. MARTIN DP, ENGELBERG R, AGEL J, SWIONTKOWSKI MF: Comparison of the Musculoskeletal Function Assessment Questionnaire 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 1997; 79. A (9, sept.).
9. DEYO RA, INUI TS: Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Services Research* 1984; 19: 275-89.
10. SWINKELS RA, OOSTENDORP RA, BOUTER LM: Which are the best instruments for measuring disabilities in gait and gait-related activities in patients with rheumatic disorders. *Clin Exp Rheumatol* 2004; 22: 25-33.
11. OOSTENDORP RAB, ELVERS JWH, SWINKELS RAHM: Quality scoring list for reading articles. (*Unpublished*) 1997.
12. ELIASZIW M, YOUNG SL, WOODBURY MG, FRYDAY-FIELD K: Statistical methodology for the concurrent assessment of inter-rater and intra-rater reliability: Using goniometric measurements as an example. *Physical Therapy* 1994; 74: 777-88.
13. NICASSIO PM, WALLSTON KA, CALLAHAN LF, HERBERT M, PINCUS T: The measurement of helplessness in rheumatoid arthritis. The development of the arthritis helplessness index. *J Rheumatol* 1985; 12: 462-7.
14. HAGGLUND KJ, ROTH DL, HALEY WE, ALARCON GS: Discriminant and convergent validity of self-report measures of affective distress in patients with rheumatoid arthritis. *J Rheumatology* 1989; 16: 1428-32.
15. MEENAN RF, ANDERSON JJ, KAZIS LE *et al*.: Outcome assessment in clinical trials. Evidence for the sensitivity of a health status measure. *Arthritis Rheum* 1984; 27: 1344-52.
16. MEENAN RF, GERTMAN PM, MASON JH: Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 1980; 23: 146-52.
17. KAZIS LE, ANDERSON JJ, MEENAN RF: Effect sizes for interpreting changes in health status. *Med Care* 1989; 27 (3 Suppl.): S178-S189.
18. ANDERSON JJ, FIRSCHEIN HE, MEENAN RF: Sensitivity of a health status measure to short-term clinical changes in arthritis. *Arthritis Rheum* 1989; 32: 844-50.
19. HIDDING A, SANTEN MV, KLERK ED, GIELEN X, BOERS M: Discordance between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. Group physical therapy in ankylosing spondylitis (thesis). *Meppel, Krips Repro*, 1993: 94-104.
20. BUCHBINDER R, BOMBARDIER C, YEUNG M, TUGWELL P: 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* 1995; 38: 1568-80.
21. ANDERSON JJ, CHERNOFF MC: Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *J Rheumatol* 1993; 20: 535-7.
22. POTTS MK, BRANDT KD: Evidence of the validity of the Arthritis Impact Measurement Scales. *Arthritis Rheum* 1987; 30: 93-6.
23. MEENAN RF, MASON JH, ANDERSON JJ, GUCCIONE AA, KAZIS LE: AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum* 1992; 35: 1-10.
24. FITZPATRICK R, ZIEBLAND S, JENKINSON C, MOWAT A, MOWAT A: A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol* 1993; 20: 429-36.
25. LIANG MH, LARSON MG, CULLEN KE, SCHWARTZ JA: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; 28: 542-7.
26. MEENAN RF, GERTMAN PM, MASON JH, DUNAIF R: The arthritis impact measurement scales. Further investigations of a health status measure. *Arthritis Rheum* 1982; 25: 1048-53.
27. KEEFE FJ, CALDWELL DS, QUEEN KT *et al*.: Pain coping strategies in osteoarthritis patients. *J Consult Clin Psychol* 1987; 55: 208-12.
28. TAAL E, SEYDEL ER, JACOBS J, WIEGMAN O, RASKER JJ: De Nederlandse Arthritis Impact Measurement Scales (DUTCH-AIMS). Bepaling van fysieke en psychosociale gezondheidsaspecten van reumatoide arthritis. *Gedrag & Gezondheid* 1989; 17: 69-74.
29. LIANG MH, FOSSEL AH, LARSON MG: Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990; 28: 632-42.
30. GOEPPINGER J, DOYLE MA, CHARLTON SL, LORIG K: A nursing perspective on the assessment of function in persons with arthritis. *Res Nurs Health* 1988; 11: 321-31.
31. JACOBS JW, OOSTERVELD FG, DEUXBOUTS N *et al*.: Opinions of patients with rheumatoid arthritis about their own functional capacity: how valid is it? *Ann Rheum Dis* 1992; 51: 765-8.
32. WALLSTON KA, BROWN GK, STEIN MJ, DOBBINS CJ: Comparing the short and long versions of the Arthritis Impact Measurement Scales. *J Rheumatol* 1989; 16: 1105-9.
33. DUFFY CM, WATANABE DUFFY KN, GLADMAN DD *et al*.: The utility of the arthritis impact measurement scales for patients with psoriatic arthritis. *J Rheumatol* 1992; 19: 1727-32.

34. HAWLEY DJ, WOLFE F: 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* 1992; 5: 130-6.
35. VLIET VLIELAND TP, VAN DER WIJK TP, JOLIE IM, ZWINDERMAN AH, HAZES JM: Determinants of hand function in patients with rheumatoid arthritis. *J Rheumatol* 1996; 23: 835-40.
36. WEINBERGER M, SAMSA GP, TIERNEY WM, BELYEA MJ, HINER SL: Generic versus disease specific health status measures: comparing the sickness impact profile and the arthritis impact measurement scales. *J Rheumatol* 1992; 19: 543-6.
37. HUSTED J, GLADMAN DD, FAREWELL VT, LONG JA: Validation of the revised and expanded version of the Arthritis Impact Measurement Scales for patients with psoriatic Arthritis. *J Rheumatol* 1996; 23: 1015-9.
38. CALLAHAN LF, KAPLAN MR, PINCUS T: 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* 1991; 4: 3-11.
39. GARRETT S, JENKINSON T, KENNEDY LG, WHITELOCK H, GAISFORD P, CALIN A: A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* 1994; 21: 2286-91.
40. CALIN A, NAKACHE JP, GUEGUEN A, ZEIDLER H, MIELANTS H, DOUGADOS M: Defining disease activity in ankylosing spondylitis: is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? *Rheumatology* (*Oxford*) 1999; 38: 878-82.
41. SPOORENBERG A, VAN DER HD, DE KLERK E *et al*.: A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. *J Rheumatol* 1999; 26: 961-5.
42. VLIET-VLIELAND TPM, ZWINDERMAN AH, BREEDVELD FC, HAZES JMW: Measurement of morning stiffness in rheumatoid arthritis clinical trials. *J Clin Epidemiol* 1997; 50.
43. RHIND VM, BIRD HA, WRIGHT V: A comparison of clinical assessments of disease activity in rheumatoid arthritis. *Ann Rheum Dis* 1980; 39: 135-7.
44. OOSTERHOF J: Research on the reproducibility of the VAS for stiffness in patients with rheumatoid arthritis (in Dutch). (*unpublished*) 1992.
45. RITCHIE DM, BOYLE JA, MCINNES JM *et al*.: Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. *Q J Med* 1968; 37: 393-406.
46. CAMP AV: An articular index for the assessment of rheumatoid arthritis. *Orthopaedics* 1971; 4: 39-45.
47. DOUGADOS M, GUEGUEN A, NAKACHE JP, NGUYEN M, MERY C, AMOR B: Evaluation of a functional index and an articular index in ankylosing spondylitis. *J Rheumatol* 1988; 15: 302-7.
48. PREVOO ML, KUPER IH, VAN'T HOF MA, VAN LEEUWEN MA, VAN DE PUTTE LB, VAN RIEL PL: Validity and reproducibility of self-administered joint counts. A prospective longitudinal followup study in patients with rheumatoid arthritis. *J Rheumatol* 1996; 23: 841-5.
49. HART LE, TUGWELL P, BUCHANAN WW, NORMAN GR, GRACE EM, SOUTHWELL D: Grading of tenderness as a source of interrater error in the Ritchie articular index. *J Rheumatol* 1985; 12: 716-7.
50. DOYLE DV, DIEPPE PA, SCOTT J, HUSKISSON EC: An articular index for the assessment of osteoarthritis. *Ann Rheum Dis* 1981; 40: 75-8.
51. FUCHS HA, BROOKS RH, CALLAHAN LF, PINCUS T: A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum* 1989; 32: 531-7.
52. SMOLEN JS, BREEDVELD FC, EBERL G *et al*.: Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. A*rthritis Rheum* 1995; 38: 38-43.
53. CALVO FA, CALVO A, BERROCAL A *et al*.: Self-administered joint counts in rheumatoid arthritis: comparison with standard joint counts. *J Rheumatol* 1999; 26: 536-9.
54. TAAL E, JACOBS JW, SEYDEL ER, WIEGMAN O, RASKER JJ: Evaluation of the Dutch arthritis Impact Measurement Scales (DUTCH-AIMS) in patients with rheumatoid arthritis. *Br J Rheumatol* 1989; 28: 487-91.
55. OOSTERVELD FGJ, JACOBS JWG, DEUXBOUTS N, RASKER JJ, DEQUEKER J, TAAL E: The DUTCH-AIMS and RA: a good measure for physical functioning (in Dutch). *Ned Tijdschr Fysioth* 1990; 100: 317-9.
56. VLIELAND TP, VAN DEN ENDE CH, BREEDVELD FC, HAZES JM: Evaluation of joint mobility in rheumatoid arthritis trials: the value of the EPM-range of motion scale. *J Rheumatol* 1993; 20: 2010-4.
57. FITZPATRICK R, NEWMAN S, LAMB R, SHIPLEY M: A comparison of measures of health status in rheumatoid arthritis. B*r J Rheumatol* 1989; 28: 201-6.
58. MOLENDIJK M, COLARD I, JACOBS JWG, OOSTERVELD FGJ: Drukpunten bij hte fibromyalgie syndroom; de betrouwbaarheid in de dagelijkse praktijk. *Ned Tijdschr Fysioth* 1991; 101: 31-6.
59. RIEMSMA RP, TAAL ERJJ, HOUTMAN PM, PAASSEN HCV, WIEGMAN O: Evaluation of a Dutch version of the AIMS2 for patients with reumatoid arthritis. *Br J of Rheumatol* 1996; 35: 755-60.
60. WIGERS SH, SKRONDAL A, FINSET A, GÖTESTAM KG: Measuring change in fibromyalgic pain: the relevance of pain distribution. *J Musculoskeletal Pain* 1997; 5: 29-41.
61. MCCARTHY DJ, JR, GATTER RA, PHELPS P: A dolorimeter for quantification of articular tenderness. *Arthritis Rheum* 1965; 44: 551-9.
62. WESSEL J: The reliability and validity of pain threshold measurements in osteoarthritis of the knee. *Scand J Rheumatol* 1995; 24: 238-42.
63. PUTTICK M, SCHULZER M, KLINKHOFF A, KOEHLER B, RANGNO K, CHALMERS A: Reliability and reproducibility of fibromyalgic tenderness, measurement by electronic and mechanical dolorimeters. *J Musculoskeletal Pain* 1995; 3: 3-14.
64. ATKINS CJ, ZIELINSKI A, KLINKHOFF AV *et al*.: An electronic method for measuring joint tenderness in rheumatoid arthritis. *Arthritis Rheum* 1992; 35: 407-10.
65. GUTHRIE NG, EDWORTHY SM, MARTIN L, FRITZLER MJ, KINSELLA TD, RYAN JP: Comparing three techniques for the evaluation of fibromyalgia. *Physiotherapy Canada* 1994; 46: 94-7.
66. DALTROY LH, LARSON MG, ROBERTS NW, LIANG MH: A modification of the Health Assessment Questionnaire for the spondyloarthropathies. *J Rheumatol* 1990; 17: 946-50.
67. SMYTHE HA, BUSKILA D, UROWITZ S, LANGEVITZ P: Control and "fibrositic" tenderness: comparison of two dolorimeters. *J Rheumatol* 1992; 19: 768-71.
68. TUNKS E, CROOK J, NORMAN G, KALAHER S: Tender points in fibromyalgia. PAIN 1988; 34: 11-9.
69. LANGLEY GB, FOWLES M, SHEPPEARD H, WIGLEY RD: A simple pressure dolorimeter for the quantification of joint tenderness in inflammatory arthritis. *Rheumatol Int* 1983; 3: 109-12.
70. MANDER M, SIMPSON JM, MCLELLAN A, WALKER D, GOODACRE JA, DICK WC: Studies with an enthesis index as a method of clinical assessment in ankylosing spondylitis. *Ann Rheum Dis* 1987; 46: 197-202.
71. LANKVELD WV, PAD-BOSCH PV', VAN DE PUTTE LB, STAAK CVD, NÄRING G: Pain from rheumatoid arthritis measured with the Visual Analugue Scale and the Dutch version of the McGill Pain Questionnaire (in Dutch). Ned Tijdschr Geneeskd 1992; 136: 1166-70.
72. NORDENSKIOLD U, GRIMBY G: Assessments of disability in women with rheumatoid arthritis in relation to grip force and pain. *Disabil Rehabil* 1997; 19: 13-9.
73. BURCKHARDT CS, BJELLE A: A Swedish version of the short-form McGill Pain Questionnaire. *Scand J Rheumatol* 1994; 23: 77-81.
74. MELZACK R: The McGill Pain Questionnaire: major properties and scoring methods. *PAIN* 1975; 1: 277-99.
75. DOWNIE WW, LEATHAM PA, RHIND VM, WRIGHT V, BRANCO JA, ANDERSON JA: Studies with pain rating scales. Annals of the *Rheumatic Diseases* 1978; 37: 378-81.
76. LANGLEY GB, SHEPPAERD H: Problems associated with pain measurement in arthritis: comparison of the visual analogue and verbal rating scales. *Clin Exp Rheumatol* 1984; 2: 231-4.
77. HURST NP, KIND P, RUTA D, HUNTER M, STUBBINGS A: Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997; 36: 551-9.

78. JENKINSON TR, MALLORIE PA, WHITELOCK HC, KENNEDY LG, GARRETT SL, CALIN A: Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. *J Rheumatol* 1994; 21: 1694-8.
79. MOLL JM, WRIGHT V: The pattern of chest and spinal mobility in ankylosing spondylitis. An objective clinical study of 106 patients. *Rheumatol Rehabil* 1973; 12: 115-34.
80. HEIKKILA S, VIITANEN JV, KAUTIAINEN H, KAUPPI M: Sensitivity to change of mobility tests; effect of short term intensive physiotherapy and exercise on spinal, hip, and shoulder measurements in spondyloarthropathy. *J Rheumatol* 2000; 27: 1251-6.
81. VIITANEN JV, HEIKKILA S, KOKKO ML, KAUTIAINEN H: Clinical assessment of spinal mobility measurements in ankylosing spondylitis: a compact set for follow-up and trials ? *Clin Rheumatol* 2000; 19: 131-7.
82. FERRAZ MB, OLIVEIRA LM, ARAUJO PM, ATRA E, WALTER SD: EPM-ROM Scale: an evaluative instrument to be used in rheumatoid arthritis trials. *Clin Exp Rheumatol* 1990; 8: 491-4.
83. ENDE CHMVD, VLIET-VLIELAND TPM, BREEDVELD FC, HAZES JM: The EPM-ROM scale: a valuable measurement instrument for the mobility of the joints of patients with rheumatoid arthritis (in Dutch). *Ned Tijdschr Fysioth* 1994; 104: 86-91.
84. REYNOLDS PM: Measurement of spinal mobility: a comparison of three methods. *Rheumatol Rehabil* 1975; 14: 180-5.
85. VIITANEN JV, KAUTIAINEN H, SUNI J, KOKKO ML, LEHTINEN K: The relative value of spinal and thoracic mobility measurements in ankylosing spondylitis. *Scand J Rheumatol* 1995; 24: 94-7.
86. VIITANEN JV: Thoracolumbar rotation in ankylosing spondylitis. A new noninvasive measurement method. *SPINE* 1993; 18: 880-3.
87. STURROCK RD, WOJTULEWSKI JA, HART FD: Spondylometry in a normal population and in ankylosing spondylitis. *Rheumatol Rehabil* 1973; 12: 135-42.
88. ROBERTS WN, LIANG MH, PALLOZZI LM, DALTROY LH: Effects of warming up on reliability of anthropometric techniques in ankylosing spondylitis. *Arthritis Rheum* 1988; 31: 549-52.
89. NORDENSKIOLD UM, GRIMBY G: Grip force in patients with rheumatoid arthritis and fibromyalgia and in healthy subjects. A study with the Grippit instrument. *Scand J Rheumatol* 1993; 22: 14-9.
90. DELLHAG B, BJELLE A: A Grip Ability Test for use in rheumatology practice. *J Rheumatol* 1995; 22: 1559-65.
91. DELLHAG B, HOSSEINI N, BREMELL T, INGVARSSON PE: Disturbed grip function in women with rheumatoid arthritis. *J Rheumatol* 2001; 28: 2624-33.
92. STUCKI G, SCHONBACHLER J, BRUHLMANN P, MARIACHER S, STOLL T, MICHEL BA: Does a muscle strength index provide complementary information to traditional disease activity variables in patients with rheumatoid arthritis? *J Rheumatol* 1994; 21: 2200-5.
93. HAMILTON GF, MCDONALD C, CHENIER TC: Measurement of grip strength: validity and reliability of the sphygmomanometer and Jamar Grip Dynamometer. *J Orthop Sports Phys Ther* 1992; 16: 215-9.
94. HELEWA A, GOLDSMITH CH, SMYTHE HA: The modified sphygmomanometer-an instrument to measure muscle strength: a validation study. *J Chronic Dis* 1981; 34: 353-61.
95. HELEWA A, GOLDSMITH CH, SMYTHE HA: 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* 1986; 13: 1044-9.
96. SPIEGEL JS, PAULUS HE, WARD NB, SPIEGEL TM, LEAKE B, KANE RL: What are we measuring? An examination of walk time and grip strength. *J Rheumatol* 1987; 14: 80-6.
97. LEE P, BAXTER A, DICK WC, WEBB J: An assessment of grip strength measurement in rheumatoid arthritis. *Scand J Rheumatol* 1974; 3: 17-23.
98. PINCUS T, BROOKS RH, CALLAHAN LF: Reliability of grip strength, walking time and button test performed according to a standard protocol. *J Rheumatol* 1991; 18: 997-1000.
99. PREVOO ML, RIEL PLCMV, HOF MAV' et al.: 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*. Nijmegen, Catholic University Nijmegen, 1995: 19-31.
100. SWINKELS RAHM, BOUTER LM, OOSTENDORP RAB, ENDE CHMVD: Impairment measures in rheumatic disorders for rehabilitation medicine and allied health care: a systematic review. *Rheum Int* 2005; 25: *In press*.
101. SWINKELS RAHM, DIJKSTRA PU, BOUTER LM: Reliability, validity and responsiveness of instruments to assess disabilities in personal care in patients with rheumatic disorders: a systematic review. *Clin Exp Rheumatol* 2005; 23: 71-9.
102. STUCKI G, CIEZA A, GEYH S *et al.*: ICF Core Sets for rheumatoid arthritis. *J Rehabil Med* 2004; (44 Suppl.): 87-93.
103. DREINHOFER K, STUCKI G, EWERT T *et al.*: ICF Core Sets for osteoarthritis. *J Rehabil Med* 2004; (44 Suppl.): 75-80.
104. CIEZA A, SCHWARZKOPF S, SIGL T *et al.*: ICF Core Sets for osteoporosis. *J Rehabil Med* 2004; (44 Suppl.): 81-6.
105. CIEZA A, STUCKI G, WEIGL M *et al.*: ICF Core Sets for low back pain. *J Rehabil Med* 2004; (44 Suppl.): 69-74.
106. CIEZA A, STUCKI G, WEIGL M *et al.*: ICF Core Sets for chronic widespread pain. *J Rehabil Med* 2004; (44 Suppl.): 63-8.
107. WADDELL G: *The Back Pain Revolution*. 1 ed., Edinburgh, Churchill Livingstone, 1998.
108. STREINER DL, NORMAN GR: *Health Status Measurement Scales. A Practical Guide to Their Development and Use*. 2nd ed., Oxford, Oxford University Press, 2003: 1-231.
109. ROORDA LD, ROEBROECK ME, LANKHORST GJ, VAN TILBURG T, BOUTER LM: Measuring functional limitations in rising and sitting down: development of a questionnaire. *Arch Phys Med Rehabil* 1996; 77: 663-9.

Appendix: 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 Scale 2 Dutch	Anx = anxiety Depri = 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
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	
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 duration Stiff-S = Stiffness severity
MS-S	Morning stiffness, severity	
P-NRS	Pain Numeric Rating Scale	
SAJ	Self Assessment Joint Count	
Shob	Shober Test	
Sphy	Sphygmomanometer	
Spond	Spondylometer (Dunham)	
SSAI	Spielberger State-Anxiety Inventory	
STAI	Spielberger Trait-Anxiety Inventory	
Stest	Stiffness test	
VAS-P	Visual Analogue Scale Pain	