Which are the best instruments for measuring disabilities in gait and gait-related activities in patients with rheumatic disorders

R.A.H.M. Swinkels[1], R.A.B. Oostendorp[1], L.M. Bouter[2]

[1]University Medical Centre Nijmegen, Centre for Quality of Care Research, Nijmegen; [2]VU University Medical Centre, Institute for Research in Extramural Medicine, Amsterdam, The Netherlands.

Raymond A.H.M. Swinkels, MSc, PT, MT; Prof. Rob A.B. Oostendorp, PhD, PT, MT; Prof. Lex M. Bouter, PhD.

This study is supported by a grant from the Health Insurance Executive Board (CvZ), no. CURE/24101/97.

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

Received on September 9, 2003; accepted in revised form on December 18, 2003.

Clin Exp Rheumatol 2004; 22: 25-33.

Key words: Measurement, gait, walking, rheumatic diseases.

ABSTRACT

Objectives. *Our first objective was to make an inventory of available instruments for the assessment of disabilities in gait and related activities in patients with rheumatic disorders. Our second aim was to investigate which of these instruments have acceptable methodological quality with regard to reliability and validity. Our third aim was to investigate the assumption that the evaluation of convergent construct validity results in stronger correlations when validated against a more similar construct.*

Materials. *A computer-aided literature search (1982-2001) of several databases was performed to identify studies focusing on the clinimetric properties of instruments to assess impairments in function in patients with rheumatic disorders. Data on intra-rater reliability, inter-rater reliability and convergent construct validity were extracted in a standardized manner and compared to a priori defined criteria.*

Results. *In total 78 instruments were eligible. Intra-observer reliability was investigated for 28 instruments and only 7 demonstrated good reliability as well as good validity. Surprisingly, the convergent construct validation against a similar construct resulted often in lower correlations than validation against a less similar construct.*

Conclusion. *Based on the available information, the Rheumatoid Arthritis Quality of Life Scale and the Health Assessment Questionnaire seem to be the best instruments for assessing disabilities in gait and related activities in patients with rheumatic disorders.*

Introduction

In rehabilitation 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 renamed the International Classification of Functioning, Disabilities and Health Problems (ICF) (3) has encouraged such a shift.

The main reason for this is that in rheumatology and allied health care, suitable treatment goals cannot be directly derived from diseases, but should be mainly based on the consequences of diseases (4, 5). The ICF-classification distinguishes the disease itself from disease consequences and influencing contextual factors (4). If the treatment goals are expressed in terms of impairment, disability or problems in participation (as in the ICF), these must be demonstrated as reliable and valid (6, 7).

From this point of view it is essential to choose adequate measurement instruments, based on evidence of their methodological quality (6, 8-13).

This article focuses on the clinimetric properties of instruments to assess disabilities in gait and related activities (running, moving around on different types of surfaces, climbing stairs and handling special aids for transportation) in patients with rheumatic disorders. The reason for focusing particularly on disabilities in gait and related activities is that this represents an essential part of health-related quality of life for these patients in the sense of their being able to manage activities independently. Therefore, improvement of the independence of the patient and the quality of gait and related activities is frequently included as a treatment goal for patients with rheumatic disorders.

The clinical variables that are examined need to be relevant for the treatment goals and need to be coherent to the prognosis of the disease (8-10, 14-21). From this point of view it is essential to use the correct and most optimal measurement instrument(s) for the

assessment of aspects of gait and related activities. We recently demonstrated that 37% of all 209 available instruments to assess people with rheumatic disorders focus on disabilities, partly on disabilities in gait and related activities. A majority of 57% of all available instruments to assess people with rheumatic disorders focus on impairments (4).

In the ideal situation the choice for use of a measurement instrument should be based on the methodological quality of the available instruments (6, 8-13). However, an earlier study has demonstrated a lack of information on the reliability and validity of the instruments in use for the assessment of musculoskeletal disorders (11).

Reliability is a basic characteristic of an instrument that is required to be adequate: if an instrument is not reliable it will not be useful. If the reliability is sufficient, it is also relevant to assess the validity of the instrument. In the literature, several aspects of validity are described, but because of the frequent lack of a gold standard, the construct validity is the most commonly assessed aspect. An instrument can be validated by studying its correlation with a similar (optimally comparable) construct (for instance the same disability) or by correlation with a less similar construct, such as age, gender, etc. In 73% of validity studies concerning construct validity, the validity is based on its correlation with constructs that measure a totally different entity, and in 11% the validity values are based on its correlation with a similar construct (22).

In this study we distinguished various levels of construct validity, because we expected to find stronger correlations if measurement instruments were validated against a similar construct than if validated against a less similar construct (22). If this were found to be so, it could be relevant for the interpretation of validity studies.

The aim of this study was three-fold. The first objective was to make an inventory of available instruments for the assessment of disabilities in gait and related activities in patients with rheumatic disorders. Our second aim was to investigate which of these instruments have acceptable methodological quality with regard to reliability and validity. Our third aim was to investigate the assumption that the evaluation of convergent construct validity results in stronger correlations when validated against a similar construct.

Methods and materials

We performed a systematic review of the literature on the clinimetric properties of measurement instruments for all disabilities in gait and related activities that are relevant in allied health care for patients with rheumatic disorders.

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 clinimetric properties of instruments to assess disabilities in gait and mobility.
- Different versions of an instrument were considered as separate measurement instruments.
- The questionnaire should contain at least one question or a separate subscale regarding gait or gait related activities (running, moving around on different types of surfaces, climbing stairs and handling special aids for transportation).
- Only instruments for the measurement of adult patients were included.

Literature search

First the Medline database was searched for the period January 1982 – April 2001, using search terms for the relevant rheumatic disorders and various search terms for clinimetric properties (the detailed search strategy is available on request from the corresponding author). The database of the Centre for Documentation of the Dutch Institute of Allied Health Care was also searched for the period January 1988 – April 2001, using the same keywords. Furthermore, the search in both databases was repeated with the names of the identified measurement instruments. The 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 also in some publications from before 1982.

Data extraction

All identified publications were assessed independently on the basis of their title and abstract by two reviewers (RS and YK). In cases of disagreement (3%) the article was also assessed by a third reviewer. The assessment was based on a standardized data collection form (23) consisting of four sections: general description (name, first author, etc.), assessment domain (according to the ICIDH-classification), methodological aspects (concerning reliability, validity, responsiveness) and aspects of feasibility. An explanation of all the abbreviations used in this article is given in the Appendix.

Methodological criteria for clinimetric properties

We investigated the following clinimetric properties: intra-rater reliability, construct validity and responsiveness. To interpret the data on reliability and validity we used criteria based on De Jong *et al.* (24), Eliasziw *et al.* (25) and Doeglas *et al.* (26) (Table I).

To investigate the influence of validat-

Table I. Cut-off points used for intra-rater reliability, inter-rater reliability and construct validity.

	Intra-rater reliability	Validation against similar construct	Validation against dissimilar construct
Good	$x \geq 0.85$	$x \geq 0.65$	$x \geq 0.50$
Moderate	$0.65 < x < 0.85$	$0.50 < x < 0.65$	$0.40 < x < 0.50$
Poor	$x < 0.65$	$x < 0.50$	$x < 0.40$

x = Pearson's r, Spearman's rho or Intra-Class Correlation Coefficient (ICC).

Table II. Categorization of comparators utilized in assessing construct validity.

Construct level	Definition
Construct 1	Validation against instruments that measure the same disability
Construct 2	Validation against instruments that measure the same disability as well as other disabilities
Construct 3	Validation against instruments that measure other disabilities than the instrument to be validated
Construct 4	Validation against generic instruments that measure disabilities as well as impairments and participation problems

ing against different constructs (varying from optimally comparable to dissimilar) the construct validity was divided into four 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 II).

'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 [for example: the HAQ-Mob (variable mobility) that is validated against the AIMS-Mob]. Validity is defined as 'construct 2' if the instrument is validated against instruments that measure the same construct, as well as other disabilities: for example, if walking time is validated against another instrument that measures disabilities in walking, but also disabilities in personal care. 'Construct 3' indicates that the construct relates to other disabilities than the disability to be validated: for example, if walking time is validated against disabilities in interpersonal relations. 'Construct 4' 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 [for example: a questionnaire for disabilities in walking (variable mobility) that is validated against an instrument that is intended for the assessment of muscle force].

We clustered 'construct 1' and 'construct 2'as 'validation against a similar construct' and 'construct 3', 'construct 4' as 'validation against a dissimilar construct'. This distinction is also reflected in the criteria for validity in the two last columns of Table I. The argument for this distinction is the fact that 'similar construct validity' comes closest to the gold standard, and is therefore expected to result in stronger correlations than the dissimilar construct validity'(22).

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 disabilities, based on the ICF-classification (27).

The values of different studies of the same instrument were pooled for each clinimetric property that was assessed. When the relevant information was available, statistical pooling of the data was performed if the measurement instrument was validated against a similar construct. The values were pooled per construct. The pooled index is composed of measurements, appropriately weighted: $\Sigma X = (n_1 x_1) / N$, where X = the pooled index, n_1 = the number of persons included in the study, x_1 = the value of methodological aspect (Pearson's r, Spearman's rho or intra-class correlation coefficient (ICC)) in the study, and N = the total number of persons in all studies included in the pooling. The pooled index was computed for the Pearson's/Spearman's correlation coefficients and the ICC values separately. Values for the convergent construct validity of multi-dimensional instruments can be strongly influenced by values of one or more sub-scales. Therefore, whenever possible, the data were also pooled for the separate subscales.

Results

For the assessment of disabilities in gait and related activities, 78 instruments with a total of 36 subscales were found. Two adapted versions of AIMS, a multi-dimensional questionnaire, were included in this analysis: AIMS2 and AIMSS (for all abbreviations see the Appendix). Furthermore, beside the HAQ also a modified version of the HAQ (MHAQ) is included. AIMS2-WaBe, AIMS2D-WaBe, FAS-Mob, HAQ-Walk, MHAQ-Walk, SW and the WT are intended only for the assessment of walking or related activities (in this case climbing stairs); all other instruments or subscales are multi-dimensional and also measure other disabilities and/or impairments. For 16 of these 78 instruments there are no data at all available regarding the methodological quality of these instruments (neither about reliability nor concerning validity based on the described constructs). The results of the remaining 61 instruments are presented in Table III.

Intra-observer reliability was investigated for 28 instruments or subscales. Sixteen out of these 28 met the criterion of r/ICC ≥ 0.85, and 37 were found to have good validity. For 32 instruments and/or subscales construct validity was investigated in studies in which they were validated against the most similar constructs (construct 1 or construct 2). Seventeen out of these 32 met the criterion of r/ICC ≥ 0.65. For 35 instruments and/or subscales construct validity was investigated in studies in which they were validated against a dissimilar construct (construct 3 or construct 4).

If both reliability and validity are required to be 'good', 7 instruments meet the criteria (shown in grey in Table III). In this study we did not report on the responsiveness of the identified instruments. So far, research into the responsiveness of measurement instruments is hampered by the lack of consensus regarding the preferred method. Of the seven instruments that were found to be valid and reliable in this study, two (the DFI, and the RAQoL) are responsive according to the conclusions of the investigators (57,87,101). For HAQ, SIP and WT there are conflicting results regarding responsiveness, while the responsiveness of ASES-FSE and

Table III. Reliability and validity of instruments for assessment of disabilities in gait or gait related activities in patients with rheumatic disorders.

Measurement instrument / subscale@	Intra-rater reliability	Construct validity if validated against instruments measuring: The same disability	The same as well as other disabilities	Other disabilities	Impairments, disabilities and participation problems or general aspects
AIMS[28-49]	0.86*				0.48*
AIMS-Disab[42]				0.88	
AIMS-GlobH[38]			0.79*		
AIMS-Mob[38]		0.74	0.74*		
AIMS-PhysA[29,39,44]	0.62	0.49	0.45	0.62	0.55
AIMS-PhysH[48]					0.76*
AIMS-PhysM[37]		0.79			
AIMSD[47]				0.43	0.65
AIMSD-PhysA[50,51]			0.71*		0.54
AIMSS-PhysA[44]	0.58		0.40	0.53	
AIMS2-WaBe[36]	0.92#				
AIMS2D-PhysA[52]					0.86
AIMS2D-WaBe[62]					0.67
ASES[53]	0.78*				-0.66
ASES-FSE[53,54]	0.85				0.61
BIM[55]					0.43
DFI[56,96]	0.94				0.60
E-QoL[57,58,97,102]	0.74#*		0.49*		-0.78
FAS[59]	0.85#*				
FAS-Aid[59]	0.60#*				
FAS-Trans[59]	0.68#*				
FIQ[60,98]		0.47			
FIQ-PhysF[61]	0.95				
FSAI[62]	0.82*				
FSI-GlobH[38]			0.66*		0.84
FSI-Mob[38,63]	0.85#*		0.61*		
GARS[64,65]		0.80			0.83*
GARS-ADL[64,65]		0.80			0.80
HAQ[42,58,64,66,97]	0.93#* 0.94*		0.80	0.82*	0.67*
HAQ-Disab[42]				0.88	
HAQ-GlobH[38]			0.71*		0.84
HAQ-Mob[38]		0.58	0.54*		
HAQ-ODI[67]		0.34			
HAQ-OthAc[67]		0.19			
HAQ-Walk[67]		0.23			
IRGL-Mob[38,68,69]		0.66*		0.12*	-0.71
IWB-GlobH[38]			0.56*		0.62
IWB-Mob[38]		0.50	0.70		
KFT[70-73]	0.96			0.71*	0.72
LDQ[74]	0.94#*				
MDR[75]	0.93#				
MFAQ[76,77,101]	0.89 0.92#				
MHAQ[78]	0.91				
MHAQ-ODI[67]		0.34			
MHAQ-OthAc[67]		0.33			
MHAQ-Walk[67]		0.25			
MPPDQ[79]			0.66*		0.40*
MUMQ[80-82]	0.73#*		0.45*	0.26*	0.46*
NHP-PhysM[83,84,105]				0.79	0.82
OSRA[85]					0.52
RAQoL[86]	0.90	0.87		0.62	
SF-36 [76,79,95,97,102,103,104]	0.78#*				
SF-36-PhysA[79]	0.77#		0.73*	0.56*	0.47*
SIP[48,87-89,99,100]	0.92* 0.97#			0.55	0.71*
SIP-GlobH[38]			0.70*		0.87
SIP-Mob[38]		0.60	0.72*		
SIP-PhysH[84]					0.76*
SRQ[90]		0.44		0.58	
SW[91]	0.81#*				
WOMAC-Disab[79,97,102,103,104,105]	0.78#		0.73*	0.60*	0.54*
WT[91,93,94,98,100]	0.89#* 0.91				0.66

@ For abbreviations see Appendix 1; # Intra-class correlation coefficient; * pooled value.In gray: instruments that meet the criteria for reliability as well as for validity. All values expressed as Pearson's-r or Spearman's-r.

KFT have not been investigated up to now.

Data on the validity of 49 instruments and/or sub-scales are available. For 18 out of these 49 instruments, there are data on the optimally comparable construct validity as well as the imperfect construct validity. In 7 of those 18 instruments the correlation for the similar constructs proved to be stronger than the values for validation against the dissimilar constructs. For one instrument the validity was the same for the optimally comparable constructs as for validation against the imperfect constructs. In 10 out of 18 instruments the correlation values for the optimally comparable constructs proved to be lower than the values for validation against the imperfect constructs.

Discussion

For the assessment of disabilities in gait or gait related activities in patients with rheumatic disorders, 78 instruments have been identified. Sixteen out of these 78 were found to have good reliability, and 37 were found to have good validity. Only 7 out of the 78 measurement instruments had good reliability as well as good validity.

The chosen methodological criteria in this study were possibly too severe. Though several authors in the literature advocate a correlation of 0.4 for (construct) validity as 'reasonably high' (106–108) we only accepted a value of 0.50–0.65 as 'moderately strong' and a value ≥ 0.65 as 'strong' for construct validity. The reason to deviate from the level of ≥ 0.40 is the fact that it might be expected that a more optimal comparable construct validity generates higher values than validation against an imperfect construct (22). However, our study did not confirm this hypothesis. This could possibly be explained by the fact that the large majority of the instruments for assessment of walking and related activities are multi-dimensional.

As demonstrated by Table III there are several other instruments that could be considered useful if the methodological criteria were moderated slightly. However, the combination of acceptable reliability as well as validity is sometimes lacking. For example, in the literature it has been demonstrated that the performance of the SF-36 in terms of its psychometric and clinical validity, may be affected by the clinical context of the patient group in which it is applied, and that satisfactory performance cannot be guaranteed for all the groups of conditions. The same point is made for the E-QoL, which has been validated in population surveys and appears to perform well there; however, its value in rheumatic disease populations remains uncertain (Table III). The problems of this scale (distribution of scores, disconcordance with VAS at lower levels of the E-QoL, and compression of item range) suggest a limited role for this instrument in rheumatologic conditions at this time.

In this study, several levels of construct validity were distinguished, based on the hypothesis that validation against an optimally comparable construct should result in stronger correlations than validation against a dissimilar construct. However, the result of this study do not confirm this hypothesis: for the validity of 10 out of 18 instruments, the values for validation against an optimally comparable construct proved to be lower than in case of validation against an imperfect construct. This is opposite to the results of a comparable study on the validity of measurement instruments for the assessment of impairments in patients with rheumatic disorders (109). The reason for this discrepancy is unclear. A possible explanation could be the fact that the large majority of instruments for the assessment of walking and related activities are multi-dimensional and for the most part focus on disabilities other than walking, whereas on the other hand instruments for the assessment of impairment more frequently only focus on the impairment to be measured, and not on other domains.

The group of what is defined as 'rheumatic disorders' (see criteria in inclusion procedure) includes a variety of diseases. We also focused on differences for subgroups in the rheumatic disorders, because some measurement instruments focus in particular on specific characteristics of those patient populations and are developed just for only those patients. For example, there have been several instruments developed for ankylosing spondylitis that are not relevant for patients with rheumatoid arthritis. The majority of the mentioned instruments are multi-dimensional. At this time, for the assessment of intervention in osteoarthritis of the lower extremities, the WOMAC is generally recommended as the most sensitive condition-specific instrument. However, the responsiveness varies by both subscale and intervention. Generally, the physical function subscale has large effect sizes, particularly in hip and knee arthroplasty patients.

Of the great number of available instruments for the assessment of walking and related activities, 'WT' and 'ST' are the only instruments that measure only walking or stair walking respectively. In all other cases it concerns a subscale of a more extensive and mostly multi-dimensional instrument that measures various disabilities, not only walking. Even in case of WT there are different interpretations for patients with rheumatoid arthritis than for patients with osteoarthritis, total hip replacement etc.

The fact that the majority of the instruments are questionnaires in which walking and related activities are only a small part is the consequence of our inclusion procedure. In other parts of the study we only included: 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. This requirement was not possible here, because applying this rule should result in only two instruments (WT and SW). In fact both instruments have moderate or good reliability, but there is lack of information concerning their validity.

Combining reliability with validity leaves 7 instruments that meet the criteria. However, this does not mean that all the others are neither reliable nor valid, because reliability was calculated only for 28 instruments. In other words, for 64% of the available instruments the reliability is unknown. Ano-

ther restriction concerning the low number of data relating to reliability is the fact that this review was restricted to studies focusing on populations with rheumatic disorders. The reliability and validity of the instruments might just as well have been investigated in other patient populations.

Based on the available information the RAQoL and HAQ seem to be the best available instruments; however, both are multi-dimensional and the influence of certain subscales is not clear. For example, it is worth noting that the subscale HAQ-Walk has a very low validity score (r = 0.23, construct 1). Both instruments could possibly result in a lot of unnecessary data for clinical use if the user is only interested in walking and related activities. For that reason the usefulness of RAQoL and HAQ is under discussion if the intention is to measure only gait and related activities. More attractive for clinical use could be the WT and the SW; however, there is a lack of data concerning the validity of those two.

In conclusion, for clinical applications the user needs to balance the importance of the available reliability, validity and amount of ballast of some instruments. Table III is intended as a guide in this process of balancing.

Acknowledment

We wish to thank Mrs. Y. Kappe (YK) for reviewing the majority of the articles, together with the first author.

References

1. RIMMER JH: Health promotion for people with disabilities: the emerging paradigm shift from disability prevention to prevention of secondary conditions. *Phys Ther* 1999; 5: 495-502.
2. WORLD HEALTH ORGANIZATION: *International Classification of Impairments, Disabilities and Handicaps: A Manual of Classification Relating to the Consequences of Disease.* Geneva, World Health Organization, 1980.
3. WORLD HEALTH ORGANIZATION: *International Classification of Functioning, Disabilities and Health Problems.* Geneva, World Health Organization, 2001.
4. SWINKELS RAHM: The ICF-classification as a system for structuring outcome measurement. Submitted, 2003.
5. STREINER WA, RYSER L, HUBER E, UEBELHART D, AESCHLIMANN A, STUCKIG: Use of the ICF model as a clinical problem solving tool in physical therapy and rehabilitation medicine. *Phys Ther* 2002; 82: 1098-107.
6. FEINSTEIN AR: *Clinimetrics*. New Haven, Yale University Press, 1987.
7. HENDRIKS HJM, BRANDSMA JW, HEERKENS YF, OOSTENDORP RAB, NELSON RM: Intraobserver and interobserver reliability of assessments of impairments and disabilities. *Phys Ther* 1997; 77: 1097-106.
8. KIRSCHNER B, GUYATT G: A methodological framework for assessing health indices. *J Chron Dis* 1985; 38: 27-36.
9. BOMBARDIER C, TUGWELL P: Methodological considerations in functional assessment. *J Rheumatol* 1987; 14 (Suppl. 15): 6-10.
10. BOMBARDIER C, TUGWELL P: A methodological framework to develop and select indices for clinical trials: statistical and judgment approaches. *J Rheumatol* 1982; 9: 753-7.
11. BECKERMAN H, BOUTER LM: Efficacy of physiotherapy for musculoskeletal disorders: what can we learn from research? *Br J Gen Pract* 1993; 43: 483-91.
12. MCDOWELL I, NEWELL C: *Measuring Health: A Guide to Rating Scales and Questionnaires.* New York, Oxford University Press, 1987.
13. STREINER DC, NORMAN GR: *Health Measurement Scales. A Practical Guide to their Development and Use*. Oxford, Oxford Medical Publications, 1989.
14. BOUTER LM, LINDEN S VAN DER, KOES BW: Effectmeting in de fysiotherapie. Een kritische beschouwing van de methodologie met speciale aandacht voor het kiezen van effectparameters. *Ned Tijdschr Fysiother* 1991; 101: 46-8.
15. TURNER RR: Rehabilitation. *In* SPILKER B (Ed.): *Quality of Life Assessments in Clinical Trials.* New York, Raven Press, 1990: 247-66.
16. SPECTOR WD: Functional disability scales. *In* SPILKER B (Ed.): *Quality of Life Assessments in Clinical Trials.* New York, Raven Press, 1990: 115-29.
17. MIETTINEN OS, ELLISON RC, PECKHAM GJ: Overall prognosis as the primary criterion of outcome in a clinical trial. *Control Clin Trials* 1983; 4: 227-37.
18. GUYATT GH, DEYO RA, CHARLSON M: Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989; 42: 403-8.
19. DEYO RA, INUI TS: Towards clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Services Research* 1984; 19: 275-89.
20. DEYO RA, CENTOR RM: Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis* 1986; 39: 897-906.
21. GUYATT G, WAITER S, NORMAN G: Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chron Dis* 1987; 40: 171-8.
22. SWINKELS RAHM, VET HCW DE, OOSTENDORP RAB, BOUTER LM: Construct validity of impairment measures in rheumatic disorders: Which constructs are selected for validation? Submitted, 2003.
23. OOSTENDORP RAB, ELVERS JWH, SWINKELS RAHM: Judgement list for reading articles (in Dutch). Dutch National Institute of Allied Health Professions, Amersfoort, The Netherlands, 1997.
24. JONG A DE, MOLENAAR IW: An application of Mokken's model for stochastic, cumulative scaling in psychiatric research. *J Psychiatr Res* 1987; 21: 137-49.
25. ELIASZIW M, YOUNG SL, WOODBURY MG: Statistical methodology for the concurrent assessment of inter-rater and intra-rater reliability: using goniometric measurements as example. *Phys Ther* 1994; 74: 777-88.
26. DOEGLAS D, KROL B, GUILLEMIN F *et al.*: 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* 1995; 22: 1834-43.
27. WORLD HEALTH ORGANIZATION: *ICIDH-2, International Classification of Functioning and Disability*. (Beta-2 draft, short version) Geneva, World Health Organization, 1999.
28. MEENAN RF, ANDERSON JJ, KAZIS LW *et al.*: Outcome assessment in clinical trials: evidence for the sensitivity of a health status measure. *Arthritis Rheum* 1984; 27: 1344-52.
29. MEENAN RF, GERTMAN PM, MASON JH: Measuring health status in arthritis. *Arthritis Rheum* 1980; 23: 146-52.
30. KAZIS LE, ANDERSON JJ, MEENAN RF: Effect sizes for interpreting changes in health status. *Med Care* 1989; 27: S178-S189.
31. 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.
32. HIDDING A, SANTEN M VAN, KLERK E DE *et al.*: Discordance between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. Chapter 9 *In* HIDDING A (Ed.): *Group Physical Therapy in Ankylosing Spondylitis*, Chapter 9, Meppel, Krips repro. 1993: 94-104.
33. BUCHBINDER R, BOMBARDIER C, YEUNG M, TUGWELL P: Which outcome measurements 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.
34. ANDERSON JJ, CHERNOFF MC: Sensitivity to change of rheumatoid arthritis clinical outcome measures. *J Rheumatol* 1993; 20: 535-7.
35. POTTS MK, BRANDT KD: Evidence of the validity of the arthritis impact measurement scales. *Arthritis Rheum* 1987; 30: 93-6.
36. 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.
37. FITZPATRICK R, ZIEBLAND S, JENKINSON C, MOWAT A: A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol* 1993; 20: 429-36.
38. 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.

39. 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.
40. KEEFE FJ, CALDWELL DS, QUEEN KT *et al.*: Pain coping strategies in osteoarthritic patients. *J Consult Clin Psychol* 1987; 55: 208-12.
41. LIANG MH, FOSSELAH, LARSON MG: Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990; 28: 632-42.
42. 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.
43. JACOBS JWG, OOSTERVELD FGJ, 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.
44. 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.
45. DUFFY CM, WATANABE KN, GLADMAN DD *et al.*: The utility of the Arthritis Impact Measurement Scales for patients with psoriatric arthritis. *J Rheumatol* 1992; 19: 1727-32.
46. 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.
47. VLIET-VLIELAND TPM, WIJK TP VAN DER, ZWINDERMAN JAH, HAZES JMW: Determinants of hand function in patients with rheumatoid arthritis. *J Rheumatol* 1996; 23: 835-40.
48. WEINBERGER M, SASMA GP, TIERNEY WM, BELEYA MJ, HINER SL: Generic versus disease specific health status measures: Comparing the Sickness Impact Measurement Scales. *J Rheumatol* 1985; 19: 543-6.
49. HUSTED JA, 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.
50. TAAL E, JACOBS JW, SEYDAL ER, WIEGMAN O, RASKER JJ: Evaluation of the Dutch Arthritis Impact Measurement Scales (Dutch-AIMS) in patients with rheumatic arthritis. *Br J Rheumatol* 1989; 28: 487-91.
51. TAALE, SEYDALER, JACOBS JW, WIEGMAN O, RASKER JJ: De Nederlandse Arthritis Impact Measurement Scales (Dutch-AIMS) – bepaling van fysieke en psychosociale gezondheidsaspecten van reumatoïde artritis. *Gedrag en Gezondheid* 1989; 17: 69-74.
52. RIEMSMA RP, TAAL E, RASKER JJ, HOUTMAN PM, PAASSEN HC VAN, WIEGMAN O: Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. *Br J Rheumatol* 1996; 35: 755-60.
53. SHOOR SM, HOLMANHR: Development of an instrument to explore psychological mediators of outcome in chronic arthritis. *Trans Assoc Am Physicians* 1985; 97: 325-31.
54. LORIG K, CHASTAIN RL, UNG E, SCHOOR S: Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum* 1989; 32: 37-44.
55. BAKHEIT A, HARRIES S, HULL R: Validity of a self-administered version of the Barthel index in patients with rheumatoid arthritis. *Clin Rehab* 1995; 9: 234-7.
56. CREEMERS MCW, HOF MA VAN T, FRANSSEN MJAM: A Dutch version of the functional index for ankylosing spondylitis: Development and validation in a long-term study. *Br J Rheumatol* 1994; 33: 842-6.
57. HURST NP, JOBANPUTRA P, HUNTER M: Validity of Euroqol: A generic health status instrument in patients with rheumatoid arthritis. *Br J Rheumatol* 1994; 33: 655-62.
58. HURST NP, KIND P, RUTA D: Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of Euroqol (EQ-5D). *Br Soc Rheumatol* 1997; 36: 551-9.
59. MARKS R: Reliability and validity of self-paced walking time measures for knee osteoarthritis. *Arthritis Care Res* 1994; 7: 51-3.
60. LEE P, MUKUNDRAI K, JASANI W, CARSON D: Evaluation of a functional index in rheumatoid arthritis. *Scand J Rheumatol* 1973; 2: 71-7.
61. BURCKHARDT CS, CLARK SR, BENNETT RM: The fibromyalgia impact questionnaire: Development and validation. *J Rheumatol* 1991; 18: 728-33.
62. JETTE AM, DENISTON OL: Interobserver reliability of a functional status assessment instrument. *J Chron Dis* 1978; 31: 573-80.
63. JETTE AM: The functional status index: Reliability and validity of a self report functional disability measure. *J Rheumatol* 1987; 14: S15-S19.
64. DOEGLAS D, KROL B, GUILLEMIN F: 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* 1995; 22: 1834-43.
65. CALLAHAN LF, MCCOY A, SMITH W: Comparison and sensitivity to change of self-report scales to assess difficulty, dissatisfaction, and pain in performing activities of daily living over one and five years in rheumatoid arthritis. *Arthritis Care Res* 1992; 5: 137-45.
66. ENDE CHM VAN DEN, BREEDVELD FC, DIJKMANS BAC: The limited value of the Health Assessment Questionnaire as an outcome measure in short term exercise trials. *J Rheumatol* 1997; 24: 1972-7.
67. BLALOCK SJ, SAUTER SVH, DEVELLIS RF: The modified Health Assessment Questionnaire difficulty scale: A health status measure revisited. *Arthritis Care Res* 1990; 3: 182-8.
68. BORSTLAP M, LAAR M VAN DER, ZANT J: Components of health: An analysis in rheumatoid arthritis using quality of life questionnaires and clinical and laboratory variables. *Ann Rheum Dis* 1993; 52: 650-5.
69. BIJLSMA JWJ, HUISKES CJAE, KRAAIMAAT FW: Relation between patients' own health assessment with clinical and laboratory findings in rheumatoid arthritis. *J Rheumatol* 1991; 18: 650-3.
70. HAKALA M, NIEMINEN P: Functional status assessment of physical impairment in a community based population with rheumatoid arthritis: Severely incapacitated patients are rare. *J Rheumatol* 1996; 23: 617-23.
71. KALLA AA, KOTZE TJ, MEYERS OL: Clinical assessment of disease activity in rheumatoid arthritis; evaluation of a functional test. *Ann Rheum Dis* 1988; 47: 773-9.
72. EBERL DR, FASCHING V, RAHLFS V: Repeatability and objectivity of various measurements in rheumatoid arthritis : a comparative study. Arthritis Rheum 1976; 19: 1278-86.
73. KALLA AA, SMITH PR, BROWN GMM: Responsiveness of Keitel Functional Index compared with laboratory measures of disease activity in rheumatoid arthritis. *Br J Rheumatol* 1995; 34: 141-9.
74. ABBOTT CA, HELLIWELL S, CHAMBERLAIN MA: Functional assessment in ankylosing spondylitis: evaluation of a new self-administered questionnaire and correlation with anthropometric variables. *Br J Rheumatol* 1994; 33: 1060-6.
75. DURHAM J, MACNEISH A, ROONEY PJ: The MDR index of function in rheumatoid arthritis. *Clin Exp Rheumatol* 1985; 3: 297-302.
76. MARTINDP: Comparison of the Musculoskeletal Function Assessment Questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile Health-status measures. *J Bone Joint Surg* 1997; 79-A: 1323-35.
77. MARTIN DP, ENGELBERG R, AGEL J: Development of a musculoskeletal extremity health status instrument: The musculoskeletal function assessment instrument. *J Orthop Res* 1996; 14: 173-81.
78. PINCUS T, SUMMAY JA, SORACI SA: Assessment of patient satisfaction in activities of daily living using a modified Stanford health assessment questionnaire. *Arthritis Rheum* 1983; 26: 1346-53.
79. WRIGHT JG, YOUNG NL: A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997; 50: 239-46.
80. BAKKER C, RUTTEN M, DOORSLAER E VAN, BENNET K, LINDEN S VAN DER: Feasibility of utility assessment by rating scale and standard gamble in patients with ankylosing spondylitis or fibromyalgia. *J Rheumatol* 1994; 21: 269-74.
81. BAKKER C: Patient utilities in fibromyalgia and the association with other outcome measures. Maastricht, University Press, 1995.
82. BAKKER C, RUTTEN M, HIDDING A, DOORSLAER E VAN, BENNETT K, LINDEN S VAN DER: Patient utilities in ankylosing spondylitis and the association with other outcome measures. *J Rheumatol* 1994; 21: 1298-304.
83. FITZPATRICK R, ZIEBLAND S, JENKINSON C, MOWATA, MOWATA: A general health status instrument in the assessment of rheumatoid arthritis. *Br J Rheumatol* 1992; 31: 87-90.
84. HOUSSIEN DA, MCHENNA SP, SCOTT DL: The Nottingham Health Profile as a measure of disease activity and outcome in rheumatoid arthritis. *Br J Rheumatol* 1997; 36: 69-73.
85. KATZ PP, ALFIERI WS: Satisfaction with abilities and well-being: Development and validation of a questionnaire for use among persons with rheumatoid arthritis. *Arthritis Care*

Res 1997; 10: 89-98.
86. JONG Z DE, VAN DER HEIJDED, MCKENNASP: The reliability and construct validity of the RAQoL: A rheumatoid arthritis-specific quality of life instrument. *Br J Rheumatol* 1997; 36: 878-83.
87. DEYO RA, INUI TS, LEININGER JD: Measuring functional outcomes in chronic disease: A comparison of traditional scales and a self-administered health questionnaire in patients with rheumatoid arthritis. *Med Care* 1983; 21: 180-92.
88. BERGNER M, BOBBITT RA, CARTER WB: The Sickness Impact Profile: Development en final revision of a health status measure. *Med Care* 1981; 19: 787-805.
89. PAPAGEORGIOU AC, BADLEY EM: 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* 1989; 16: 106-12.
90. PINCUS T, CALLAHAN LF, BROOKS RH: Self-report questionnaire scores in rheumatoid arthritis compared with traditional physical, radiographic and laboratory measures. *Ann Intern Med* 1989; 110: 259-66.
91. MARKS R: Walking time measures for evaluating OAof the knee: reliability and sensitivity of self-paced walking time for assessing functional capacity in persons with osteoarthritis of the knee. *S Afr J Physiother* 1994; 50: 5-8.
92. WEBB J, WILSON DW, CARSON DW: Evaluation of digital joint circumference measurements in rheumatoid arthritis. *Scand J Rheumatol* 1973; 2: 127-31.
93. BORSTLAP M, ZANT JL, SOESBERGEN SM VAN, KORST JK VAN DER: Quality of life assessment: A comparison of four questionnaires for measuring improvement after total hip replacement. *Clin Rheumatol* 1995; 14: 15-20.
94. GRACE EM, CERECZ E, KASSAM Y, BUCHANAN WW: Fifty foot walking time: An inappropriate outcome measure. *Arthritis Rheum* 1986; 29: 516.
95. HUSTED JA, GLADMAN DD, FAREWELLVT, LONG JA, COOK RJ: Validating the SF-36 Health Survey Questionnaire in patients with psoriatic arthritis. *J Rheumatol* 1997; 24: 511-7.
96. HIDDING A, SANTEN M VAN, KLERK E DE *et al.*: Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. *J Rheumatol* 1994; 21: 818-23.
97. BRAZIER JE, HARPER R, MUNRO J, WALTERS SJ, SNAITH ML: Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Br Soc Rheumatol* 1999; 38: 870-7.
98. KING S, WESSEL J, BHAMBHANI Y, MAIKALA R, SHOLTER D, MAKSYMOWYCH W: Validity and reliabilityof the 6 minute walk in persons with fibromyalgia. *J Rheumatol* 1999; 26: 2233-7.
99. KOSINSKI M, KELLER SD, HATOUM HT, KONG SX, WARE JE: The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis. *Med Care* 1999; 37: MS10-MS22.
100. KOSINSKI M, KELLER SD, WARE JE, HATOUM HT, KONG SX: The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis. *Med Care* 1999; 37: MS23-MS39.
101. SWIONTKOWSKI MF, ENGELBERG R, MARTIN DP, AGEL J: Short musculoskeletal function assessment questionnaire: Validity, reliability and responsiveness. *J Bone Joint Surg* 1999; 81-A: 1245-60.
102. FRANSEN M, EDMONDS J: Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Br Soc Rheumatol* 1999; 38: 807-13.
103. DOMSIC RT, SALTZMAN CL: Ankle Osteoarthritis Scale. *Foot & Ankle* 1998; 19: 466-71.
104. ROOS EM, KLÄSSBO M, LOHMANDEOR LS: WOMAC Osteoarthritis Index. *Scand J Rheumatol* 1999; 28: 210-5.
105. SÖDERMAN P, MALCHAU H: Validity and reliability of Swedish WOMAC Osteoarthritis Index. *Acta Orthop Scand* 2000; 71: 39-46.
106. GHISELLEEE: *The Measurement of Occupational Aptitude*. Berkeley, Los Angeles, 1966.
107. HUISKES CJAE, KRAAIMAAT FW, BIJLSMA JWJ: Development of a self-report questionnaire to assess the impact of rheumatic diseases on health and lifestyle. *J Rehab Sciences* 1990; 3: 65-70.
108. STEWART AL, WARE JE: *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach.* Durham, North Carolina, Duke University Press, 1992.
109. SWINKELS RAHM, OOSTENDORP RAB, BOUTER LM, ENDE CHM VAN DEN: Measuring impairments in rheumatic disorders: What are the best instruments for allied health care? Submitted, 2003.

Appendix. List of abbreviations of measurement instruments and sub-scales.

Abbreviation	Name of Measurement Instrument	Abbreviations of Sub-Scales	
AIMS	Arthritis Impact Measurement Scale	Disab GlobH Mob PhysA PhysH PhysM	Disabilities Global Health Mobility Physical Activities Physical Health Physical Mobility
AIMSD	Arthritis Impact Measurement Scale – Dutch	PhysA	Physical Activities
AIMS2	Arthritis Impact Measurement Scale 2	WaBe	Walking and Bending
AIMS2D	Arthritis Impact Measurement Scale2 Dutch	PhysA WaBe	Physical Activities Walking and Bending
AIMSS	AIMS short version	PhysA	Physical Activities
ASES	Arthritis Self Efficacy Scale	FSE	Functional Self Efficacy
BIM	Barthel Index Modified		
DFI	Functional Index (Dutch)		
E-QoL	Euro Quality of Life Scale		
FAS	Functional Assessment Survey	Aids Trans	Aid-tools Transport
FIQ	Functional Index Questionnaire	PhysF	Physical Function
FSAI	Functional Status Assessment Instrument		
FSI	Functional Status Index	GlobH Mob	Global Health Mobility
GARS	Groninger Activity Restriction Scale	ADL	Activities of Daily Living
HAQ	Health Assessment Questionnaire	Disab GlobH Mob ODI OthAc Walk	Disabilities Global Health Mobility Overall Disability Index Other Activities Walking
IRGL	Invloed van Rheuma op Gezondheid en Leefwijze	Mob	Mobility
IWB	Index of Well Being	GlobH Mob	Global Health Mobility
KFT	Keitel Functional Test		
LDQ	Leeds Disability Questionnaire		
MDR	MDR-Index		
MFAQ	Musculoskeletal Function Assessment Questionnaire		
MHAQ	Modified Health Assessment Questionnaire	ODI OthAc Walk	Overall Disability Index Other Activities Walking
MPPDQ	MACTAR Patient Preference Disability Questionnaire		
MUMQ	Maastricht Utility Measurement Questionnaire		
NHP	Nothingham Health Profile	PhysM	Physical Mobility
OSRA	Overall Status in Rheumatoid Arthritis		
RAQoL	Rheumatoid Arthritis Quality of Life Scale		
SF-36	Short Form 36	PhysA	Physical Activities
SIP	Sickness Impact Profile	GlobH Mob PhysH	Global Health Mobility Physical Health
SRQ	Self-Report Questionnaire		
SW	Stair Walk		
WOMAC	WOMAC-Osteoarthritis Index	Disab	Disabilities
WT	Walking time-50 foot		