

Psychometric properties of a Dutch short form of the Arthritis Impact Measurement Scales 2 (Dutch-AIMS2-SF)

E. Taal¹, J. J. Rasker^{1,2} and R. P. Riemsma³

Objective. To evaluate the reliability and validity of a Dutch version of the Arthritis Impact Measurement Scales 2 short form (AIMS2-SF) and examine the agreement between the AIMS2 and AIMS2-SF in rheumatoid arthritis (RA) patients.

Methods. Data were collected from 587 RA patients from three studies. Patients completed the Dutch-AIMS2, Modified Health Assessment Questionnaire (M-HAQ), and Visual Analogue Scale for pain (VAS-pain), and clinical data were collected to calculate the Disease Activity Score 28 (DAS28). Short-form component scores were calculated from the AIMS2 long-form data. In addition, a Modified Symptom component score was calculated by replacing item 42 with item 38 as was suggested by Haavardsholm *et al.* [7] for the Norwegian version.

Results. The internal consistency of the Physical, Symptom and Affect components was good (Cronbach's $\alpha = 0.75$ – 0.87), moderate for the Role component ($\alpha = 0.62$) but rather low for the Social Interaction (0.51) component. Replacing item 33 with item 31 of the long-form AIMS2 increased internal consistency for the Social Interaction component to 0.63. Test–retest reliability of the AIMS2-SF components was high (intraclass correlation coefficients >0.70). Mean scores of the AIMS2-SF were generally close to those from the AIMS2, but the limits of agreement were rather wide. Both the Modified Symptom and Modified Social Interaction components showed better agreement than the original short-form components. Plots of differences between AIMS2 and AIMS2-SF against the mean of the two scores for the five components showed that the differences varied over the range of the measurements. Factor analysis confirmed the three-factor structure, with a physical, psychological and social dimension that has been found for the Dutch-AIMS2 long form. Correlations of the AIMS2-SF components with M-HAQ total score, functional class, VAS-pain and DAS28 were very similar to the correlations for the original AIMS2.

Conclusion. The Dutch-AIMS2-SF, with Modified Symptom and Social Interaction components has good psychometric properties, similar to those of the Dutch-AIMS2 long form.

KEY WORDS: Rheumatoid arthritis, AIMS2-SF, Outcome assessment, Psychometrics.

The Arthritis Impact Measurement Scales 2 (AIMS2) is a reliable and valid self-report questionnaire to assess health-related quality of life in arthritis patients [1]. It

has been translated and validated for use in The Netherlands and other countries [2–7]. The main part of the AIMS2 consists of 57 questions in 12 scales to

¹Department of Communication Studies, University of Twente, ²Department of Rheumatology, Medisch Spectrum Twente Enschede, The Netherlands and ³Centre for Reviews and Dissemination, University of York, UK.

Submitted 1 July 2002; revised version accepted 16 September 2002.

Correspondence to: E. Taal, University of Twente, Faculty of Behavioral Sciences Department of Communication Studies, Iustituteuweg, PO Box 217, 7500 AE Enschede, The Netherlands. E-mail: e.taal@wvm.utwente.nl

assess five components of quality of life: Physical, Symptom, Affect, Social Interaction and Role. Most patients need more than 20 min to complete all questions. This is a burden to patients and may be a limitation for its use in clinical research or to monitor patients in routine practice.

A French short form of the AIMS2 (AIMS2-SF) has been developed [8]. Two panels of patients and experts reached consensus on the items to be kept for each component in the AIMS2-SF. Choice of items was mainly based on patients' and experts' judgements of the relevance of individual items to the concept to be measured to preserve content validity. Information about psychometric properties (reproducibility, sensitivity to change) of individual items was available as an aid in selecting items when consensus could not be reached based on judgements of relevance alone. The number of items in the AIMS2-SF was reduced by 54.4% from 57 to 26, and it assesses the same five components as the long form of the AIMS2 (Table 1). In a study with 127 French RA patients the AIMS2-SF showed similar reliability, construct and convergent validity and sensitivity to change as the long form of the AIMS2. An exception was the Social Interaction component, which had low internal consistency (Cronbach's $\alpha=0.32$) [8].

The AIMS2-SF was shown to be a reliable and valid instrument among patients with osteoarthritis in the USA [9]. It has also been evaluated among RA patients in Norway [7]. The AIMS2 and AIMS2-SF showed good agreement according to the Bland-Altman approach [10]. The Symptom component showed better agreement when item 42 'How often did your pain make it difficult for you to sleep?' was replaced with item 38 'How would you describe the arthritis pain you usually had?'. The Norwegian study showed that the AIMS2-SF and AIMS2 had similar convergent validity and similar sensitivity to change.

In this study we examined the psychometric properties of a Dutch version of the AIMS2 short form

(Dutch-AIMS2-SF) and compared the agreement between the Dutch-AIMS2 and the Dutch-AIMS2-SF in RA patients. [A copy of the Dutch-AIMS2-SF is available from the first author.]

Methods

Patients

Data were used from three studies among out-patients with RA according to the 1987 American College of Rheumatology (ACR) criteria [11]. In all three studies data were collected by postal survey.

Study 1. Ten rheumatologists from three hospitals asked out-patients with RA who were not excluded by the following criteria, to participate in a survey on volunteer aid for arthritis patients. The exclusion criteria were: residence in a nursing home, disease duration of less than 5 yr, and age less than 16 yr. In each of the three hospital out-patient rheumatology clinics, consecutive male and female patients were asked until 40 women had agreed to participate, after which only consecutive male patients were asked until a total of 40 men had agreed. Two hundred and thirty-one patients returned correctly completed questionnaires. For details see Riemsma *et al.* [2].

Study 2. Three rheumatologists from two hospitals selected 160 consecutive out-patients. Inclusion criteria were a minimum age of 20 yr and a diagnosis of RA assessed by a rheumatologist according to the 1987 ACR criteria [11]. Completed questionnaires were returned by 138 patients. For details see Evers *et al.* [3].

Study 3. All out-patients from the clinics of seven rheumatologists from two hospitals with a diagnosis of RA according to the 1987 ACR criteria were asked by their rheumatologists to participate in a study on education for arthritis patients. Out of 825 RA patients, 238 returned a form stating that they agreed to participate. A composite questionnaire was posted to each participating patient. A total of 218 respondents returned the questionnaires. For this study we used data from the baseline assessment. For details see Riemsma *et al.* [12].

TABLE 1. AIMS2 and AIMS2-SF: scales, components and number of items

Component	Scale	Number of items		
		AIMS2	AIMS2-SF	
		Scale	Component	
Physical	Mobility	5	28	12
	Walking and bending	5		
	Hand and finger	5		
	Arm function	5		
	Self-care tasks	4		
	Household tasks	4		
Symptom	Arthritis pain	5	5	3
Affect	Level of tension	5	10	5
	Mood	5		
Social Interaction	Social activity	5	10	4
	Support family/friends	5		
Role	Work		4	2
Total number of items			57	26

Measures

AIMS2 and AIMS2-SF. Participants in all three studies completed the Dutch-AIMS2 questionnaire [2, 3]. All items in the Dutch-AIMS2 are measured on 5-point Likert scales scored from 1 to 5. To obtain scale scores for the AIMS2 scales, the scores on the individual items in each scale are summed. The result is then converted into a score ranging from 0 (good health) to 10 (poor health). AIMS2 and AIMS2-SF component scores for Physical, Symptom, Affect, Social Interaction and Role components were calculated. Regarding the assessment of arthritis pain we computed scores for the original Symptom component as well as for the Modified Symptom component by replacing item 42 'How often did your pain make it difficult for you to sleep?' with item 38 'How would you describe the arthritis pain you usually had?' as suggested by Havaardsholm *et al.* [7]. Component scores are calculated by the same methods as used by Guillemin *et al.*, Haavardsholm *et al.* and Ren *et al.* [7–9]. AIMS2 component scores are the averages of the scores of the corresponding scales. AIMS2-SF component scores are calculated by summing the individual item scores, and converting these sum-scores into scores ranging from 0 to 10.

Functional class. The functional classification according to Steinbrocker *et al.* [13] was obtained of all participants in study 2 and of a subsample of 77 patients from one hospital in study 1 ($n=237$).

M-HAQ. All participants in study 1 and study 3 ($n=449$) completed a Dutch version of the Modified Health Assessment Questionnaire (M-HAQ) [14, 15].

DAS28. Disease activity was calculated in study 3 only using the Disease Activity Score (DAS28) [16]. The DAS28 comprises the erythrocyte sedimentation rate (ESR) (Westergren), number of tender joints (total of 28 joints), number of swollen joints (total of 28 joints) and a 100 mm Visual Analogue Scale for general health status (range 0–100). The DAS28 ranges from 0 to 10, where 0 represents the lowest level of disease activity and 10 the highest. DAS28 scores were available for 192 patients.

VAS-pain. In study 1 and study 3 ($n=439$) pain was assessed with a 100 mm Visual Analogue Scale ranging from 0 (no pain) to 100 (severe pain).

Statistics

All statistical analyses were performed using SPSS version 9.0.

Reliability of the Dutch-AIMS2-SF scales was assessed by determining the internal consistency (Cronbach's α) [17]. Test-retest reliability was assessed by determining intraclass correlation (ICC) coefficients within a subsample of 67 patients from one hospital from study 2 who completed the Dutch-AIMS2-SF twice with a time interval of 1 month.

Floor and ceiling of the scores (% of patients with minimum and maximum scores) of each of the AIMS2-SF component scores were examined. The presence of ceiling or floor effects (excess of maximum or minimum values) indicates that the components will have reduced sensitivity for changes and sensitivity for differences between groups [18].

Agreement between Dutch-AIMS2 and Dutch-AIMS2-SF was examined according to the Bland–Altman [10] method as was done by Haavardsholm *et al.* [7] for the Norwegian version of the AIMS2-SF. In sunflower plots, in which each spike represents one case, the difference between Dutch-AIMS2 and Dutch-AIMS2-SF was plotted against the mean of the two scores for each of the components. The mean differences between AIMS2 and AIMS2-SF, and limits of agreement were plotted as lines. Limits of agreement ($d-1.96$ s.d. and

$d+1.96$ s.d.) were calculated using the mean difference (d) and the standard deviation of the differences (s.d.). If the differences are normally distributed (Gaussian), 95% of differences will lie between these limits [10]. The proportion of patients outside the limits of agreement was calculated. The degrees of correlation between the means and the differences of the two scores for each component were calculated using linear regression, and plotted as a line (r^2 of these 'lines of best fit' are presented with the sunflower plots).

Construct validity was explored by conducting a principal component factor analysis with varimax rotation on the items of the Dutch-AIMS2-SF. Items of the Role component were excluded from this analysis because these items are only applicable for patients who still work (38%). Before conducting factor analysis, variables that did not have a normal distribution were first transformed into variables with a normal distribution by means of a power transformation. Exponentiation with powers >1 reduces negative skew and exponentiation with powers <1 reduces positive skew [19].

Convergent validity of the components of the AIMS2 and AIMS2-SF was assessed by computing Pearson's correlation coefficients with Steinbrocker functional class, M-HAQ total scores, VAS-pain and DAS28.

Results

The AIMS2 questionnaires were completed by 587 people with RA. The mean age of the patients was 60.6 yr (s.d. = 11.7; range: 25–91), mean disease duration was 15.6 yr (s.d. = 11.2; range: 0–70), 8.0% of the patients had early RA (disease duration ≤ 2 yr) and 63.2% of the patients were female. Most respondents answered all questions, the percentage of missing values was below 3% for all items. An exception to this were the items of the Role component; these were only applicable to 260 patients because the other patients were unemployed, disabled or retired.

Reliability

Internal consistency of the Physical, Symptom, Modified Symptom and Affect components was good (Table 2). The internal consistency of the Role component was only 0.62, but this component consists of only two items. Internal consistency of the Social Interaction component was low (0.51), but higher than was found for the French version (0.32) [8]. Examination of the item-total correlations showed that item 33 'How often did you go to a meeting of a church, club, team or other group' correlated only 0.17 with the scale composed of the other three

TABLE 2. Internal consistency (Cronbach's α) of the Dutch-AIMS-SF components

Component	<i>n</i>	Cronbach's α
Physical	559	0.87
Symptom	582	0.75
Modified Symptom	571	0.76
Affect	574	0.78
Social Interaction	572	0.51
Modified Social Interaction	573	0.63
Role	231	0.62

items. After examination of the results of the item selection procedure in the French study about the development of AIMS2-SF we constructed a Modified Social Interaction component by replacing item 33 with item 31 'Visit friends or relatives at their homes'. This item was selected to be kept for the short form by the patient panel, but not by the expert panel in the French study [8]. In that study this item showed acceptable reproducibility (ICC = 0.63) and the highest sensitivity to change (standardized response mean = 0.33) of all social interaction items of the AIMS long form. The Modified Social Interaction component had a much better internal consistency (Cronbach's α = 0.63) than the original Social Interaction component. In further analyses in this study we used both the original and Modified Social Interaction components.

Test-retest reliability of the AIMS2-SF components was very similar to that of the original long form of the

TABLE 3. Test-retest reliability at a 1-month interval of the AIMS2-SF and AIMS2, by intraclass correlation coefficient (ICC)

Component	AIMS2-SF			AIMS2		
	<i>n</i>	ICC	95% CI	<i>n</i>	ICC	95% CI
Physical	66	0.89	0.82–0.93	65	0.92	0.88–0.95
Symptom	67	0.71	0.56–0.81	65	0.75	0.62–0.84
Modified Symptom	65	0.72	0.58–0.82	65	0.75	0.62–0.84
Affect	67	0.83	0.73–0.89	66	0.84	0.75–0.90
Social Interaction	66	0.78	0.66–0.86	64	0.76	0.63–0.85
Modified Social Interaction	65	0.72	0.58–0.82	64	0.76	0.63–0.85
Role	22	0.46	0.05–0.73	22	0.40	–0.01–0.70

TABLE 4. Per cent lowest (floor) and per cent highest (ceiling) possible scores on the AIMS2-SF components

Component	<i>n</i>	Floor (%)	Ceiling (%)
Physical	561	1.6	0.0
Symptom	584	3.9	2.6
Modified Symptom	571	1.6	3.7
Affect	574	1.9	0.0
Social Interaction	574	0.5	0.2
Modified Social Interaction	573	0.3	0.2
Role	231	14.7	2.2

TABLE 5. Agreement between AIMS2 and AIMS2-SF scores

	AIMS2 score	AIMS2-SF score	Mean (s.d.) differences	Lower limit (% patients below)	Upper limit (% patients above)	r^a (<i>P</i> value)
Physical	3.28	3.40	–0.10 (0.57)	–1.22 (1.7)	1.02 (5.5)	–0.23 (<0.01)
Symptom	5.60	4.93	0.70 (0.67)	–0.61 (2.3)	2.01 (2.6)	–0.33 (<0.01)
Modified Symptom	5.60	5.85	–0.26 (0.61)	–1.46 (2.3)	0.94 (4.4)	–0.14 (<0.01)
Affect	3.40	3.10	0.30 (0.56)	–0.80 (0.3)	1.40 (4.7)	0.08 (0.06)
Social Interaction	4.18	4.60	–0.41 (0.80)	–1.97 (2.3)	1.16 (3.4)	0.24 (<0.01)
Modified Social Interaction	4.18	4.33	–0.15 (0.74)	–1.60 (0.7)	1.30 (3.5)	0.30 (<0.01)
Role	4.67	4.11	0.54 (1.11)	–1.64 (3.5)	2.72 (1.7)	0.01 (0.87)

^a r = Pearson correlation between difference and mean of AIMS2 and AIMS2-SF.

AIMS2 (Table 3). For all components the test-retest reliability was good, except for the Role component in both the short (ICC = 0.46) and long (ICC = 0.40) forms, but this scale was only applicable for 22 patients.

Floor and ceiling effects

For most short-form component scores there were very small floor and ceiling effects. There was only a moderate but not excessive floor effect (14.7% minimum scores) for the Role component (Table 4).

Agreement between AIMS2 and AIMS2-SF

As can be seen in Table 5 and Fig. 1 the scores from the AIMS2-SF were generally close to the scores of the AIMS2. Best agreement was shown for the Physical component (mean difference = –0.10). Although mean differences between AIMS2 and AIMS2SF scores were small, the limits of agreement were rather wide. For instance the lower limit for the Physical component score on the short form is –1.22, which is 37% below the mean score on the AIMS2 long-form Physical component score, and the upper limit of the short-form score (1.02) is 31% above this mean.

The modified scores for the Symptom component as well as the Social Interaction component showed better agreement with AIMS2 scores than the original AIMS2-SF Symptom and Social Interaction scores, as can be seen by the smaller mean differences, and slightly smaller limits of agreement.

Significant correlations were found between mean scores and differences for the Physical, Symptom, Modified Symptom, Social Interaction and Modified Social Interaction components. Thus, the differences varied over the range of measurements. This can also be seen in Fig. 1 by the lines of best fit.

Construct validity

Principal component factor analysis resulted in three factors, one mirroring physical dimensions with high loadings of items from the Physical and Modified Symptom components, one mirroring psychological dimensions with high loadings of items from the Affect and Modified Symptom components and one mirroring social dimensions with high loading of items from the Modified Social Interaction component (Table 6).

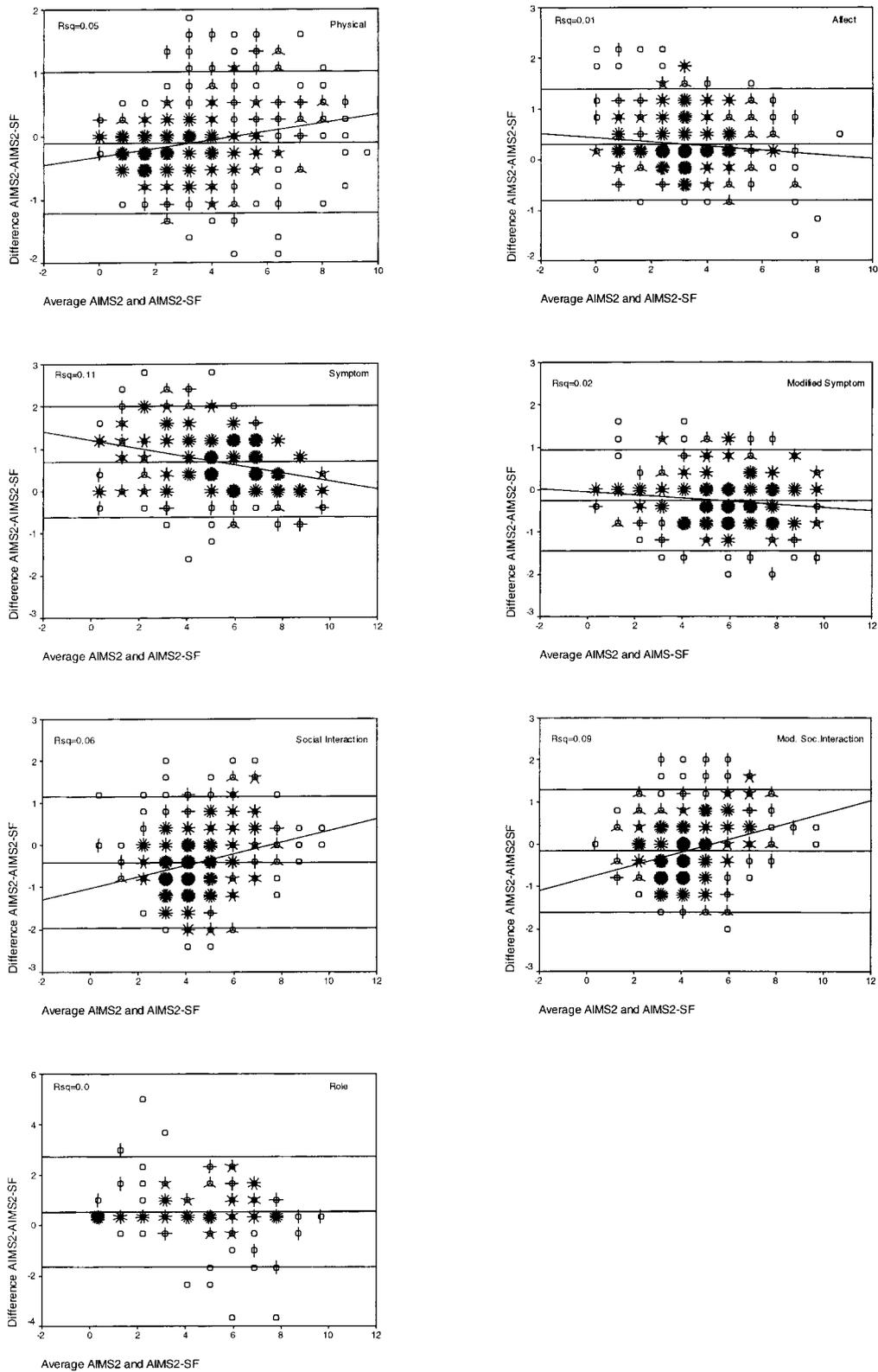


FIG. 1. Sunflower plots of difference against mean for AIMS2 and AIMS2-SF components. Each spike of a sunflower represents one case.

TABLE 6. Principal component factor analysis with varimax rotation on the items of the Dutch-AIMS2-SF (excluding two role-component items)^a

	Factor (% variance) ^b		
	Factor 1 'physical' (30.1)	Factor 2 'psychological' (9.2)	Factor 3 'social' (7.9)
1. Use a car or public transportation	0.51	–	–
2. In a bed or chair for most of the day	0.56	0.32	–
6. Trouble doing vigorous activities	0.43	–	–
7. Trouble walking several blocks or climbing a few flights of stairs	0.52	0.40	–
10. Unable to walk unless assisted	0.34	–	–
11. Write with a pen or pencil	0.69	–	–
12. Button a shirt or blouse	0.78	–	–
13. Turn a key in a lock	0.75	–	–
18. Comb or brush your hair	0.73	–	–
20. Reach a shelf that was above your head	0.78	–	–
22. Need help to get dressed	0.74	–	–
24. Need help to get in or out of bed	0.58	–	–
39. Severe pain from your arthritis	0.30	0.60	–
41. Morning stiffness lasts more than 1 h	0.47	0.46	–
38. Usual pain from your arthritis	0.39	0.54	–
48. Felt tense or high strung	–	0.82	–
49. Bothered by nervousness or your nerves	–	0.79	–
54. In low or very low spirits	–	0.72	–
53. Enjoyed the things you do	–	0.48	–
56. Others better off if you were dead	–	0.46	–
29. Get together with friends or relatives	–	–	0.78
32. On the telephone with close friends or relatives	–	–	0.72
31. Visit friends or relatives at their homes	–	–	0.67
35. Family and friends sensitive to your personal needs	–	–	0.59

^aFactor loadings ≥ 0.30 are reported in the table.^bTotal variance explained 47.1%.

TABLE 7. Pearson correlations of Dutch-AIMS2 and Dutch-AIMS2-SF components with clinical and health status measures

	Functional class <i>n</i> = 237	M-HAQ <i>n</i> = 449	VAS-pain <i>n</i> = 439	DAS-28 <i>n</i> = 192
Physical				
AIMS2	0.67**	0.87**	0.51**	0.42**
AIMS2-SF	0.63**	0.84**	0.49**	0.39**
Symptom				
AIMS2	0.43**	0.45**	0.66**	0.51**
AIMS2-SF	0.41**	0.44**	0.59**	0.47**
Modified AIMS2-SF	0.41**	0.43**	0.64**	0.46**
Affect				
AIMS2	0.34**	0.41**	0.44**	0.26**
AIMS2-SF	0.30**	0.38**	0.41**	0.20*
Social Interaction				
AIMS2	0.22**	0.23**	0.25**	0.07
AIMS2-SF	0.26**	0.25**	0.27**	0.06
Modified AIMS2-SF	0.24**	0.24**	0.23**	0.03
Role				
AIMS2	0.38**	0.50**	0.44**	0.46**
AIMS2-SF	0.43**	0.42**	0.40**	0.40**

* $P < 0.01$; ** $P < 0.001$; other correlations have P values > 0.05 .

Convergent validity

Convergent validity showed similar correlation coefficients for the AIMS2-SF components as for the long-form components (Table 7). The SF Modified Symptom component correlation with VAS-pain showed better similarity to the long-form Symptom component correlation with VAS-pain than the SF original Symptom

component. The convergent validity of the original and Modified Social Interaction components was very similar.

Discussion

Guillemin *et al.* [8] have developed a French short form of the Arthritis Impact Measurement Scales. In this

study we evaluated the psychometric properties of a Dutch version of the AIMS2-SF. The Dutch-AIMS2-SF has satisfactory psychometric properties, similar to the Dutch-AIMS2 long form after some small modifications in the Symptom and Social Interaction components. The Symptom component was modified by replacing the item 'How often did your pain make it difficult for you to sleep?' with the item 'How would you describe the arthritis pain you usually had?' as was suggested in a Norwegian study [7]. This modification led to improved agreement between short- and long-form Symptom scores and also to improved convergent validity (correlation with VAS-pain score). The Social Interaction component was modified by replacing the item 'How often did you go to a meeting of a church, club, team or other group' with the item 'Visit friends or relatives at their homes'. This modification was made because of the unacceptable low internal consistency of the Social Interaction component ($\alpha=0.51$). In the study by Guillemin *et al.* [8] an even lower internal consistency was found ($\alpha=0.42$). To preserve content validity of this scale we replaced the unsatisfactory item with an item that was felt to be relevant for assessment of social aspects by the patient panel in the French study [8]. In the French study this item also had satisfactory test-retest reliability and sensitivity to change.

Internal consistency of the Dutch-AIMS2-SF in its modified form was satisfactory. Cronbach's α coefficients for the Physical, Symptom, Modified Symptom and Affect components were above 0.70, which is generally regarded as the minimum acceptable level for psychometric scales. The α 's for the Role component ($\alpha=0.62$) and the Modified Social Interaction component ($\alpha=0.63$) are below this threshold, but above 0.60. Values of $\alpha \geq 0.60$ indicate sufficient reliability for research purposes [17]. However, the value of Cronbach's α is related to the number of items in a scale and decreases when the number of items decreases. The Role component has only two items, so an α of 0.62 is quite acceptable.

Test-retest reliability of AIMS2-SF was very similar to that of the long form. ICCs for most short- and long-form components are above the value of 0.70, which is usually recommended when scales are to be used for discriminating between groups of patients, as in clinical trials [18]. The test-retest reliability for the Role component in both the long and short forms was low, but the questions in this component were only applicable to 22 patients because many patients were unemployed, disabled or retired. ICC values for test-retest reliability were quite similar, although slightly lower, compared with the values found for the French AIMS2, but we used a rather long time interval of 1 month between assessments while Guillemin *et al.* [8] assessed patients with a short 10-day interval. It is to be expected that scores are less stable over longer time intervals.

Scores from the AIMS2-SF were generally close to the scores of the AIMS2, but limits of agreement were rather wide. The limits of the scores on the short form ranged

from 10 to 60% below or above the values of the mean scores of the AIMS2 long-form components. Such differences are certainly clinically relevant, which means that the same exact scores of AIMS2 and AIMS2-SF do not have the same exact interpretation. However, this does not mean that the AIMS2-SF is less reliable or valid in assessing health status. The Modified Symptom and Social Interaction scores for the AIMS2-SF showed better agreement with AIMS2 and improved validity than the original AIMS2-SF Symptom and Social Interaction components.

We performed principal component analysis to evaluate construct validity. This analysis resulted in a three-factor structure mirroring physical, psychological and social dimensions. The loading of items from the Modified Symptom component on the physical and psychological factors reflects the multidimensional nature of pain [20]. A comparable structure has been found for the Dutch-AIMS2 long form [2, 3]. Also convergent validity of the AIMS2-SF was similar to the validity of the AIMS2 long form.

One important property of health status scales, responsiveness or sensitivity to change, has not been assessed in this study. The AIMS2 long form has been shown to be sensitive to changes in studies in the USA, France and Sweden [4, 5, 21]. The French AIMS2-SF has similar sensitivity to change after treatment with methotrexate as the long-form AIMS2 [8]. The responsiveness of both the short and long form was high in the Physical and Symptom components, moderate in the Affect and Social Interaction components and low in the Role component. Havaardsholm *et al.* [7] also found no differences in responsiveness between the Norwegian AIMS2 short and long forms. The Physical and Symptom components were most sensitive to change. The AIMS2 long and short Physical component was shown to have similar sensitivity to change as the M-HAQ and the Physical function scale of the SF-36 [7]. Recently we have compared responsiveness of the Physical and Symptom components of the long and short forms of the Dutch-AIMS2, the M-HAQ and a VAS-Pain for changes in the patient's perception of general health. The long and short Physical and Symptom components showed similar responsiveness, while M-HAQ and VAS-pain proved to be less sensitive than both AIMS2 short and long form [22]. We used self-reported changes in general health as an indicator of responsiveness, which limits the value of our results. Sensitivity to change of the Dutch-AIMS2 short and long forms for specific interventions such as treatment with methotrexate or infliximab remains to be proven.

A limitation of our and the other studies with the AIMS2-SF is the use of data gathered with the long form of the AIMS2 [7-9]. This may act as a 'framing' effect, and may lead to an overestimation of the similarity between the two forms [7, 8]. The psychometric properties of the AIMS2-SF should also be investigated with data gathered by the short form itself.

Conclusion

In data gathered with the long form of the Dutch-AIMS2, the Dutch-AIMS2-SF was shown to have good psychometric properties after small modifications in the Symptom component, as suggested by Havaardsholm *et al.* [7], and in the Social Interaction component. We recommend use of the AIMS2-SF in its modified form.

Although the short form is as reliable and valid as the long form, there is no exact agreement between the scores of the two forms, and scores from the two forms cannot be substituted for each other. One should keep this in mind when interpreting the scores of the AIMS2-SF.

Psychometric properties of the Dutch-AIMS2-SF should still be evaluated in a study where data are gathered with the short form itself instead of the long form. Responsiveness of the Dutch-AIMS2-SF remains to be assessed in the context of effective interventions, such as treatment with methotrexate.

The number of items in the AIMS2-SF was reduced by 54.4%, so it can be expected that completion of the AIMS2-SF will take less than 10 min. Use of the AIMS2-SF, instead of the long form, reduces the burden on patients, and the short form makes it easier and less costly to collect data in clinical research and for monitoring patients in routine practice.

Acknowledgements

This study was supported by grants from the Dutch League against Rheumatism, and the Ministry of Health, Welfare and Sports of The Netherlands. We wish to thank all our respondents. We also gratefully acknowledge the co-operation of the rheumatologists H. J. Bernelot Moens, G. A. W. Bruyn, J. H. G. Bürer, J. J. M. Festen, E. N. Griep, C. J. Haagsma, P. M. Houtman, M. W. M. Kruijssen, M. A. F. J. van de Laar, P. J. H. Lanting, J. C. M. Oostveen, H. C. van Paassen and J. M. G. W. Wouters in the selection of the patients.

References

1. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status questionnaire. *Arthritis Rheum* 1992;35:1–10.
2. 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.
3. Evers AW, Taal E, Kraaijaat FW *et al.* A comparison of two recently developed health status instruments for patients with arthritis: Dutch-AIMS2 and IRGL. *Br J Rheumatol* 1998;37:157–64.
4. Pouchot J, Guillemin F, Coste J, Bregeon C, Sany J. Validity, reliability, and sensitivity to change of a French version of the Arthritis Impact Measurement Scales 2 (AIMS2) in patients with rheumatoid arthritis treated with methotrexate. *J Rheumatol* 1996;23:52–60.
5. Archenholtz B, Bjelle A. Reliability, validity, and sensitivity of a Swedish version of the revised and expanded Arthritis Impact Measurement Scales (AIMS2). *J Rheumatol* 1997;24:1370–7.
6. Brandao L, Ferraz MB, Zerbini CA. Health status in rheumatoid arthritis: cross cultural evaluation of a Portuguese version of the Arthritis Impact Measurement Scales 2 (Brasil-AIMS2). *J Rheumatol* 1998;25:1499–501.
7. Haavardsholm EA, Kvien TK, Uhlig T, Smedstad LM, Guillemin F. A comparison of agreement and sensitivity to change between AIMS2 and a short form of AIMS2 (AIMS2-SF) in more than 1000 rheumatoid arthritis patients. *J Rheumatol* 2000;27:2810–6.
8. Guillemin F, Coste J, Pouchot J, Ghezail M, Bregeon C, Sany J. The AIMS2-SF: a short form of the Arthritis Impact Measurement Scales 2. French Quality of Life in Rheumatology Group. *Arthritis Rheum* 1997;40:1267–74.
9. Ren XS, Kazis L, Meenan RF. Short-form Arthritis Impact Measurement Scales 2: tests of reliability and validity among patients with osteoarthritis. *Arthritis Care Res* 1999;12:163–71.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307–10.
11. Arnett FC, Edworthy SM, Bloch DA *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
12. Riemsma RP, Taal E, Rasker JJ. Perceptions about perceived functional disabilities and pain of people with rheumatoid arthritis: Differences between patients and their spouses and correlates with well-being. *Arthritis Care Res* 2000;13:255–61.
13. Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. *J Am Med Assoc* 1949;140:659–62.
14. Taal E, Riemsma RP, Brus HLM, Seydel ER, Rasker JJ, Wiegman O. Group education for patients with rheumatoid arthritis. *Patient Educ Couns* 1993;20:177–87.
15. Pincus T, Callahan LF, Brooks RH, Fuchs HA, Olsen NJ, Kaye JJ. Self-report questionnaire scores in rheumatoid arthritis compared with traditional physical, radiographic, and laboratory measures. *Ann Intern Med* 1989;110:259–66.
16. Prevoo ML, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44–8.
17. Nunnally JC, Bernstein JH. *Psychometric theory*, 2nd edn. New York: McGraw Hill, 1994.
18. Fayers PM, Machin D. *Quality of life. Assessment, analysis and interpretation*. Chichester, UK: John Wiley & Sons, Inc., 2000.
19. Hamilton LC. *Regression with graphics*. Pacific Grove, CA: Brooks/Cole Publishing Company, 1992.
20. Brown GK. A causal analysis of chronic pain and depression. *J Abnorm Psychol* 1990;99:127–37.
21. Anderson JJ, O'Neill A, Woodworth T, Haddad J, Sewell KL, Moreland LW. Health status response of rheumatoid arthritis to treatment with DAB486IL-2. *Arthritis Care Res* 1996;9:112–9.
22. Taal E, Rasker JJ, Riemsma RP. Sensitivity to change of AIMS2 and AIMS2-SF physical and symptom components in comparison to M-HAQ and VAS-pain. *Arthritis Rheum* 2001;44(Suppl. 9):S388.