

Ceiling effects of the Health Assessment Questionnaire and its modified version in some ambulatory rheumatoid arthritis patients

Gerold Stucki, Susanne Stucki, Pius Brühlmann, Beat A Michel

Abstract

Objective—To examine if the reduced number of items in the modified version of the Health Assessment Questionnaire (MHAQ) concerning difficulty of performing activities of daily living may lead to a reduced ability to detect clinical changes compared with the original HAQ.

Methods—In 56 consecutive ambulatory patients with rheumatoid arthritis, we examined the mean change in clinical and laboratory parameters for those who recorded improved, unchanged, or worse MHAQ scores one year after a baseline assessment.

Results—At baseline, about 50% of the patients had an MHAQ score < 0.3 and clustered at the normal end of the scale. Because of a ceiling effect, the MHAQ failed to detect clinical improvement in 18% of the patients. Changes in clinical and laboratory parameters were associated with improved, unchanged, or worse scores with the HAQ but not the MHAQ.

Conclusion—Although the format of the MHAQ has the advantage of eliciting a 'satisfaction' score, limitations in its sensitivity to detect clinical improvement in patients with relatively little difficulty in activities of daily living may not justify the use of this particular version of shorter questionnaire in certain clinical settings.

(*Ann Rheum Dis* 1995; 54: 461–465)

Self reported measures of health related quality of life are increasingly used to complement clinical information in clinical research, whereas their usefulness in clinical practice is still debated.^{1,2} One barrier to more widespread use of such questionnaires is their length;³ consequently, existing questionnaires have been shortened by omission of subscales,⁴ omission of items, or both.^{1,5} However, short forms have inherent weaknesses relative to the longer versions.

The Health Assessment Questionnaire (HAQ)⁶ is one of the most widely used physical function scales in rheumatic disease.⁷ Pincus *et al* produced the modified HAQ (MHAQ) by reducing the questionnaire from 20 to eight questions¹ and supplemented the original questions assessing level of difficulty with additional questions assessing patient satis-

faction regarding the same activities of daily living—a unique and useful feature.⁸ In a cross sectional study the eight items framed in the original 'difficulty' format captured most of the information represented by the 20 HAQ items.¹ The only study that assessed the difficulty format of the MHAQ longitudinally⁴ found the MHAQ to be as responsive as the physical dimensions of the Sickness Impact Profile,⁹ the SF-36,⁵ the Functional Status Questionnaire,¹⁰ and the short version of the Arthritis Impact Measurement Scale.¹¹ The study did not compare the responsiveness of the MHAQ with that of the full HAQ, nor did it assess if the questionnaire was sensitive to small but clinically meaningful changes.

The objective of this study was to evaluate whether the difficulty section of the MHAQ is sensitive to clinical changes in ambulatory rheumatoid arthritis (RA) patients.

Methods

PATIENTS

Sixty two consecutive patients with RA fulfilling the American Rheumatism Association 1987 revised criteria¹² attending the rheumatology outpatient clinic at the University Hospital Zurich were included in the study.

DATA COLLECTION

At a scheduled outpatient visit, each patient was asked to complete the HAQ.^{6,13} The MHAQ was not administered separately, but was scored from the same questionnaire with use of the algorithm for the MHAQ. The patient was then evaluated clinically, and laboratory tests were performed. One year later, the patient was invited for a follow up examination.

MEASURES

The HAQ is one of the most widely used measures of limitation in activities of daily living.⁶ It addresses difficulty in eight domains including dressing, arising, eating, walking, hygiene, reaching, gripping, and other activities. Whereas each domain is represented by two to four items in the original HAQ, one question represents each domain in the MHAQ.¹ Patient self report was used for the measurement of pain (numerical rating scale (NRS) 0–10) and duration of morning stiffness (minutes).

Department of
Rheumatology and
Physical Medicine,
University Hospital
Zürich,
Switzerland
G Stucki
S Stucki
P Brühlmann
B A Michel

Correspondence to:
G Stucki,
Rheumaklinik und Institut
für physikalische Therapie,
Universitätsspital,
Gloriastrasse 25,
8091 Zürich,
Switzerland.

Accepted for publication
30 January 1995

Clinical evaluation included swollen and tender joint count, grip strength (mean of both sides), and muscle strength (muscle strength index: mean of elbow and knee extension and flexor strength of both sides expressed as percentage of the maximal strength in the population.¹⁴ Laboratory assessment included erythrocyte sedimentation rate (ESR) and haemoglobin. Overall disease activity was measured by physician's global assessment with an NRS and two pooled indices (disease activity score (DAS)¹⁵ and Mallya index.¹⁶ Tender joint count was used instead of the Ritchie articular count for both indices (for the calculation of the DAS, the tender count was multiplied by 2).

ANALYSES

To compare the cross sectional discriminative ability of the MHAQ with that of the full HAQ, we examined the distribution in stem and leaf diagrams at baseline and one year follow up and examined the correlation of the questionnaires with clinical parameters using Spearman's rank correlation coefficient.

To examine if the MHAQ detects clinically meaningful change, we compared the mean change in clinical and laboratory parameters for patients who recorded improved, unchanged, or worse scores at the one year follow up. We hypothesised that patients whose HAQ or MHAQ scores improved would have a decrease in disease activity and pain, and an increase in muscle strength. A significant difference between the change in clinical para-

meters for patients who improved, remained unchanged, or deteriorated in instrument scores (Kruskal-Wallis test) was considered relevant only if the change was in same direction.

Data were analysed with use of the Statistical Analysis System (SAS Institute, Cary, NC). Values of *p* (two tailed) less than 0.05 were considered significant.

Results

One year follow up data were available for 56 of the 62 patients included in the baseline survey; six patients either refused to participate in the follow up or could not be contacted. Findings for the 56 patients with both baseline and follow up data are presented. Sixty one percent of the patients were female; median age was 62 years (25th percentile: 51; 75th percentile: 70) and median disease duration 5.1 years (25th percentile: 1.3; 75th percentile: 10.7); 57% had a positive rheumatoid factor titre (Singer-Plotz \geq 1.40). All patients were white; 97% graduated from high school, and 64% graduated from college or had professional training. Median MHAQ score was 0.25 (0; 0.75), whereas median HAQ score was 1.19 (0.31; 1.75), which indicates moderate disability. Eighty eight percent of the patients were treated with slow acting drugs, 45% were taking corticosteroids, and 94% were taking non-steroidal antirheumatic drugs. Table 1 shows the baseline clinical characteristics (25th, 50th, and 75th percentile values).

Table 1 Comparison of the rank correlation of the Health Assessment Questionnaire (HAQ) and the difficulty section of the modified version of the Health Assessment Questionnaire (MHAQ) with clinical and laboratory parameters in 56 patients with RA at baseline

Variable	Percentile values (25th, 50th, 75th)	Correlation with	
		HAQ	MHAQ
Disease activity physician (0-10)	2, 3, 5	0.55**	0.45**
Disease activity score ¹⁵	2.7, 3.6, 4.9	0.53**	0.53**
Mallya index (0-4) ¹⁶	1.8, 2.3, 2.7	0.74**	0.59**
Swollen joint count (0-78)	2, 5, 11	0.25	0.21
Tender joint count (0-68)	2, 6, 15	0.55**	0.51**
Grip strength (kp/cm ²)	0.19, 0.3, 0.42	-0.62**	-0.51**
Strength index (%)	29, 39, 51	-0.61**	-0.52**
Pain (NRS 0-10)	3, 5, 6	0.54**	0.52**
Morning stiffness (min)	0, 15, 60	0.55**	0.36*
ESR (mm/1st h)	12, 18, 30	0.23	0.33*
Haemoglobin (mg/l)	121, 129, 143	-0.17	-0.11

NRS = Numerical rating scale; ESR = erythrocyte sedimentation rate.
p* < 0.05; *p* < 0.01.

Table 2 Stem and leaf diagrams¹⁷ of the Health Assessment Questionnaire (HAQ) scores and the short version of the Health Assessment Questionnaire (MHAQ) at baseline

Baseline HAQ*	n	Baseline MHAQ*	n
26 5	1	26	
24		24	
22 588	3	22	
20 0000222	7	20 5	1
18		18 8	1
16 2555555	7	16 5	1
14 000	3	14	
12 5588888	7	12 55	2
10 0000022	7	10 002	3
8		8 888	3
6 25	2	6 22255555555	10
4 0	1	4 0000	4
2 555558888	9	2 5555588	8
0 000000022	9	0 000000000000000002222222	23

*The stem ranges from 0 to 26 (corresponding to a score from 0 to 2.6). Each observation is separated in a stem and a leaf component. n = Number of patients with equal stem scores.

CROSS SECTIONAL ANALYSES

The distributions of both the MHAQ and the HAQ at baseline and at one year follow up were bimodal (table 2). Thirteen percent of patients had an HAQ score of 0 and 32% had a score < 0.3 and clustered in the first mode; the second mode of the HAQ distribution was at 1.4. Twenty nine percent of the patients had an MHAQ score of 0, and 52% had a score < 0.3 and clustered in the first mode; the second mode of the MHAQ distribution was at 0.7.

With the exception of swollen joint count and ESR, there was a significant moderate-to-strong association between the questionnaires and clinical parameters; correlations were slightly lower for the MHAQ (table 1). The correlation with clinical parameters did not increase when patients clustering at the normal end of the scale (HAQ or MHAQ < 0.3) were excluded from the analysis.

LONGITUDINAL ANALYSES

HAQ

Forty six percent of the patients had worse HAQ scores at one year, 20% had equal scores, and 34% had improved HAQ scores. The proportion of patients who showed worsening in HAQ scores was similar to the proportion showing change in clinical parameters: 46% had a worse HAQ score at one year follow up, 32% had a worse Mallya score, 41% a worse

disease activity score, 48% worse pain, 27% worse morning stiffness, and 45% of the patients perceived their condition as worse at follow up.

Patients who had improved HAQ scores on average also had decreased disease activity in terms of physician's global assessment and the Mallya index, whereas patients who had worse HAQ scores at one year had no improvement in terms of physician's global assessment and a worse Mallya index. Similarly, changes in pain and muscle strength index were significantly associated with the direction of change in HAQ scores. Changes in grip strength, the DAS¹⁵ and its components, swollen and tender joint count, and ESR were in the direction of change in HAQ scores, but the association was not significant.

In a parametric analysis (assuming interval characteristic of the HAQ) we found a significant Pearson correlation between change in HAQ and changes in physician's estimate of disease activity ($r=0.27$, $p<0.05$), Mallya index ($r=0.30$, $p<0.05$), pain ($r=0.44$, $p<0.01$), strength index ($r=-0.36$, $p<0.01$), and patient's perception of change ($r=0.29$, $p<0.05$). The correlations with morning stiffness, the DAS, swollen and tender joint counts, ESR, and haemoglobin were not significant.

Of the seven patients with an HAQ score of 0 at baseline, four improved in their Mallya score and reported improvement in the global transition question, while their HAQ scores stayed the same in three and worsened in one. Of the three patients whose Mallya score worsened and who reported deterioration of their health, two also had a worse HAQ score and one still had a score of 0 at one year follow up. Thus, in terms of the external criteria, 7% of the study population could not improve in score despite clinical improvement. These findings indicate a slight ceiling effect of the HAQ.

MHAQ

At one year follow up, 68% of the patients had worse MHAQ scores, 18% had equal scores, and 14% had improved MHAQ scores. The proportion of patients who showed worsening in MHAQ scores was much greater than the proportion showing change in clinical

parameters: whereas 68% had a worse MHAQ score at one year follow up, only 32% had a worse Mallya score, 41% a worse disease activity score, 48% worse pain, and 27% worse morning stiffness. Also, only 45% of the patients perceived themselves as worse at follow up.

In 16 patients, the direction of change was discrepant for the HAQ and the MHAQ: 14 had worsening in the MHAQ but an unchanged or improved HAQ score, whereas only two patients had a worsening in HAQ score while unchanged or improving MHAQ scores. Compared with the HAQ, the MHAQ thus classified 12 more patients as deteriorated and 12 fewer as improved ($p<0.01$, chi square test).

Change in clinical parameters did not differ significantly for patients that recorded improved, unchanged, or worse MHAQ scores (table 3). Only mean changes in Mallya index, pain, grip strength, and haemoglobin were in the direction of change in MHAQ scores; the differences were not significant. The changes in all the other clinical parameters were not associated with the direction of change in MHAQ scores; for example, patients who had improved MHAQ scores had greater duration of morning stiffness and a higher disease activity score than patients with unchanged or worse MHAQ scores.

In a parametric analysis (assuming interval characteristic of the MHAQ) change in MHAQ correlated only with change in pain ($r=0.32$, $p<0.05$), but not with the other external criteria.

Of 16 patients with an MHAQ score of 0 at baseline, 10 improved in their Mallya score and reported improvement in the global transition question, but the MHAQ scores stayed the same in six and worsened in four. Of the six patients whose Mallya score worsened and who reported deterioration of their health, five also had a worse MHAQ score and only one still had a score of 0 at one year follow up. These findings indicate an important ceiling effect, which prevented detection of improvement in terms of the external criteria in about 18% of the study patients.

Discussion

In this ambulatory population of consecutive patients with RA, the MHAQ, and to a lesser

Table 3 Mean change in clinical parameters for patients who recorded improved, unchanged or worse instrument scores after one year

	Mean change in clinical parameters							
	Improved HAQ (n = 19)	Unchanged HAQ (n = 11)	Worse HAQ (n = 26)	p*	Improved MHAQ (n = 8)	Unchanged MHAQ (n = 10)	Worse MHAQ (n = 37)	p*
Disease activity physician (0-10)	-1.11	0.1	0.0	0.02	-1.25	0.20	-0.32	0.41
Disease activity score ¹⁵	-0.95	-0.04	1.16	0.81	3.2	-0.73	-0.18	0.19
Mallya index (0-4) ¹⁶	-1.59	-1.0	1.12	<0.01	-0.45	-0.12	-0.06	0.29
Swollen joint count (0-78)	-1.47	-1.70	-0.12	0.68	-1.75	-2.4	-0.27	0.34
Tender joint count (0-68)	-2.0	0.4	2.2	0.63	4.12	-1.00	-0.18	0.43
Grip strength (kp/cm ²)	0.063	0.060	0.052	0.95	0.183	0.081	0.022	0.10
Strength index (%)	5.7	4.5	-1.4	0.02	3.5	8.9	0.1	0.02
Pain NRS (0-10)	-0.37	0.19	1.52	0.01	0.1	0.4	0.8	0.77
Morning stiffness (min)	-16.2	10.3	9.0	0.08	12.5	-5.0	-0.8	0.80
ESR (mm/1st h)	-4.4	-6.4	0.42	0.53	-10.3	0.3	-1.6	0.65
Haemoglobin (mg/l)	2.3	2.1	-1.5	4.1	2.3	1.0	-0.1	6.4

NRS = Numerical rating scale; ESR = erythrocyte sedimentation rate. *Kruskal-Wallis test (significance under the condition that the mean change is in the same direction).

extent the HAQ, had a ceiling effect. Cross sectionally, the MHAQ did not discriminate physical functional disability in about 50% of the patients, while the HAQ did not discriminate physical functional disability in about 33% of the patients. Longitudinally, the MHAQ did not detect clinical improvement in about 20% of the patients, and the HAQ could not improve further despite clinical improvement in about 7% of the patients. The much stronger ceiling effect of the MHAQ compared with that of the HAQ is probably responsible for the large discrepancy between the two questionnaires in the ratio of patients who improved, remained the same, or worsened over the course of one year. Comparison of the number of patients recording a worse MHAQ score at one year with the proportion of patients who recorded worse HAQ scores, and the proportion who had worse clinical parameters, indicates an important and directional misclassification of patients.

Clinical parameters are not the gold standard for the validation of physical functional status and change in physical functional status. However, because clinical impairment and physical ability are closely related, one would expect an association both cross sectionally and longitudinally. Consistent with this expectation and the result of previous studies,^{1 6} we found a strong relationship between both the HAQ and the MHAQ and clinical parameters in the cross sectional assessment. However, only change in HAQ, not that in MHAQ scores, was significantly related to changes in clinical parameters in both non-parametric and parametric analyses. Several clinical parameters improved significantly for the group of patients with improved HAQ compared with the change in these parameters for the groups of patients with unchanged or worse HAQ scores at one year. The reason why the MHAQ did not distinguish patients who improved clinically from those who remained unchanged or worsened was probably a strong ceiling effect, which led to misclassification of patients as worsened or unchanged despite clinical improvement.

Our findings illustrate that a health status measure that has been validated in one setting may not be valid in another population. Whereas the MHAQ was valid in the population for whom it was developed, it did not detect clinical improvement in an important subset of ambulatory RA patients. The level of disability in our study population was not exceptional: the mean HAQ score of 1.1 is only slightly less than the 1.3 observed in a study by the American College of Rheumatology (ACR) subcommittee for the validation of the ACR functional classes (mean HAQ 1.3, indicating moderate disability).¹⁸ To be useful in ambulatory settings similar to ours, health status measures will need to include items with a high degree of 'difficulty' scoring to permit discrimination and detection of clinical change.

Our study also illustrates that short forms, although more practical, may have serious limitations. Shortening of questionnaires may result in restriction of the range of patients who

can be assessed. One simple test to detect underrepresentation of difficult items is the mean value of the total score. The greater median score of the HAQ (1.2) compared with the MHAQ (0.25) indicates that the HAQ contains more difficult items than the MHAQ. The smaller mean MHAQ score is consistent with a previous report¹⁹ and indicates that HAQ and MHAQ scores are not interchangeable. Thus it would seem critical to choose items from the full spectrum of difficulty when developing short versions of existing questionnaires.

Cross sectional examination of the association of the MHAQ with clinical and laboratory parameters did not reveal its inability to discriminate a relatively large group of patients with little physical disability. Only inspection of the instrument distribution allowed detection of a ceiling effect, and only longitudinal assessment allowed detection of the inability of the MHAQ to discriminate between patients who improved and those who did not.

This analysis emphasises the importance of studying patients in a 'real life' setting. Only if an instrument is able to detect small but clinically meaningful changes can it be useful in clinical practice. Sensitivity to change assessed in an 'assay system' such as a surgical intervention^{20 21} provides important information on the relative responsiveness of measures; however, as demonstrated, a measure that proved to be sensitive in one setting⁴ may not be sensitive to clinical change in another. Of particular concern, an instrument that is sensitive in detecting deterioration may fail to detect improvement.

Several limitations and decisions with respect to the analyses require comment. First, we did not administer the MHAQ as a separate questionnaire but scored it using HAQ forms. Thus we cannot exclude the possibility that independent administration of the MHAQ alone would have changed the result. Second, in common with other health status measures, the HAQ and MHAQ are ordinal scales^{22 23} and a non-parametric analysis may be more appropriate. Interestingly, both a non-parametric analysis using the direction of change and a parametric analysis using the magnitude of change resulted in virtually identical results. Third, we used the Mallya index and patients' perception of change as external criteria for the assessment of a longitudinal ceiling effect. We chose the Mallya index because it contains six clinical and laboratory variables and is a broad measure of overall disease activity and severity. Transition questions have face validity and have been shown to correlate with change in clinical parameters.⁸ However, neither criterion is a gold standard for change in physical functional disability. The estimates of the number of patients showing a ceiling effect therefore need to be interpreted with caution and in a qualitative rather than a quantitative way.

In summary, the MHAQ, and to a lesser extent the HAQ, did not discriminate patients according to their physical functional ability in cross sectional assessment, and failed to detect

clinical improvement in an important subset of ambulatory patients with RA. Although the format of the MHAQ has the advantage of eliciting a 'satisfaction' score,⁸ limitations in its sensitivity in detecting clinical improvement in patients with relatively little difficulty in activities of daily living may not justify the use of this particular version of shorter questionnaire in certain clinical settings.

We thank Matthew H Liang for a critical review of the manuscript and Lawren Daltroy for helpful discussions. Dr G Stucki is a recipient of a fellowship of the Swiss Science National Foundation, and grants from the EULAR and the Swiss Associations of Physical Medicine, Rehabilitation and Rheumatology.

- 1 Pincus T, Summey J A, Soraci S A, Wallston K A, Hummon N P. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. *Arthritis Rheum* 1983; **26**: 1346-53.
- 2 Wolfe F, Pincus T. Standard self-report questionnaires in routine clinical and research practice—An opportunity for patients and rheumatologists. *J Rheumatol* 1991; **18**: 643-4.
- 3 McHorney C A, Ware J E Jr, Rogers W, Raczek A E, Lu J F R. The validity and relative precision of MOS short- and long form health status scales and Dartmouth COOP charts. Results from the medical outcomes study. *Med Care* 1992; **30**: MS253-65.
- 4 Katz J N, Larson M G, Phillips C B, Fossel A H, Liang M H. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; **30**: 917-25.
- 5 Ware J E Jr, Sherbourne C D. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992; **30**: 473-83.
- 6 Fries J F, Spitz P W, Kraines R G, Holman H R. Measurements of patient outcome in arthritis. *Arthritis Rheum* 1980; **23**: 137-45.
- 7 Ramey D R, Raynauld J P, Fries J F. The Health Assessment Questionnaire 1992. Status and review. *Arthritis Care Res* 1992; **5**: 119-29.
- 8 Ziebland S, Fitzpatrick R, Jenkinson C, Mowat A, Mowat A. Comparison of two approaches to measuring change in health status in rheumatoid arthritis: The Health Assessment Questionnaire (HAQ) and modified HAQ. *Ann Rheum Dis* 1992; **51**: 1202-5.
- 9 Bergner M, Bobbitt R A, Pollard W E, *et al.* The Sickness Impact Profile: validation of a health status measure. *Med Care* 1976; **14**: 57-67.
- 10 Jette A M, Davies A R, Cleary P D, *et al.* The Functional Status Questionnaire: Reliability and validity when used in primary care. *J Gen Intern Med* 1986; **1**: 143.
- 11 Wallston K A, Brown G K, Stein M J, Dobbins C J. Comparing the short and long versions of the Arthritis Impact Measurement Scales. *J Rheumatol* 1989; **16**: 1105-9.
- 12 Arnett F C, Edworthy S M, Bloch D A, *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; **31**: 315-24.
- 13 Brühlmann P, Stucki G, Michel B A. Evaluation of a German version of the physical dimension of the health assessment questionnaire in patients with rheumatoid arthritis. *J Rheumatol* 1994; **21**: 1245-9.
- 14 Stucki G, Schönbacher J, Brühlmann P, Mariacher S, Stoll T, Michel B A. Does a muscle strength index provide complementary information to traditional disease activity variables in patients with rheumatoid arthritis? *J Rheumatol* 1994; **21**: 2200-5.
- 15 van der Heijde D M F M, van't Hof M A, van Riel P L C M, van Leeuwen M A, van Rijswijk M H, van de Putte L B A. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. *Ann Rheum Dis* 1992; **51**: 177-81.
- 16 Mallya R K, Mace B E W. The assessment of disease activity in rheumatoid arthritis using a multivariate analysis. *Rheumatol Rehabil* 1981; **20**: 14-7.
- 17 Tukey J W. *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley publishing company, 1977.
- 18 Hochberg M C, Chang R W, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. *Arthritis Rheum* 1992; **35**: 498-502.
- 19 Blalock S J, Sauter S V H, DeVellis R F. The modified health assessment questionnaire difficulty scale. A health status measure revisited. *Arthritis Care Res* 1990; **3**: 182-8.
- 20 Liang M H, Larson M G, Cullen K E, Schwartz J A. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; **28**: 542-7.
- 21 Liang M H, Fossel A H, Larson M G. Comparison of five health status instruments for orthopedic evaluation. *Med Care* 1990; **28**: 632-42.
- 22 Hurst N P. An evaluation of the health assessment questionnaire (HAQ) in a long-term longitudinal follow-up of disability in rheumatoid arthritis [letter]. *Br J Rheumatol* 1993; **33**: 195.
- 23 Merbitz C, Morris J, Grip J C. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989; **70**: 308-12.