

# Reliability, validity and responsiveness of instruments to assess disabilities in personal care in patients with rheumatic disorders. A systematic review

R.A.H.M. Swinkels<sup>1</sup>, P.U. Dijkstra<sup>2</sup>, L.M. Bouter<sup>3</sup>

---

<sup>1</sup>Free University Brussels, Faculty of Medicine and Pharmacology,  
Postgraduate Education Manual Therapy, Brussel, Belgium;

<sup>2</sup>Centre for Rehabilitation, University Hospital Groningen, Groningen;

<sup>3</sup>VU University Medical Centre, Institute for Research in Extramural Medicine,  
Amsterdam, The Netherlands.

---

## Abstract

### Objectives

*The first aim was to make an inventory of available instruments and questionnaires for the assessment of disabilities in personal care in patients with rheumatic disorders. The second aim was to investigate which of these instruments have acceptable methodological quality with regard to reliability, validity and responsiveness. The third aim was to investigate the assumption that convergent validity results in stronger correlations when validated against a more similar construct.*

---

### Methods

*A computer-aided literature search (1982-2001) in 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 were extracted in a standardised way and compared to a priori defined criteria.*

---

### Results

*In total, 19 measurement instruments were included. Five out of these 19 were found to have acceptable reliability, while 12 had acceptable validity. Only three questionnaires met both criteria. Results concerning the responsiveness of these three questionnaires were conflicting. No difference was found in the strength of correlation between validation against the most similar construct versus validation against the least similar construct.*

---

### Conclusion

*It is concluded that the Arthritis Impact Measurement Scale (AIMS) is the most suitable instrument for the assessment of disabilities in personal care.*

---

### Key words

Measurement instrument, personal care, rheumatic disorders, methodological quality, assessment, systematic review.

Raymond A.H.M. Swinkels, MSc; Pieter U. Dijkstra, PhD; Prof. Lex M. Bouter, PhD.

This study was 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 May 10, 2004; accepted in revised form on November 4, 2004.

© Copyright CLINICAL AND EXPERIMENTAL RHEUMATOLOGY 2005.

## Introduction

Personal care is an essential aspect of health-related quality of life for patients with rheumatic disorders. Due to the progressive nature of rheumatic disorders, it is important that patients are able to manage their daily life independently for as long as possible. This literature review focuses on the clinimetric properties of instruments that assess personal care disabilities (washing, dressing, eating and drinking) in patients with rheumatic disorders. To evaluate treatments aimed at reducing personal care disabilities or slowing their progress over time, measurement instruments should be used to objectify this.

In a search for the optimal measurement instrument, knowledge about the domain to be measured is the first requirement. The domain to be measured must be relevant for treatment goals. In addition, the measurement instruments need to be reliable and valid, and for the evaluation of treatment they should also be responsive (1-5). There is no 'gold standard' for the assessment of (construct-) validity of instruments measuring different levels of personal care disability; therefore, construct validity is evaluated (6). However, there is no consensus on how construct validity should be evaluated (7, 8). Such an evaluation can be based on comparison with a very similar construct (optimal construct comparison) or a different construct (imperfect construct comparison), such as age, gender, or an impairment, for instance range of motion, instead of a disability (9). It can be hypothesised that the outcome of construct validity evaluation may be influenced by the similarity or dissimilarity of the construct against which the measurement instrument is validated. Comparison with a similar construct might result in a stronger correlation than comparison with a dissimilar construct. In the latter case the validity of the instrument is under-valued.

Rehabilitation medicine, physiotherapy and occupational therapy focus on the consequences of diseases for a patient, and aim at reducing the burden of these consequences. Hence, in rehabilitation medicine there is a need to describe the

functioning (consequences of disease) of the patient in daily life. The consequences of a disease for a specific patient and the patient's complaints are not only related to the impairments due to the disease (biomedical model), but also to coping, illness perception, etc. (bio-psychosocial model). A classification system which describes these consequences, should include diseases and symptoms, as well as the complaints and the ability of the patient to function. However, the most frequently used system for the classification of diseases, the International Classification of Diseases and Related Health Problems (ICD) (10), is not appropriate for the classification of health problems according to the bio-psychosocial model. The ICD focuses only on impairments and/or diseases, and does not focus on the different domains of the bio-psychosocial model, which are a much broader context. This shortcoming of the ICD led to the development of the International Classification of Impairments, Disabilities and Handicaps (ICIDH) (11), later modified as the International Classification of Functioning, Disability and Health (ICF) (12).

This shift of focus from disease to the consequences of disease (which took place mainly in the late 1980s and the 1990s), implied that the focus of measurement instruments also had to change. Not only impairments should be measured, but also disabilities to perform activities and problems in participation. 'Impairment' is defined as loss or abnormality of psychological, physiological or anatomical structure or function (for instance reduction in range of motion or loss of muscle strength). 'Disability' is defined as any restriction or lack of ability to perform an activity (for instance washing or dressing) in the manner or within the range considered normal for a human being, as a result of an impairment. 'Handicap' is defined as a disadvantage for a given individual that limits the fulfilment of a role that is normal for that person, related to age, gender, and social and cultural factors, as a result of an impairment or disability. However, the majority of instruments were devel-

oped before the above-mentioned shift in focus. As a consequence for patients with rheumatic disorders, 57% of the available measurement instruments (n = 209) focus on impairments in function, approximately 37% measure disabilities to perform activities, and only 5% of all instruments deal with problems in participation (13).

As Stucki and Sigl emphasize, from the rehabilitation perspective, the measurement of functioning and health is relevant not only for evaluating the impact of the disease or the outcome of intervention. It is also used for the diagnosis and interventional management of limitations in functioning and health. Thus, from a rehabilitation perspective, measures of functioning, disability and health are examined much more closely at both the level of individual problems and the level of instrument scales. In rheumatology, both perspectives are important in the management of most patients.

Health status instruments usually include the dimensions of physical functioning, social function, emotional function, pain and the perception of well-being. However, there is unfortunately no clear relationship with the ICF framework (14)

The aim of this literature review was 3-fold. The first aim was to make an inventory of available instruments and questionnaires for the assessment of personal care disabilities in patients with rheumatic disorders. The second aim was to investigate which of these instruments have acceptable methodological quality with regard to reliability, validity and responsiveness. The third aim was to investigate the assumption that convergent construct validity results in stronger correlations when validated against a more similar construct.

## Method and materials

### 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 detailed search strategy is available on request from the corresponding

author). 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, using the same keywords. The names of the measurement instruments that were 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 number of papers written before 1982.

### Inclusion criteria

The following inclusion criteria were applied:

- Papers had to focus on patients suffering from rheumatoid arthritis, seronegative polyarthritis (including psoriatic arthritis), osteoarthritis, ankylosing spondylitis, polymyositis or fibromyalgia.
- Papers had to contain information about the clinimetric properties of instruments to assess personal care disabilities.
- Different versions of an instrument were considered as separate measurement instruments.
- Because many questionnaires focus on more than one domain of the ICDH or ICF classification, or measure more than one variable, papers were included if the instruments focused mainly (50% or more of the items) on the disability to be measured, or if questionnaires had a subscale for personal care disability so that it can be interpreted separately as a single entity.
- Only instruments for the measurement of adult patients were included.

### Data extraction

All papers that were identified were selected independently on the basis of title and abstract by two reviewers (RS and YK). A standardised scoring form (15) was used to assess reliability, validity and responsiveness. The domains according to the ICDH classification were also assessed. In case of disagreements, which occurred in 3% of the assessments, the paper was also assessed by a third reviewer. An explanation of all the abbreviations of instruments and sub-scales mentioned in this article is given in the Appendix.

### Methodological criteria for clinimetric properties

Analysis and interpretation of the intra-observer reliability, construct validity and responsiveness were performed according to the criteria presented in Table I (16-18).

Construct validity was divided into five levels, in which the constructs against which the instruments were validated were ranked according to their degree of similarity to the instrument at issue, a method previously used by Swinkels *et al.* (19) (Table II). Thus the five levels indicate the degree of convergence. 'Level 1 construct' is the most convergent construct, which means that the instrument is validated against an instrument which is very similar to the instrument to be validated. For example, the HAQ-Mob (variable mobility) is validated against the AIMS-Mob (variable mobility). 'Level 2 construct' indicates that the instrument is validated against an instrument that measures the same construct, as well as other types of disabilities. For example, the HAQ-eat is validated against the total score of the AIMS (including more disabilities other than eating). 'Level 3

**Table I.** Cut-off points for intra-rater reliability and construct validity.

|          | Intra-rater reliability | Optimal convergent construct validity | Least convergent construct validity |
|----------|-------------------------|---------------------------------------|-------------------------------------|
| Good     | x 0.85                  | x 0.65                                | x 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.** Levels of construct convergence used to assess construct validity

| Level of convergence of constructs | Definition   |                                   |
|------------------------------------|--|-----------------------------------|
| Level 1 construct                  | Validation against instruments that measure the same disability  | Optimal convergent construct      |
| Level 2 construct                  | Validation against instruments that measure the same disability as well as other disabilities                      | Optimal convergent construct      |
| Level 3 construct                  | Validation against instruments that measure other disabilities than the instrument to be validated                 | Less optimal convergent construct |
| Level 4 construct                  | Validation against instruments that measure impairments instead of the disability at issue                         | Less optimal convergent construct |
| Level 5 construct                  | Validation against generic instruments that measure impairments as well as disabilities and participation problems | Less optimal convergent construct |

'Instrument' can mean a 'measurement instrument' (the SODA or a questionnaire) as well as a 'sub-scale' of a questionnaire (see Appendix)

construct 'indicates that the instrument is validated against an instrument that measures other disabilities than the disability to be validated. For example, the HAQ-eat is validated against the SIP-mobility. 'Level 4 construct' indicates that the instrument is validated against instruments that measure impairments instead of disabilities. For example, the HAQ-mobility is validated against the Larsen score. 'Level 5 construct' is the least convergent validity, indicating that the instrument is validated against generic instruments that measure impairments as well as disabilities and problems in participation, or instruments that measure global aspects of the patients such as gender, age, etc.

In order to be able to interpret the results and to investigate the assumption that construct validity results in stronger correlations when validated against a more similar construct, the levels 1 to 5 construct were dichotomised. Level 1 and level 2 construct were clustered into 'optimal convergent construct validity' and levels 3 through 5 construct were clustered into 'least convergent construct validity' (Table I). This dichotomy was made because an optimal convergent construct validity resembles the gold standard, and is therefore expected to result in stronger correlations than the least convergent construct validity (9, 13).

There is no consensus about the most appropriate method to evaluate responsiveness (20). Therefore, the results are

merely described as reported by the authors.

*Data analysis*

The instruments were classified according to the type of disabilities, based on the ICF classification (12).

When the relevant information was available, statistical pooling of the data was performed if the measurement instrument was validated against the same construct. The values were pooled per construct. A pooled index of the reliability and validity of the instruments was compiled, weighted according to the formula:  $X = (n_1 x_1) / N$  where X = pooled estimate,  $n_1$  = number of persons included in the study,  $x_1$  = correlation (Pearson's r, Spearman's rho or Intra Class Correlation Coefficient (ICC)) reported in the study, N = total number of persons in all studies included in the pooling). The pooled estimate was separately computed for Pearson's r, Spearman's rho and the ICCs. Values for the convergent construct validity of multidimensional instruments can be strongly influenced by values of one or more sub-scales (21-23). Therefore, whenever possible, the data were also pooled for the separate sub-scales.

The data were analysed in the Statistical Package for Social Sciences (SPSS) 8.0.

**Results**

In total, 19 different measurement instruments or sub-scales for the assess-

ment of personal care disabilities were identified: five integral instruments (AIMS, AIMS-D, HAQ, MHAQ and SODA) and 14 sub-scales of multi-dimensional instruments. All integral instruments were, in fact, multi-dimensional questionnaires, in which only some of the questions focused on personal care.

Five out of these 19 instruments had good reliability (AIMS, HAQ, MHAQ, SIP and SODA), and eleven questionnaires or sub-scales were found to have good validity (Table III).

The construct validity of eight instruments and/or sub-scales was investigated in studies in which they were validated against the most optimally convergent constructs (level 1 or 2 construct). Three out of these eight met the criterion of  $r/ICC \geq 0.65$ . The construct validity of 16 instruments and/or sub-scales had been investigated in studies in which they were validated against a less convergent construct (level 3, 4 or 5 construct). Ten of these 16 instruments met the criterion for good validity.

Good reliability as well as good validity was found for three instruments (AIMS, HAQ and SIP, marked grey in Table III). With regard to the responsiveness of the three instruments that were found to be both reliable and valid, the conclusions drawn in the various studies were conflicting.

Data on validity were available for all 19 instruments or sub-scales. For eight instruments there were data on compar-

**Table III.** Results for intra-rater reliability, validity and responsiveness of instruments for the assessment of disabilities in personal care in patients with rheumatic disorders.

| Measurement<br>Instrument / sub-scale@     | Intra-rater<br>reliability | Construct validity if validated against instruments measuring |  |                  |   |  | Respon-<br>siveness |
|--|----------------------------|---|--|------------------|---|--|---------------------|
|  |                            | The same<br>disability  | The same<br>disability as<br>well as other<br>disabilities | Impair-<br>ments | Impairments as<br>well as disabilities<br>and participa-<br>tion problems | General<br>aspects, such<br>as gender,<br>age etc. |                     |
|  |                            | Level 1   | Level 2  | Level 3          | Level 4   | Level 5  |                     |
| <b>AIMS</b> (21;22;29-49)                  | <b>0.86*</b>               |   |  |                  | <b>0.71*</b>  |  | <b>1,2</b>          |
| AIMS-ADL (21;22;29-49)                     | 0.59                       |   | 0.30   | 0.34*            | 0.39*   | 0.29*  |                     |
| AIMS2D-Selfcare (37;48-50)                 |                            |   |  |                  | 0.77  | 0.22*  | 1                   |
| AIMSD (43;46;51)                           |                            | 0.65  |  | 0.43             | 0.65  | -0.56  | 1                   |
| AIMSD-ADL (43;46;51;52)                    |                            |   | 0.69*  |                  | 0.36  | 0.37*  |                     |
| AIMSS-ADL (22)                             | 0.56                       |   | 0.33   | 0.40             | 0.29  | 0.29*  |                     |
| GARS-ADL (18;53;54)                        |                            |   | 0.64   | 0.78*            | 0.63*   |  |                     |
| <b>HAQ</b> (18;21;23;34;38;41;42;45;54-71) | <b>0.94* 0.93#*</b>        |   | <b>0.80</b>  | <b>0.82*</b>     | <b>0.78</b>   | <b>0.34*</b>                                       | <b>1,2</b>          |
| HAQ-hygie (18;21;23;34;38;41;42;45;54-71)  |                            |   |  | 0.10             | 0.42  | 0.15*  | 1,2                 |
| HAQ-eat (18;21;23;34;38;41;42;45;54-71)    |                            |   | 0.43   | 0.65             |   | -0.09*   | 1,2                 |
| HAQ-dresG (18;21;23;34;38;41;42;45;54-71)  |                            |   | 0.21   | 0.73             |   | 0.19*  | 1,2                 |
| IRGL-Selfcare (72-75)                      |                            |   |  | 0.28*            | 0.77  | -0.18*   |                     |
| MHAQ (23;59;60;63;64;67-69)                | 0.91                       |   |  | 0.05             |   | 0.24*  | 1,2                 |
| MHAQ-hygie (49;50;53;54;57-60)             |                            |   |  | 0.49             | 0.42  |  | 1,2                 |
| MHAQ-eat (23;59;60;63;64;67-69)            |                            |   |  |                  | 0.65  | -0.05*   | 1,2                 |
| MHAQ-dresG (23;59;60;63;64;67-69)          |                            |   |  | 0.26             | 0.73  | -0.06*   | 1,2                 |
| <b>SIP-BC&amp;M</b> (38;41;47;76-84)       | <b>0.92*</b>               |   |  |                  |   | <b>0.59*</b>                                       | <b>1,2</b>          |
| SIP-eat (38;41;47;76-84)                   | 0.92*                      |   |  |                  |   | 0.22   | 1,2                 |
| SODA (85)                                  | 0.93                       |   |  |                  |   | 0.26*  |                     |

@ For explanation of abbreviations see Appendix 1. All values expressed in Pearson's-r or Spearman's-rho

# Intra-class Correlation Coefficient; \* Pooled value

Responsiveness 1: studies that indicate that the instrument is responsive; Responsiveness 2: studies that indicate that the instrument is not responsive  
In grey: instruments that meet the criteria for reliability as well as for validity.

isons with an optimal convergent construct and with a less optimal convergent construct. In four of the instruments the strength of the correlations with an optimal convergent construct was similar or higher than the strength of the correlations with a less convergent construct, and in four of the instruments the strength of the correlations with an optimal convergent construct was less than the strength of the correlations with a less convergent construct.

### Discussion

Novel in our systematic review is firstly its focus on reliability,(construct) validity and responsiveness of all avail-

able measurement instruments for one single domain (disabilities in personal care in patients with rheumatic disorders). Secondly, we grouped the data on content validity optimally by classifying all constructs, that are used to validate a measurement instrument, according the ICF-classification. In this way, we divided optimal convergent constructs from less convergent constructs.

Based on the results of this study, it is concluded that the AIMS and the HAQ are the most suitable instruments for the assessment of personal care disabilities in patients with rheumatic disorders. The reliability and validity of both of these questionnaires was good

to very good, based on the results of the evidence retrieved in the literature review. Although the validity of the integral HAQ is very good, analysis of its validity at the level of the sub-scales demonstrates, in general, unacceptably low values (in the majority < 0.50) for HAQ-eat and for HAQ-hygie. The AIMS was responsive, in particular for the self-care sub-scale, whereas for the HAQ more conflicting conclusions were found for responsiveness. Therefore, the AIMS is more appropriate than the HAQ for measuring personal care disabilities, because the AIMS has a stronger validity for almost all of the sub-scales than the HAQ. The reliability of most instruments was good, with

the exception of the ADL sub-scale of the AIMS and the AIMSS (Table III). Since only questionnaires which focus mainly (50% or more of the items) on disabilities in personal care were included, this reduced the total number of available questionnaires (n = 69) to 19. In fact, in the review six different instruments were assessed; the other instruments were sub-scales or modifications of one of these six instruments (Table III). The reliability of five of these instruments was good, but the reliability of the GARS was not investigated. There is a difference between the SODA and the other instruments, in that the SODA is a performance-based instrument that measures manual activities, whereas the other instruments are questionnaires based on the patient's perception of (dis)ability to perform activities.

Generic measures have some disadvantages, for example, items are often irrelevant to persons with certain kinds of disability. Another disadvantage of most generic measures is the fact that people with low levels of physical functioning are not asked about high-levels activities, and people with high levels of physical functioning are not asked about easy activities (24).

Therefore, it is recommended to supplement generic measures with a targeted measure whenever possible (Hays *et al.*) (24)

From the clinical point of view one of the contributions of this paper is, that we grouped data on the content validity optimally by classifying all constructs, that are used to validate a measurement instrument, according to the ICF-classification. In this way, we divided optimal convergent constructs from less convergent constructs. Several levels of construct validity were distinguished in this review because it was hypothesized that validation against an instrument with an optimal convergent construct may result in stronger correlations than validation against a less convergent construct. However, the results do not support this hypothesis. In four out of eight validations, comparison with an optimal convergent construct resulted in stronger correlations than comparison with a less convergent construct. In

the other 4 validations the results were just the opposite. Even when a more strict cut-off point for good validity (all

0.65) was applied, the results with regard to the assumed hypothesis did not change (data not shown). Similar results concerning validation against an optimal or a less optimal construct have been found for instruments to measure disabilities in gait and related activities in patients with rheumatic disorders (19). In that study it was found that in 7 out of 18 instruments the correlations with optimal constructs were stronger than the correlations with less optimal constructs. For one instrument the correlations were the same for both the optimal and the less optimal construct. In 10 out of 18 instruments the correlations with optimal constructs were weaker than the correlations with less optimal constructs.

In contrast to the results reported here are the findings of an earlier study that in 11 out of 18 instruments the correlations with optimal constructs were higher than the correlations with less optimal constructs (25). However, in that study the validity of instruments to measure impairments in patients with rheumatic disorders was reviewed. The explanation for these differences in outcome might be the following. The majority of Instruments to assess impairments are uni-dimensional, while instruments to assess personal care disabilities are multi-dimensional, indicating that the total score of disabilities may be strongly influenced by one or more sub-scales (positive as well as negative) (21,23). In other words: the heterogeneity of questions and sub-scales in multi-dimensional questionnaires could explain these differences, because instruments to measure impairments are more homogeneous and uni-dimensional. Another explanation for the differences in review results could be that questionnaires to measure disabilities reflect the patient's perception of the disability at issue, whereas the majority of instruments to measure impairments are anthropometric, and reflect the assessment made by an observer.

The cut-off points for construct validity,

as presented in Table I, were chosen more or less arbitrarily, because in the literature a wide range of criteria are described, depending on the discipline underlying the study. In general, the highest cut-off criteria are applied in pharmacology (26, 27), whereas lower cut-off criteria (0.21 – 0.54) are applied in disciplines such as psychology and sociology, where there is often no gold standard available and the heterogeneity of constructs is usually greater (28). The responsiveness of the three instruments that were found to be both reliable and valid (AIMS, HAQ and SIP) is not clear. Conflicting outcomes are reported for responsiveness, probably because there is lack of consensus regarding the best method for the evaluation of responsiveness. Because AIMS, HAQ, and SIP are reliable and valid questionnaires that are widely used, the responsiveness of these instruments should be evaluated in people with rheumatic disorders. Therefore, there is need for additional and standardised research on the responsiveness of these instruments in homogeneous populations with rheumatic disorders, based on the same method for the measurement of responsiveness outcomes.

Based on the results of this review, it is now concluded that the AIMS is the most suitable instrument for the measurement of disabilities in personal care in patients with rheumatic disorders. We would like to emphasize that this conclusion is based solely on the clinimetric properties that we evaluated, and that other aspects, like availability, usefulness, difficulties to handle and interpret, time required to complete, etcetera are not assessed in this review.

### Acknowledgment

We wish to thank Mrs. Y. Kappe for reviewing the articles, together with the first author.

### References

1. FEINSTEIN AR: *Clinimetrics*. New Haven, Yale University Press, 1987.
2. KIRSCHNER B, GUYATT GH: A methodological framework for assessing health indices. *J Chron Dis* 1985; 38:27-36.
3. BOMBARDIER C, TUGWELL P: Methodological considerations in functional assessment. *J Rheumatol* 1987; 14 (Suppl. 15): 6-10.

4. BECKERMAN H, BOUTER LM: Efficacy of physiotherapy for musculoskeletal disorders: what can we learn from research? *Br J Gen Pract* 1993; 43: 483-91.
5. McDOWELL I, NEWELL C: *Measuring Health: A Guide To Rating Scales and Questionnaires*. New York, Oxford University Press, 1987.
6. PORTNEY LG, WATKINS MP: *Foundations of Clinical Research. Applications to Practice*. Appleton and Lange, 2000: 1-722.
7. 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.
8. JENSEN MP: Questionnaire validation: a brief guide for readers of the research literature. *Clin J Pain* 2003; 19: 345-52.
9. SWINKELS RAHM, VET HCWD, OOSTENDORP RAB, BOUTER LM: Construct validity of impairment measures in rheumatic disorders: Which constructs are selected for validation? 2004; submitted.
10. WORLD HEALTH ORGANIZATION: *International Statistical Classification of Diseases and Related Health Problems (ICD-10)*, 10th revision. Geneva, World Health Organisation, 1993.
11. WORLD HEALTH ORGANIZATION: *International Classification of Impairments, Disabilities and Handicaps: A Manual of Classification Relating to the Consequences of Disease*. Geneva, World Health Organisation, 1980.
12. WORLD HEALTH ORGANIZATION: *International Classification of Functioning, Disabilities and Health Problems*. Geneva, World Health Organization, 2001 (report).
13. SWINKELS RAHM: The ICF-classification as a system for structuring outcome assessment. *Physiotherapy Singapore* 2004; 7: 7-13.
14. STUCKI G, SIGL T: Assessment of the impact of disease on the individual. *Best Pract Res Clin Rheumatol* 2003; 17: 451-73.
15. OOSTENDORP RAB, ELVERS JWH, SWINKELS RAHM: Judgement list for reading articles. Amersfoort 1997, unpublished.
16. DEJONG A, MOLENAAR IW: An application of Mokken's model for stochastic cumulative scaling in psychiatric research. *J Psychiatr Res* 1987; 21: 137-49.
17. ELIASZIW M, YOUNG SL, WOODBURY MG, FRYDAY-FIELD K: Statistical methodology for the concurrent assessment of inter-rater and intrarater reliability: using goniometric measurements as an example. *Phys Ther* 1994; 74: 777-88.
18. 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.
19. 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.
20. TERWEE CB, DEKKER FW, WIERSINGAWM, PRUMMEL MF, BOSSUYT PM: On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003; 12: 349-62.
21. 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.
22. 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.
23. ZIEBLAND S, FITZPATRICK R, JENKINSON C, MOWAT A, MOWAT A: Comparison of two approaches to measuring change in health status in rheumatoid arthritis: the Health Assessment Questionnaire (HAQ) and modified HAQ. *Ann Rheum Dis* 1992; 51: 1202-5.
24. HAYS RD, HAHN H, MARSHALL G: Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. *Arch Phys Med Rehabil* 2002; 83 (Suppl. 2): S4-S9.
25. SWINKELS RAHM, OOSTENDORP RAB, BOUTER LM, ENDE CHMVD: Measuring impairments in rheumatic disorders: what are the best instruments for allied health care? 2004; submitted.
26. COX-REIJVEN PL, VAN KREEL B, SOETERS PB: Bioelectrical impedance measurements in patients with gastrointestinal disease: validation of the spectrum approach and a comparison of different methods for screening for nutritional depletion. *Am J Clin Nutr* 2003; 78: 1111-9.
27. CHRISTENSEN AE, TOBIASSEN M, JENSEN TK, WIELANDT H, BAKKETEIG L, HOST A: Repeated validation of parental self-reported smoking during pregnancy and infancy: a prospective cohort study of infants at high risk for allergy development. *Paediatr Perinat Epidemiol* 2004; 18: 73-9.
28. VLAEYEN JWS, KOLE-SNIJDERS AMJ, ROTTEVEEL AM, RUESINK R, HEUTS PHTG: The role of fear of movement/(re)injury in pain disability. *J Occup Rehabil* 1995; 5: 235-52.
29. 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.
30. MEENAN RF, GERTMAN PM, MASON JH: Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 1980; 23: 146-52.
31. KAZIS LE, ANDERSON JJ, MEENAN RF: Effect sizes for interpreting changes in health status. *Med Care* 1989; 27 (Suppl.): S178-S189.
32. 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.
33. 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.
34. 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.
35. ANDERSON JJ, CHERNOFF MC: Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *J Rheumatol* 1993; 20: 535-7.
36. POTTS MK, BRANDT KD: Evidence of the validity of the Arthritis Impact Measurement Scales. *Arthritis Rheum* 1987; 30: 93-6.
37. 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.
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, CALDWELLS, QUEEN KT *et al.*: Pain coping strategies in osteoarthritis 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. GOEPFINGER 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 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.
44. 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.
45. HAWLEYDJ, 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.
46. 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.
47. WEINBERGER M, SAMS A 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.
48. 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.

49. GOOSSENS PH, HEEMSKERK B, VAN TONGEREN J, ZWINDERMAN AH, VLIELAND TP, HUIZINGA TW: Reliability and sensitivity to change of various measures of hand function in relation to treatment of synovitis of the metacarpophalangeal joint in rheumatoid arthritis. *Rheumatology (Oxford)* 2000; 39: 909-13.
50. RIEMSMA RP, TAAL E, RASKER JJ, HOUTMAN PM, VAN PAASSEN HC, WIEGMAN O: Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. *Br J Rheumatol* 1996; 35: 755-60.
51. OOSTERVELD FGI, JACOBS JW, DEUXBOUITS N, RASKER JJ, DEQUEKER J, TAAL E: De DUTCH-AIMS bij RA: een goede maat voor fysiek functioneren. *Ned Tijdschr Fysioth* 1990; 100: 317-9.
52. 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.
53. KEMPEN GIJM, DOEGLAS DM, SUURMEIJER THBM: *Groningen Activiteiten Restrictie Schaal (GARS). Handleiding. 1-31*. Groningen, Noordelijk Centrum voor Gezondheidsvraagstukken 1993 (report).
54. SUURMEIJER TP, DOEGLAS DM, MOUM T et al.: The Groningen Activity Restriction Scale for measuring disability: its utility in international comparisons. *Am J Public Health* 1994; 84: 1270-3.
55. LEIGH JP, FRIES JF: Predictors of disability in a longitudinal sample of patients with rheumatoid arthritis. *Ann Rheum Dis* 1992; 51: 221-7.
56. WOLFE F, CATHEY MA: The assessment and prediction of functional disability in rheumatoid arthritis. *J Rheumatol* 1991; 18: 1298-306.
57. VAN DEN ENDE CH, BREEDVELD FC, DIJKMANS BA, HAZES JM: The limited value of the Health Assessment Questionnaire as an outcome measure in short term exercise trials. *J Rheumatol* 1997; 24: 1972-7.
58. 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.
59. PINCUS T, SUMMEY JA, SORACI SA JR, WALLSTON KA, HUMMON NP: Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. *Arthritis Rheum* 1983; 26: 1346-53.
60. DALTRY LH, LARSON MG, ROBERTS NW, LIANGMH: A modification of the Health Assessment Questionnaire for the spondyloarthropathies. *J Rheumatol* 1990; 17: 946-50.
61. SIEGERT CE, VLEMING LJ, VANDENBROUCKE JP, CATS A: Measurement of disability in Dutch rheumatoid arthritis patients. *Clin Rheumatol* 1984; 3: 305-9.
62. WOLFE F, KLEINHEKSEL SM, CATHEY MA, HAWLEY DJ, SPITZ PW, FRIES JF: The clinical value of the Stanford Health Assessment Questionnaire Functional Disability Index in patients with rheumatoid arthritis. *J Rheumatol* 1988; 15: 1480-8.
63. HUSTED JA, GLADMAN DD, LONG JA, FAREWELL VT: A modified version of the Health Assessment Questionnaire (HAQ) for psoriatic arthritis. *Clin Exp Rheumatol* 1995; 13: 439-43.
64. GARDINER PV, SYKES HR, HASSEY GA, WALKER DJ: An evaluation of the Health Assessment Questionnaire in long-term longitudinal follow-up of disability in rheumatoid arthritis. *Br J Rheumatol* 1993; 32: 724-8.
65. FITZPATRICK R, NEWMAN S, LAMB R, SHIPLEY M: A comparison of measures of health status in rheumatoid arthritis. *Br J Rheumatol* 1989; 28: 201-6.
66. 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.
67. BLALOCK SJ, SAUTER SV, DEVELLIS RF: The Modified Health Assessment Questionnaire Difficulty Scale. A health status measure revisited. *Arthritis Care Res* 1990; 3: 182-8.
68. STUCKI G, STUCKI S, BRUHLMANN P, MICHEL BA: Ceiling effects of the Health Assessment Questionnaire and its modified version in some ambulatory rheumatoid arthritis patients. *Ann Rheum Dis* 1995; 54: 461-5.
69. BLACKMORE MG, GLADMAN DD, HUSTED J, LONG JA, FAREWELL VT: Measuring health status in psoriatic arthritis: the Health Assessment Questionnaire and its modification. *J Rheumatol* 1995; 22: 886-93.
70. BRAZIER JE, HARPER R, MUNRO J, WALTERS SJ, SNAITH ML: Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology (Oxford)* 1999; 38: 870-7.
71. FRIES JF, SPITZ P, KRAINES RG, HOLMAN HR: Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980; 23: 137-45.
72. HUISKES CJAE, KRAAIMAAT FW, BIJLSMA JW: De ontwikkeling van de IRGL. *Gedrag & Gezondheid* 1990; 18: 78-89.
73. HUISKES CJAE, KRAAIMAAT FW, BIJLSMA JW: Development of a self-report questionnaire to assess the impact of rheumatic diseases on health and lifestyle. *J Rehabil Sci - ences* 1990; 3: 65-70.
74. BORSTLAP M, ZANT JL, VAN SOESBERGEN RM, VAN DER KORST JK: Quality of life assessment: a comparison of four questionnaires for measuring improvements after total hip replacement. *Clin Rheumatol* 1995; 14: 15-20.
75. BIJLSMA JW, HUISKES CJ, KRAAIMAAT FW, VANDERVEEN MJ, HUBER-BRUNING O: Relation between patients' own health assessment and clinical and laboratory findings in rheumatoid arthritis. *J Rheumatol* 1991; 18: 650-3.
76. MACKENZIE CR, CHARLSON ME, DIGIOIAD, KELLY K: Can the Sickness Impact Profile measure change? An example of scale assessment. *J Chron Dis* 1986; 39: 429-38.
77. MATSSON E, BROSTROM LA: The physical and psychosocial effect of moderate osteoarthritis of the knee. *Scand J Rehabil Med* 1991; 23: 215-8.
78. DEYO RA, INUI TS, LEININGER JD, OVERMAN SS: Measuring functional outcomes in chronic disease: a comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. *Med Care* 1983; 21: 180-92.
79. BERGNER M, BOBBITT RA, CARTER WB, GILSON BS: The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* 1981; 19: 787-804.
80. DE BRUIN AF, BUYS M, DE WITTE LP, DIEDERIKS JP: The sickness impact profile: SIP-68, a short generic version. First evaluation of the reliability and reproducibility. *J Clin Epidemiol* 1994; 47: 863-71.
81. 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.
82. MARTIN DP, ENGELBERG R, AGEL J, SWINTKOWSKI 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.).
83. WITTE LD, JACOBS H, HORST FVD, LUTTIK A, JOOSTEN J, PHILIPSEN H: De waarde van de Sickness Impact Profile als maat voor het functioneren van patiënten. *Gezondheid & Samenleving* 1987; 8: 120-7.
84. SULLIVAN M, AHLMEN M, BJELLE A: Health status assessment in rheumatoid arthritis. I. Further work on the validity of the sickness impact profile. *J Rheumatol* 1990; 17: 439-47.
85. VAN LANKVELD W, VAN'T PAD BP, BAKKER J, TERWINDT S, FRANSSSEN M, VAN RIEL P: Sequential occupational dexterity assessment (SODA): A new test to measure hand disability. *J Hand Ther* 1996; 9: 27-32.

## Appendix

List of abbreviations of measurement instruments and sub-scales focusing on personal care

| Abbreviation | Measurement instrument                       | Abbreviations of sub-scales |                            |
|--------------|--|-----------------------------|----------------------------|
| AIMS         | Arthritis Impact Measurement Scale           | ADL                         | Activities of Daily Living |
| AIMSD        | Arthritis Impact Measurement Scale - Dutch   | ADL                         | Activities of Daily Living |
| AIMS2        | Arthritis Impact Measurement Scale 2         |                             |                            |
| AIMS2D       | Arthritis Impact Measurement Scale2 Dutch    | Selfcare                    | Self-care                  |
| AIMSS        | AIMS short version                           | ADL                         | Activities of Daily Living |
| GARS         | Groninger Activity Restriction Scale         | ADL                         | Activities of Daily Living |
| HAQ          | Health Assessment Questionnaire              | Hygie                       | Hygiene                    |
|              |  | Eat                         | Eating                     |
|              |  | DressG                      | Dressing & grooming        |
| IRGL         | Impact of Arthritis on Health and Lifestyle  | Selfcare                    | Self-care                  |
| MHAQ         | Modified Health Assessment Questionnaire     | Hygie                       | Hygiene                    |
|              |  | Eat                         | Eating                     |
|              |  | DressG                      | Dressing & grooming        |
| SIP          | Sickness Impact Profile                      | BC&M                        | Body Care & Movement       |
| SODA         | Sequential Occupational Dexterity Assessment |                             |                            |