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

R.A.H.M. Swinkels¹, P.U. Dijkstra², L.M. Bouter³

¹Free University Brussels, Faculty of Medicine and Pharmacology, Postgraduate Education Manual Therapy, Brussel, Belgium; ²Centre for Rehabilitation, University Hospital Groningen, Groningen; ³VU 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.
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 ICIDH 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 ICIDH 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 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>
<thead>
<tr>
<th>Description</th>
<th>Intra-rater reliability</th>
<th>Optimal convergent construct validity</th>
<th>Least convergent construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>x ≥ 0.85</td>
<td>x ≥ 0.65</td>
<td>x ≥ 0.50</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.65 ≤ x &lt; 0.85</td>
<td>0.50 ≤ x &lt; 0.65</td>
<td>0.40 ≤ x &lt; 0.50</td>
</tr>
<tr>
<td>Poor</td>
<td>x &lt; 0.65</td>
<td>x &lt; 0.50</td>
<td>x &lt; 0.40</td>
</tr>
</tbody>
</table>

\[ x = \text{Pearson’s r, Spearman’s rho or Intra-Class Correlation Coefficient (ICC).} \]
Instruments to assess disabilities in rheumatic disorders / R.A.H.M. Swinkels et al.

Table II. Levels of construct convergence used to assess construct validity

<table>
<thead>
<tr>
<th>Level of convergence of constructs</th>
<th>Definition</th>
<th>Optimal convergent construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 construct</td>
<td>Validation against instruments that measure the same disability</td>
<td>Optimal convergent construct</td>
</tr>
<tr>
<td>Level 2 construct</td>
<td>Validation against instruments that measure the same disability as well as other disabilities</td>
<td>Optimal convergent construct</td>
</tr>
<tr>
<td>Level 3 construct</td>
<td>Validation against instruments that measure other disabilities than the instrument to be validated</td>
<td>Less optimal convergent construct</td>
</tr>
<tr>
<td>Level 4 construct</td>
<td>Validation against instruments that measure impairments instead of the disability at issue</td>
<td>Less optimal convergent construct</td>
</tr>
<tr>
<td>Level 5 construct</td>
<td>Validation against generic instruments that measure impairments as well as disabilities and participation problems</td>
<td>Less optimal convergent construct</td>
</tr>
</tbody>
</table>

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

Results

In total, 19 different measurement instruments or sub-scales for the assessment of personal care disabilities were identified: five integral instruments (AIMS, AIMS-D, HAQ, MHAQ and SODA) and 14 sub-scales of multidimensional 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-
Novel in our systematic review is first of all its focus on the assessment of disabilities 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

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.

<table>
<thead>
<tr>
<th>Measurement Instrument / sub-scale@</th>
<th>Construct validity if validated against instruments measuring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intra-rater reliability</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td>Level 1</td>
</tr>
<tr>
<td>AIMS (25;22;28;49)</td>
<td>0.86*</td>
</tr>
<tr>
<td>AIMS-ADL (21;22;29;49)</td>
<td>0.59</td>
</tr>
<tr>
<td>AIMS2D-Selfcare (37;48;50)</td>
<td>0.65</td>
</tr>
<tr>
<td>AIMS-ADL (45;48;51;52)</td>
<td>0.69*</td>
</tr>
<tr>
<td>AIMS2D-Selfcare (37;48;50)</td>
<td>0.56</td>
</tr>
<tr>
<td>HAQ (18;21;23;34;41;42;45;54;71)</td>
<td>0.94*</td>
</tr>
<tr>
<td>HAQ-hygie (18;21;23;34;41;42;45;54;71)</td>
<td>0.10</td>
</tr>
<tr>
<td>HAQ-eat (18;21;23;34;41;42;45;54;71)</td>
<td>0.43</td>
</tr>
<tr>
<td>HAQ-dresG (18;21;23;34;41;42;45;54;71)</td>
<td>0.21</td>
</tr>
<tr>
<td>IRGL-Selfcare (72-75)</td>
<td>0.28*</td>
</tr>
<tr>
<td>MHAQ (23;59;60;63;64;67-68)</td>
<td>0.91</td>
</tr>
<tr>
<td>MHAQ-hygie (49;50;53;54;57-60)</td>
<td>0.49</td>
</tr>
<tr>
<td>MHAQ-eat (23;59;60;63;64;67-69)</td>
<td>0.26</td>
</tr>
<tr>
<td>MHAQ-dresG (23;59;60;63;64;67-69)</td>
<td>0.92*</td>
</tr>
<tr>
<td>SODA (85)</td>
<td>0.93</td>
</tr>
</tbody>
</table>

@ For explanation of abbreviations see Appendix 1. All values expressed in Pearson’s-r or Spearman’s-rho
# Intra-class Correlation Coefficient; * Pooled value

In grey: instruments that meet the criteria for reliability as well as for validity.

Discussion

Novel in our systematic review is firstly its focus on reliability, construct validity and responsiveness of all available 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 \(\geq 0.65\)) 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 there 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, that is, in homogeneous populations the reliability 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
Instruments to assess disabilities in rheumatic disorders / R.A.H.M. Swinkels et al.


Instruments to assess disabilities in rheumatic disorders / R.A.H.M. Swinkels et al.


79. MARTIN DP, ENGELBERG R, AGEL J, SWINTOKWKO MJ: 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* 1979; 79/A(9, Sept.).


## Appendix

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

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