

THE AIMS2-SF

A Short Form of the Arthritis Impact Measurement Scales 2

FRANCIS GUILLEMIN, JOËL COSTE, JACQUES POUCHOT, MARC GHÉZAIL, CHRISTIAN BREGEON, JACQUES SANY, and THE FRENCH QUALITY OF LIFE IN RHEUMATOLOGY GROUP

Objective. To develop a short form of the Arthritis Impact Measurement Scales 2 (AIMS2) questionnaire, preserving content validity as the priority criterion.

Methods. A 2-step reduction procedure was used: 1) Delphi technique, with 1 panel of patients and 1 panel of experts each selecting 1 set of items independently; and 2) nominal group technique, where members of both panels reached consensus on the final selection of items, using information derived from item analysis. Psychometric properties of the AIMS2-Short Form (AIMS2-SF) and AIMS2 were compared using data from a cohort of 127 rheumatoid arthritis patients who completed the AIMS2 twice prior to the initiation of methotrexate (MTX) treatment and 3 months post-initiation of MTX treatment.

Results. The 2 panels reached consensus on a 26-item AIMS2-SF (54.4% reduction from the AIMS2). Factor analysis showed preservation of the 5-component structure. Convergent validity (Physical and Symptom components with clinical variables: $r = 0.24-0.59$), test-retest reproducibility (intraclass correlation coefficient >0.7), and sensitivity to change at 3 months (standardized response mean $0.36-0.8$, except Social Interaction component [0.08]) were very close to the values for the original AIMS2.

Conclusion. The AIMS2-SF is a shorter version of the AIMS2 (i.e., available in 2-page format) and has psychometric properties similar to those of the AIMS2.

Francis Guillemin, MD, PhD: Hôpital Marin, Nancy, France; Joël Coste, MD, PhD: Hôpital Cochin, Paris, France; Jacques Pouchot, MD, PhD: Hôpital Louis Mourier, Colombes, France; Marc Ghézail, MD: Hôpital Henri Mondor, Créteil, France; Christian Brégeon, MD: Centre Hospitalier Universitaire, Angers, France; Jacques Sany, MD: Hôpital Lapeyronie, Montpellier, France.

Address reprint requests to Francis Guillemin, MD, PhD, SIMES-Unité d'Epidémiologie, Hôpital Marin, 54035 Nancy, France. The AIMS2-SF form is available upon request from Dr. Guillemin.

Submitted for publication November 18, 1996; accepted in revised form January 29, 1997.

The increasing interest in using health-related quality-of-life outcome measures in rheumatoid arthritis (RA) has been made possible with the development of measures specific to this condition (Health Assessment Questionnaire [1], Arthritis Impact Measurement Scales [AIMS] [2]). These instruments have been extensively validated. They have proven to be relevant for cohort studies and clinical trials, detecting health status changes over both short (3) and long (4) periods of time. Recently, the original AIMS was revised. The new instrument, the AIMS2 (5), entailed an expansion of the AIMS with additional questions and dimensions, and a modified format and time frame of response options. The revision procedure combined experience gained from the former version and reevaluation of the underlying concepts explored, thus increasing the content validity. Other aspects of validity and sensitivity to change of this questionnaire were recently documented (5,6). However, one of its potential limitations, especially when repeated assessment is required or when rapid measurement is sought, is its length. The average time to complete the instrument is estimated to be ~20-30 minutes. This may hinder its use in clinical research or routine practice because of the burden imposed on patients. To reduce the completion time, it would be useful to have a short form of the AIMS2 questionnaire for RA, in which the content validity and psychometric properties of the original are preserved.

The procedure to shorten questionnaires usually relies on statistical approaches, using item analysis methods such as correlation, item-remainder correlation, regression scores, etc. Such an approach is aimed at preserving the items with the highest correlation, or yielding the better prediction of the long form, in each of the dimensions (7). Another approach, based on the judgment of relevance of individual items to the concept, is less frequently used (8). This approach, which favors

the preservation of content validity, is closer to the methodology used to generate items when constructing a new questionnaire (9). The procedure consists of selection of the minimal number of relevant items among a broader list of items pertaining to the concepts being explored, in order to cover the area of interest without measurement overlap.

In the absence of a gold standard measure of the concept being explored, i.e., health-related quality of life in RA, the questionnaire reduction process must put much emphasis on the judgment of how close the new questionnaire should be to the concept explored (10). The goal of the present study was to shorten the AIMS2 while preserving its content validity, and to maintain equivalent properties of validity, reliability, and sensitivity to change for RA. Statistical techniques were used to supplement the clinical approach and to document the validity of items retained in the final short-form version.

PATIENTS AND METHODS

Population sample. Data were obtained from a cohort study of 127 RA patients who met the American College of Rheumatology (ACR; formerly, the American Rheumatism Association) criteria for RA (11) and were starting therapy with low-dose methotrexate (MTX) (6). All patients who had a clinical indication for such treatment were included. The first visit for this study was 10 days before MTX was started (day 0), the second was at the initiation of MTX (day 10), and there was a followup visit at 3 months (day 90). At each visit, the following variables were recorded: tender and swollen joint counts according to the definitions of the ACR Glossary Committee (12), and duration of morning stiffness. Patients completed an AIMS2 questionnaire (long form) at each visit. Data from this cohort were used to document the properties of each item of the original questionnaire, and to assess the psychometric properties of the short form resulting from the reduction process.

Original questionnaire. The AIMS2 is a self-report health status questionnaire with 57 core items organized in 12 scales (dimensions), further aggregated into 5 components, i.e., Physical (mobility level, walking and bending, hand and finger function, arm function, self-care, household tasks), Symptom (arthritis pain), Role (work), Social Interaction (social activities, support from family and friends), and Affect (level of tension, mood). Each item measurement is on a 5-point Likert scale from 0 to 4. In each dimension and component, item values are summed and the scores are normalized so as to range from 0 (perfect health) to 10 (worst health) (5). A French version of this questionnaire has been developed and shown to be equivalent in validity and reliability to its American version. Its sensitivity to change has been documented (6).

Study design. Two panels were formed to select items to be kept in the short form. A panel of experts included 19 rheumatologists, rehabilitation specialists, and methodologists

from the French Quality of Life in Rheumatology Group.* A panel of patients included 12 RA patients (members of a self-help group) who volunteered to participate in the study. In a first step, a selection of items was made independently by each panel, using a Delphi technique (13). This technique has been developed to help groups reach agreement on complex decisions. Members of each panel were asked individually to provide a written selection of the items they considered to be most relevant to each concept or construct explored by the 5 components in the long-form questionnaire. Their opinion about each item could be rated as "must be absolutely kept," "should be kept preferably," or "could be discarded." The overall panel responses indicating the proportion of members in favor of preserving each individual item were provided to all members of the panel. With these results in hand, each member had to restate his or her own written opinion individually. This procedure was iterated until no significant improvement toward an agreement was reached. No communication between members in a panel, or between panels, was allowed. The process ended with 2 independent sets of items issued by each panel.

In a second step, a nominal group process was used to reach final consensus. This consisted of a structured discussion among all members of the 2 panels, which took place until consensus on which items to retain, from the perspective both of RA patients and of health professionals, was reached. This discussion also took into account the results of item analysis (see below). The objective was to preserve the original 5-component structure in the final questionnaire.

Item selection. During the second step of the reduction procedure, information on psychometric properties of individual items was taken into account by the nominal group. An order of priority was set for item selection at the start of the procedure, favoring content validity first, reproducibility second, and sensitivity to change last. Content validity was assessed by experts and patients (see above). Reproducibility was evaluated by intraclass correlation coefficients (ICC) derived from a 2-way analysis of variance (ANOVA) in a random-effect model obtained from a test-retest procedure (day 0–day 10). Sensitivity to change was assessed by computation of standardized response means (SRM) at 3 months. SRM is the mean change in score divided by the standard deviation of the change in score (14). This information was available for each item of the original long form and was used as an aid in selecting the items to be kept in the short form when panel members could not reach consensus based on content analysis alone.

Psychometric properties of the short form. The psychometric properties obtained for the short form of the AIMS2 (AIMS2-SF) were assessed in several respects. Construct validity was explored through principal-component analysis, with the hypothesis that the 5 components of the original instrument would be preserved. Factors were retained if their eigenvalue (a statistical measure of their power to explain variation between patients) exceeded 1.2. Varimax orthogonal

* Members of the French Quality of Life in Rheumatology Group: B Avouac, D Barbazanges, I Bono, C Brégeon, C Cohen, J Coste, RL Dreiser, R Dropsy, D Dumont, M Ghézail, F Guillemin, C Hanotin, JM Joubert, JP Larbre, M Marty, J Pouchot, S Rozenberg, J Sany, A Thévenon.

rotation was used to identify items with loadings >0.4 on each factor. Internal consistency was calculated in each component using Cronbach's alpha coefficient (15). Convergent validity of the Physical and Symptom components was assessed by Pearson's correlation coefficients with the number of swollen and tender joints, using a multitrait method. Reproducibility (test-retest reliability) was tested using the ICC derived from a 2-way ANOVA, in a random-effect model. Sensitivity to change was assessed by the SRM after 3 months of MTX therapy. All of these psychometric properties were expected to be similar to those of the long form (6). All calculations were performed using BMDP software (16).

RESULTS

At the first step of the reduction process, using the Delphi technique in each group independently, expert and patient panels did not reach similar item selection. In the expert panel, 29 of the 57 items of the core questionnaire were rated by the majority of experts "must be absolutely kept," 23 items "could be discarded," and 4 items "should be preferably kept"; the remaining item (no. 43), asking to describe the type of role activity, was maintained in the final version without consideration in the content analysis. In the patient panel, 40 items were rated by the majority of patients "must be absolutely kept," 5 items "could be discarded," and 11 items "should be preferably kept." The number of items agreed on for selection or rejection was 49 and 47 at the first round of the Delphi technique in the expert and patient panels, respectively, and 45 and 52 at the second round. Since no further significant improvement in agreement was expected, no additional round was conducted in either panel. A synthesis of these group decisions showed a concordance between panels for rejection of 5 items and retention of 23 items, and a discordance regarding 28 items (Table 1).

At the second step of the reduction procedure, using a nominal group method in which patients and experts met together, the final decision was to keep 26 items. This decision excluded 9 items that both panels had agreed should be kept after the first step (39.1%), 17 items for which there was discordance between the panels (60.7%), and 4 items that both panels agreed should be rejected (80%). The nominal group choices relied mainly on item content analysis (Table 1). Although item statistical properties, i.e., reproducibility by ICC and sensitivity to change by SRM, were available to participants during the session, they were not used by the group as a basis to make decisions. The median ICC was 0.66 (range 0.48–0.90) for items kept in the AIMS2-SF and 0.64 (range 0.34–0.88) for items rejected. The median absolute SRM value was 0.28 (range

0.18–0.73) for items kept in the AIMS2-SF and 0.27 (range 0.01–0.60) for items rejected.

Item content was balanced in terms of choice of items within each component under consideration. In the AIMS2-SF, the number of core items selected from the original dimensions was as follows: 3 (of 5) items from walking and bending, 3 (of 5) items from hand and finger function, 3 (of 5) items from arthritis pain, 3 (of 5) items from social activities, 3 (of 5) items from mood, 2 (of 5) items from mobility, 2 (of 5) items from arm function, 2 (of 4) items from self-care, 2 (of 5) items from level of tension, 2 (of 4) items from work, 1 (of 4) items from support from family and friends, and 0 (of 4) items from household tasks. Thus, the selection of items in each of the 5 components was as follows: 12 items (42.8%) in the Physical component, 4 items (44.4%) in the Social Interaction component, 3 items (60%) in the Symptom component, 5 items (50%) in the Affect component, and 2 items (50%) in the Role component. The overall reduction rate from the long to the short form was 54.4%. Time frame being addressed, response format for each item, and the scoring method with normalization from 0 (perfect health status) to 10 (worst health status) were kept identical to the original questionnaire.

Scores on the AIMS2-SF were very close to those on the long form (Table 2). Departures from the original scores were higher in the Social Interaction component (+11.5%) and slightly lower in the Symptom (–7.8%) and Affect (–3.9%) components. The principal-component analysis, performed after excluding the 2 items of the Role component (because ~50% of the patients were unemployed), identified 5 factors explaining 57.2% of the cumulated variance, namely, upper limb function, lower limb function, and the Affect, Pain, and Social components, with 8, 6, 4, 3, and 3 items, respectively (Table 3). Two items originally included in the self-care dimension, involving upper and lower limbs, loaded equally on upper and lower limb function factors. This factorial structure was close to the 5 components conceptualized in the long form, with the Physical component divided into 2 parts, the Role component being excluded from this factor analysis.

Internal consistency was high in all components of the AIMS2-SF except the Social Interaction component (Table 4). Convergent validity showed similar correlation coefficients to those found using the long form (Table 5). Reproducibility was also very similar to that of the original form, ranging from 0.76 in the Social Interaction component to 0.8 in the Physical component (Table 6). The 3-month sensitivity to change of the short form proved also to be very close to that of the long form (Table 7). Mean

Table 1. Final decision on items kept in the AIMS2-SF by the nominal group, using independent decisions by expert and patient panels (Delphi technique), and item statistical analysis*

Component, item	Decisions by each panel using Delphi technique†				Item analysis		
	Discordant		Concordant		Reproducibility (ICC)	Sensitivity to change (SRM)	Final decision
	Item kept by expert panel	Item kept by patient panel	Item kept	Item rejected			
Physical							
1. Drive a car or use public transportation	–	–	X	–	0.76	0.21	Kept
2. Out of the house for at least part of the day	–	–	X	–	0.72	0.28	Rejected
3. Do errands in the neighborhood	–	X	–	–	0.72	0.43	Rejected
4. Have someone to assist you to get around outside your home	–	–	X	–	0.57	0.02	Rejected
5. In a bed or a chair for most or all of the day	X	–	–	–	0.65	0.43	Kept
6. Trouble doing vigorous activities	–	–	X	–	0.51	0.41	Kept
7. Trouble walking several blocks or climbing a few flights of stairs	–	X	–	–	0.59	0.34	Kept
8. Trouble bending, lifting, or stooping	–	X	–	–	0.68	0.31	Rejected
9. Trouble walking one block or climbing a flight of stairs	X	–	–	–	0.66	0.39	Rejected
10. Unable to walk unless assisted	–	–	X	–	0.90	0.26	Kept
11. Write with a pen or pencil	–	–	X	–	0.48	0.24	Kept
12. Button a shirt or blouse	–	–	X	–	0.62	0.26	Kept
13. Turn a key in a lock	–	X	–	–	0.55	0.28	Kept
14. Tie a knot or a bow	–	–	–	X	0.64	0.23	Rejected
15. Open a new jar of food	–	X	–	–	0.66	0.32	Rejected
16. Wipe your mouth with a napkin	–	–	–	X	0.35	0.27	Rejected
17. Put on a pullover sweater	–	–	X	–	0.57	0.54	Rejected
18. Comb or brush your hair	–	–	X	–	0.63	0.44	Kept
19. Scratch your low back with your hand	–	X	–	–	0.57	0.20	Rejected
20. Reach shelves that were above your head	–	X	–	–	0.62	0.52	Kept
21. Need help to take a bath or shower	X	–	–	–	0.84	0.27	Rejected
22. Need help to get dressed	–	–	X	–	0.77	0.46	Kept
23. Need help to use the toilet	–	–	X	–	0.80	0.07	Rejected
24. Need help to get in or out of bed	–	–	X	–	0.76	0.14	Kept
25. Go shopping for groceries without help	–	X	–	–	0.64	0.40	Rejected
26. Prepare your own meals without help	X	–	–	–	0.64	0.24	Rejected
27. Do your own housework without help	–	–	X	–	0.51	0.49	Rejected
28. Do your own laundry without help	–	–	–	X	0.59	0.36	Rejected
Social Interaction							
29. Get together with friends or relatives	X	–	–	–	0.56	0.24	Kept
30. Have friends or relatives over to your home	–	X	–	–	0.62	0.15	Rejected
31. Visit friends or relatives at their home	–	X	–	–	0.63	0.33	Rejected
32. On the telephone with close friends or relatives	X	–	–	–	0.68	0.09	Kept
33. Go to a meeting of a church, club, team or other group	–	X	–	–	0.76	0.26	Kept
34. Family or friends would be around if you needed assistance	–	–	X	–	0.68	0.02	Rejected
35. Family and friends sensitive to your personal needs	–	X	–	–	0.68	0.18	Kept
36. Family and friends interested in helping you solve problems	–	X	–	–	0.68	0.01	Rejected
37. Family and friends understood the effects of your arthritis	–	–	X	–	0.69	0.13	Rejected

* AIMS2-SF = Arthritis Impact Measurement Scales 2–Short Form; ICC = intraclass correlation coefficient; SRM = standardized response mean.

† X indicates that the item falls under the definition in the respective column heading (i.e., discordant/kept by expert panel; discordant/kept by patient panel; concordant/kept; concordant/rejected).

‡ From the French adaptation of the item originally formulated “Feel that others would be better off if you were dead” in the AIMS2.

Table 1. Continued

Component, item	Decisions by each panel using Delphi technique†				Item analysis		
	Discordant		Concordant		Reproducibility (ICC)	Sensitivity to change (SRM)	Final decision
	Item kept by expert panel	Item kept by patient panel	Item kept	Item rejected			
Symptom							
38. Describe the arthritis pain you usually had	-	-	X	-	0.63	0.58	Rejected
39. Severe pain from your arthritis	-	X	-	-	0.66	0.61	Kept
40. Pain in two or more joints at the same time	-	X	-	-	0.61	0.45	Rejected
41. Morning stiffness lasts more than one hour from the time you woke up	-	-	X	-	0.79	0.73	Kept
42. Pain makes it difficult for you to sleep	-	-	X	-	0.72	0.63	Kept
Role							
44. Unable to do any paid work, home work, or school work	-	-	X	-	0.73	0.23	Kept
45. Have to work a shorter day	-	X	-	-	0.88	0.02	Rejected
46. Unable to do your work as carefully and accurately as you would like	-	X	-	-	0.68	0.42	Rejected
47. Have to change the way your work is usually done	-	X	-	-	0.83	0.25	Kept
Affect							
48. Felt tense or high strung	-	X	-	-	0.73	0.40	Kept
49. Bothered by nervousness or your nerves	-	-	X	-	0.72	0.20	Kept
50. Able to relax without difficulty	-	-	X	-	0.34	0.13	Rejected
51. Felt relaxed and free of tension	-	-	-	X	0.34	0.23	Rejected
52. Felt calm and peaceful	-	X	-	-	0.64	0.21	Rejected
53. Enjoy the things you do	-	-	X	-	0.56	0.19	Kept
54. In low or very low spirits	-	-	X	-	0.65	0.32	Kept
55. Feel that nothing turned out the way you wanted it to	-	X	-	-	0.62	0.60	Rejected
56. Feel a burden to others‡	X	-	-	-	0.79	0.26	Kept
57. Feel so down in the dumps that nothing could cheer you up	-	-	-	X	0.65	0.33	Rejected

differences of scores over time were similarly significant, and the SRM was low in the Role component, moderate in the Affect and Social Interaction components, and high in the Physical and Symptom components.

DISCUSSION

In order to address recent recommendations for shortening composite measurement scales, instruments

Table 2. Baseline score in each component by the short-form (Arthritis Impact Measurement Scales 2-Short Form [AIMS2-SF]) and long-form (AIMS2) questionnaires

Component	n	AIMS2-SF, mean ± SD	AIMS2, mean ± SD	Difference, %
Physical	127	3.6 ± 1.9	3.6 ± 2.0	+1.0
Symptom	127	6.7 ± 2.4	7.3 ± 2.1	-7.8
Affect	127	4.1 ± 1.9	4.2 ± 1.8	-3.9
Social Interaction	127	4.9 ± 1.3	4.4 ± 1.6	+11.5
Role	78	4.3 ± 3.4	4.4 ± 3.2	+1.0

with established validity and relevance to the context of use must be available (10). The construction of a health-related quality-of-life instrument usually consists of generating numerous items to explore the concepts at stake, and further selecting the most relevant items in terms of content, adequate level of measurement of the scale, and avoidance of overlap. With further reduction, there is the risk of generating an unbalanced selection of items and losing some relevance and validity. The challenge in the shortening procedure is to maintain such properties while preserving the original content.

The AIMS2 has a long development history and extensive testing, is relevant to clinical research, and has shown responsiveness in clinical trials (17) and in situations of expected health-related quality-of-life changes (6). The use of simple statistical correlational methods to reduce the number of items would lead to misleading results if this were the sole method for reduction: it would reinforce major statistical traits without preserv-

Table 3. Factor analysis of 24 items (excluding 2 Role-component items) of the Arthritis Impact Measurement Scales 2–Short Form

Item	Factor (% variance)*				
	Factor 1: upper limb function (28.3)	Factor 2: lower limb function (9.1)	Factor 3: Affect (8.1)	Factor 4: Symptom (6.2)	Factor 5: Social Interaction (5.5)
12. Button a shirt or blouse	0.824	–	–	–	–
13. Turn a key in a lock	0.812	–	–	–	–
11. Write with a pen or pencil	0.794	–	–	–	–
20. Reach shelves that were above your head	0.652	–	–	–	–
18. Comb or brush your hair	0.558	–	–	–	–
1. Drive a car or use public transportation	0.540	–	–	–	–
22. Need help to get dressed	0.531	0.472	–	–	–
6. Trouble doing vigorous activities	–	0.789	–	–	–
7. Trouble walking several blocks or climbing a few flights of stairs	–	0.709	–	–	–
5. In a bed or a chair for most or all of the day	–	0.520	–	–	–
10. Unable to walk unless assisted	–	0.473	–	–	–
24. Need help to get in or out of bed	0.413	0.401	–	–	–
35. Family and friends sensitive to your personal needs	–	0.340	–	–	–
48. Felt tense or high strung	–	–	0.830	–	–
49. Bothered by nervousness or your nerves	–	–	0.817	–	–
54. In low or very low spirits	–	–	0.674	–	–
56. Feel a burden to others	–	–	0.513	–	–
41. Morning stiffness lasts more than one hour	–	–	–	0.756	–
42. Pain makes it difficult for you to sleep	–	–	–	0.742	–
39. Severe pain from your arthritis	–	–	–	0.727	–
32. On the telephone with close friends or relatives	–	–	–	–	0.804
29. Get together with friends or relatives	–	–	–	–	0.731
53. Enjoy the things you do	–	–	–	–	0.420
33. Go to a meeting of a church, club, team, or other group	–	–	–	–	0.262

* Total variance explained: 57.2%.

ing content validity and clinical relevance. Our primary concern of content validity led us to favor an expert-based approach for the selection of items. We used statistical item analysis only as complementary information in the process of content analysis and item selection, i.e., to assist in the choice between items of equal relevance that could not be chosen by consensus.

The contribution of 2 panels allowed us to incorporate a global judgment with input from 2 perspectives: the points of view of patients and of health care providers. These different perspectives were clearly expressed in the different numbers and types of items kept/rejected by each panel using the Delphi technique. In the second

step of the reduction process, pressure was put on the nominal group members to significantly reduce the questionnaire, i.e., by >50%. The main reasons for keeping items in the AIMS2-SF were appropriateness to clinical and patient experience and usefulness for exploring the broad spectrum of quality-of-life impairment, while reasons for rejecting items were redundancy or closeness between items, persistent disagreement in the whole group, or low statistical properties. This explained why some concordant decisions from the first step were revised in this second step of the reduction process.

There are limitations to our methodology. First, the members of the patient panel may not be representative: the fact that they were participants in a self-help group could possibly have caused a bias toward more severe impairment. However, the patients were well aware of the potential severity of the disease to be captured by the questionnaire. Similarly, rheumatologists and rehabilitation specialists are well aware of the potential articular destruction and severity of functional disability in RA patients, and might have been influenced in their choice of items aimed at more severe conditions. In order to limit this bias, the focus of the

Table 4. Internal consistency of each component of the Arthritis Impact Measurement Scales 2–Short Form questionnaire

Component	Cronbach's alpha
Physical	0.87
Symptom	0.74
Affect	0.75
Social Interaction	0.32
Role	0.74

Table 5. Convergent validity of each AIMS2-SF and AIMS2 component with clinical variables*

	AIMS2-SF			AIMS2		
	Duration of morning stiffness	Tender joint count	Swollen joint count	Duration of morning stiffness	Tender joint count	Swollen joint count
Physical Symptom	0.35†	0.39†	0.24†	0.36†	0.36†	0.16
Affect	0.59†	0.43†	0.28†	0.60†	0.39†	0.14
Social Interaction	0.33†	0.42†	0.29†	0.29†	0.28†	0.23†
Role	0.06	0.16	0.15	0.004	0.02	0.07
	0.23	0.32†	0.09	0.16	0.18	0.03

* Correlation coefficients, by a multitrait method. AIMS2-SF = Arthritis Impact Measurement Scales 2-Short Form.

† $P < 0.05$.

panels was directed toward considering the full spectrum of the disease explored.

Second, our use of data gathered with the long-form questionnaire to test for the properties of the short form may present a methodologic problem. One may assume that the long form used to collect data acts as a framing effect. Thus, data collected with the short form itself, i.e., within another frame, could slightly differ. One can expect that our results could overestimate the similarity between the long and short forms. Further independent testing of the administration of the short form is in progress.

Our strategy was to maintain the content validity within the 5-component structure identified in the original AIMS2. This provided a useful framework to balance the number of items selected in each component and concept explored (but not in the 12-dimension structure). Although no dimension was to be kept purposely in the process, most of the original dimensions were represented in the AIMS2-SF by 1-3 items; only 1 dimension (household tasks) was no longer represented. We maintained the same scoring method as the original one for each component, i.e., original Likert scale answers, item recoding, items averaging within each component, and normalizing scale scores from 0 to 10.

In the selection process, we assigned a hierarchi-

cal priority to content over reproducibility and sensitivity to change of items. This was in accordance with the rationale to not consider the original as a gold standard. Thus, it replicates the natural process of development of a new instrument with a predetermined framework, represented by the targeted content of the original instrument.

Previously, there have been 2 reports of studies attempting to reduce the original AIMS questionnaire (45 items in 9 dimensions) (18,19). Both used an approach based on statistical item analysis methodology. These studies yielded selection of items that were remarkably similar to one another. In both studies, the main focus was on preserving internal consistency, maximizing short-long scale and item-total correlations, and preserving the number of scales. The preservation of the domain content was not explicitly stated as a priority. Long and short forms had 16 items in common among 18 and 22, respectively, in the 2 studies, and similar relationships. Our expert-based approach resulted in a short form with 9 and 10 items, respectively, in common with these short forms in the previous 2 studies (1 [or 2] from each of 9 dimensions of the first version of the AIMS). Among the 17 (or 16) remaining items, 6 originated from the expanded AIMS2, emphasizing the important contribution in the revision of the AIMS.

Table 6. Reproducibility of the Arthritis Impact Measurement Scales 2-Short Form (AIMS2-SF) and the AIMS2, by test-retest at a 10-day interval*

Component	AIMS2-SF			AIMS2		
	n	ICC	95% CI	n	ICC	95% CI
Physical Symptom	94	0.81	0.73-0.86	93	0.73	0.64-0.79
Affect	94	0.79	0.72-0.84	94	0.79	0.66-0.88
Social Interaction	93	0.80	0.74-0.85	92	0.80	0.73-0.85
Role	94	0.76	0.68-0.82	94	0.75	0.67-0.81
	53	0.80	0.69-0.87	47	0.90	0.82-0.94

* ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval.

Table 7. Sensitivity to change of the Arthritis Impact Measurement Scales 2-Short Form (AIMS2-SF) and the AIMS2 after 3 months of methotrexate treatment

Component	n	AIMS2-SF			AIMS2		
		SRM*	Paired <i>t</i> -test	<i>P</i>	SRM	Paired <i>t</i> -test	<i>P</i>
Physical Symptom	101	0.61	5.64	0.00001	0.61	6.06	0.0001
Affect	101	0.80	8.04	0.00001	0.77	7.72	0.0001
Social Interaction	101	0.41	4.15	0.0001	0.47	4.71	0.0001
Role	57	0.08	0.78	0.44	0.12	1.19	0.25
		0.36	2.76	0.008	0.32	2.37	0.02

* SRM = standardized response mean.

When comparing properties of the long and short forms using Cronbach's alpha, it should be kept in mind that reducing the number of items mathematically induces a decrease of the alpha coefficient value. Construct and convergent validity, reproducibility, and responsiveness are less dependent on the number of items. Reproducibility and responsiveness are very close in the short and long forms. Reproducibility is a prerequisite for responsiveness, but does not necessarily guarantee it (20). Responsiveness in the context of initiation of MTX therapy, a well-calibrated situation of health-related quality-of-life improvement (21,22), is a key aspect of the usefulness of the AIMS2-SF: it proved satisfactory for future use in clinical trials, just as the full AIMS2 is. The use of such instruments is generally applicable at the group level, and mainly in research. Their relevance as an outcome measure at the individual level for clinical practice remains, however, questionable (23). Finally, although the original AIMS2 questionnaire has been used successfully in other rheumatic diseases, the performance characteristics of this short form are documented only for RA and should be further confirmed in other diseases.

In conclusion, a reduced form of the validated AIMS2 questionnaire has been developed using a comprehensive expert-based approach, and supported by psychometric properties testing. This new version, which is available in a 2-page format, reduces the burden to patients and the length of investigation for research, while its validity, reproducibility, and sensitivity to change are preserved. This shortened instrument is easier to use for repeated measures in longitudinal designs of RA studies, and less costly for use in large clinical trials and cohort investigations.

ACKNOWLEDGMENTS

We are indebted to Paul Fortin and Mario Cardiel-Ros for their critical reading and fruitful suggestions on an earlier draft of the manuscript.

REFERENCES

- Fries JF, Spitz P, Kraines RG, Holman HR: Measurement of patient outcome in arthritis. *Arthritis Rheum* 23:137-145, 1980
- Meenan RF, Gertman PM, Mason JH: Measuring health status in arthritis: the Arthritis Impact Measurement Scales. *Arthritis Rheum* 23:146-152, 1980
- Meenan RF, Anderson JJ, Kazis LE, Egger MJ, Altz-Smith M, Samuelson CO Jr, Willkens RF, Solsky MA, Hayes SP, Blocka KL, Weinstein A, Guttadauria M, Kaplan SB, Klippel J: Outcome assessment in clinical trials: evidence for the sensitivity of a health status measure. *Arthritis Rheum* 27:1344-1352, 1984
- Wolfe F, Hawley D, Cathey MA: Clinical and health status measures over time: prognosis and outcome assessment in rheumatoid arthritis. *J Rheumatol* 18:1290-1297, 1991
- 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* 35:1-10, 1992
- Pouchot J, Guillemin F, Coste J, Brégeon C, Sany J, and the French Quality of Life in Rheumatology Group: 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* 23:52-60, 1996
- Silverstein AB: Agreement between a short form and the full scale as a function of the correlation between them. *J Clin Psychol* 45:929-931, 1989
- Parkerson GR, Broadhead WE, Tse CKJ: The Duke Health Profile: a 17 item measure of health and dysfunction. *Med Care* 28:1056-1072, 1990
- Streiner DL, Norman GR: Health Measurement Scales: A Practical Guide to Their Development and Use. Toronto, Oxford Medical Publications, 1989
- Coste J, Guillemin F, Pouchot J, Fermanian J: Methodological approaches to shortening composite measurement scales. *J Clin Epidemiol* (in press)
- Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA Jr, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 31:315-324, 1988
- ARA Glossary Committee: Dictionary of the Rheumatic Diseases. Vol. 1. Signs and Symptoms. New York, Contact Associates International, 1982
- Dalkey NC, Helmer O: An experimental application of the Delphi method to the use of experts. *Manage Sci* 9:456-457, 1963
- Liang MH, Fossel AH, Larson MG: Comparison of five health status instruments for orthopedic evaluation. *Med Care* 28:632-642, 1990
- Nunnally JC: Psychometric Theory. New York, McGraw-Hill, 1978
- Dixon WJ: BMDP Statistical Software Manual. Bognor Regis, CA, John Wiley and Sons, 1992
- Liang MH, Larson MG, Cullen KE, Schwartz JA: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 28:542-547, 1985
- Wallston KA, Brown GK, Stein MJ, Dobbins CJ: Comparing the short and long versions of the Arthritis Impact Measurement Scales. *J Rheumatol* 16:1105-1109, 1989
- Lorish CD, Abraham N, Austin JS, Bradley LA, Alarcón GS: A comparison of the full and short versions of the Arthritis Impact Measurement Scales. *Arthritis Care Res* 4:168-173, 1991
- Guyatt GH, Kirshner B, Jaeschke R: Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 45:1341-1345, 1992
- Tugwell P, Bombardier C, Buchanan WW, Goldsmith C, Grace E, Bennett KJ, Williams HJ, Egger M, Alarcon GS, Guttadauria M, Yarboro C, Polisson RP, Szydlo L, Luggen ME, Billingsley LM, Ward JR, Marks C: Methotrexate in rheumatoid arthritis: impact on quality of life assessed by traditional standard-item and individualized patient preference health status questionnaires. *Arch Intern Med* 150:59-62, 1990
- Wolfe F, Cathey MA: Analysis of methotrexate treatment effect in a longitudinal observational study: utility of cluster analysis. *J Rheumatol* 18:672-677, 1991
- McHorney CA, Tarlov AR: Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 4:293-307, 1995