Validity and Responsiveness of the Dutch McMaster Toronto Arthritis Patient Preference Questionnaire (MACTAR) in Patients with Osteoarthritis of the Hip or Knee

DI-JANNE J.A. BARTEN, MARTIJN F. PISTERS, TIM TAKKEN, and CINDY VEENHOF

ABSTRACT. ***Objective.*** To determine the content validity, the construct validity, and the responsiveness of the Dutch McMaster Toronto Arthritis Patient Preference Questionnaire (MACTAR) in patients with osteoarthritis (OA) of the hip or knee.

Methods. The MACTAR comprises 2 parts: a transitional part and a status part. Content validity was investigated by comparing patient-elicited activities to items on the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the Medical Outcomes Study Short-Form 36 (SF-36). Construct validity was determined by correlating MACTAR outcomes with WOMAC/SF-36 outcomes. Responsiveness was investigated by correlating MACTAR, WOMAC, and SF-36 change scores with patient global assessment (PGA) scores and plotting a receiver-operating characteristics (ROC) curve.

Results. Eleven percent of the 894 impaired activities, identified by 192 patients, were not represented in either the WOMAC or the SF-36. The correlations (r_s) investigated for the MACTAR transitional part varied between 0.27 and –0.40; the status part correlated moderately with the general health scale of the SF-36 (r_s = 0.44). MACTAR change scores correlated better with PGA than with WOMAC/SF-36 change scores. The area under the ROC curve amounted to 0.90.

Conclusion. Our results suggest that the MACTAR exhibits moderate construct validity and good responsiveness in a population of patients with OA of the hip or knee. The MACTAR is potentially better able to detect changes over time in activities that are important to individual patients compared to other tools measuring physical function (WOMAC, SF-36). Clinicians could use the MACTAR to evaluate clinically relevant changes over time in patient-specific physical functioning. (First Release April 1 2012; J Rheumatol 2012;39:1064–73; doi:10.3899/jrheum.110876)

Key Indexing Terms:
OSTEOARTHRITIS HIP KNEE PHYSICAL FITNESS REPRODUCIBILITY OF RESULTS

Osteoarthritis (OA) is a common chronic musculoskeletal disorder[1], which can result in moderate to severe limitations in physical functioning[2]. Limited physical functioning can lead to a diminished quality of life[3,4,5]. OA treatment guidelines recommend exercise therapy to reduce impairments in physical function due to OA[6,7]. Exercise therapy can thus enable individuals to better meet the demands of daily living[8,9,10,11].

A number of tools are available to clinicians to evaluate the effect of exercise therapy on physical function. General, disease-specific, and patient-specific tools can be applied as either (self-reported) questionnaires or performance-based tests. A systematic review of the psychometric quality of both questionnaires and performance-based tests in patients with OA of the hip or knee has been published[12,13]. The reviews recommended the application of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[14], the Medical Outcomes Study Short-Form 36 (SF-36)[15,16,17], and multiactivity tests when evaluating physical function in patients with OA[12,13].

Standardized tools, applied to all patients in an identical manner, are recommended for the evaluation of physical functioning. Data produced by these tools may be conveniently and relatively easily categorized and compared between patients and across settings[18]. However, standardized tools are often difficult to interpret at the individual level and fail to take account of individual preferences and variation in the performance of particular activities[18]. Patient-specific tools measuring physical function have been developed based on

From the Netherlands Institute for Health Services Research (NIVEL); Program in Physiotherapy Sciences, Utrecht School of Clinical Health Sciences, University Medical Center Utrecht; and Child Development and Exercise Center, University Medical Center Utrecht, Utrecht, The Netherlands.

D-J.J.A. Barten, MSc, NIVEL, and Program in Physiotherapy Sciences, Utrecht School of Clinical Health Sciences, University Medical Center Utrecht; M.F. Pisters, PhD, Program in Physiotherapy Sciences, Utrecht School of Clinical Health Sciences, and Child Development and Exercise Center, University Medical Center Utrecht; T. Takken, PhD, Child Development and Exercise Center, University Medical Center Utrecht; C. Veenhof, PhD, NIVEL.

Address correspondence to D-J.J.A. Barten, Netherlands Institute for Health Services Research, PO Box 1568, 3500 BN Utrecht, The Netherlands. E-mail: d.barten@nivel.nl

Accepted for publication January 12, 2012.

the need for a more patient-centered approach as set out in healthcare policies and to enable clinicians to measure changes in activities that really matter to individual patients[19]. In contrast with standardized tools, patient-specific instruments can identify the relevant issues at an individual level and allow evaluation to focus on what is important to each individual patient[18]. Although the possibilities to compare statistical data between patients are minimal, the application of patient-specific tools may improve the validity and responsiveness for the assessment of physical function[18,19].

The McMaster Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) is one example of a patient-specific scale measuring physical function[20,21] (Appendix). The objective of the MACTAR is to identify individual disabilities due to the disease and their relative importance to the patient, complemented by questions on general health status[20]. The MACTAR has been described as a highly responsive and valid tool for the evaluation of physical function in patients with rheumatoid arthritis (RA)[21]. A recent psychometric evaluation of the questionnaire in patients with chronic lower back pain and patients with systemic sclerosis (SSc) showed moderate correlations with general and disease-specific tools that measure physical function[22,23].

To enable clinicians to use the MACTAR when evaluating physical function in patients with OA of the hip or knee, the psychometric properties of the questionnaire in this specific population must be determined. Therefore, our objective was to determine the content validity, construct validity, and responsiveness of the MACTAR in patients with OA of the hip or knee.

MATERIALS AND METHODS

Study design. Data reported in this study were collected from a cluster-randomized controlled trial of 200 patients with OA of the hip or knee over a 12-week period (maximum 18 sessions) that compared behavioral graded activity with usual care in accord with the Dutch physical therapy guidelines[24]. The content of the interventions has been described elsewhere[24]. The Medical Ethics Committee of the VU University Medical Center, Amsterdam, approved the study. For the purposes of this validation study, data on "physical function" were used, as well as descriptive data on the study population.

Study population. Participants were recruited between November 2001 and May 2003 through participating physical therapists and local newspapers. Dutch-speaking patients with OA of the hip or knee (based on the criteria of the American College of Rheumatology[25,26]) aged between 50 and 80 years who experienced diminished physical function were included in the study[24]. Participants who completed both baseline and followup (Week 13) measurements were eligible for inclusion in the present psychometric evaluation.

Measurements. Demographic and clinical data. Demographic and clinical data, including age, sex, duration of symptoms, OA location, and OA grade according to Kellgren and Lawrence[27], were collected from participating patients.

Physical function. Dutch MACTAR: The objective of this interview-based measurement tool is to evaluate changes in patient-specific physical function over time. It comprises 2 parts. The baseline interview starts with a transitional part. In this part, a trained interviewer asks the patient to identify up to 10 activities in which he/she experiences difficulties because of OA, such as activities in domestic care, professional life, and social interaction. The identified activities are ranked by the patient from 1 to 10 in order of importance: 1 for the activity the patient most wishes to be able to do without pain or discomfort due to OA, 2 for the next most important activity and so on. The top 5 prioritized activities are evaluated at followup. The second part of the MACTAR (status) collects information on health status. Perceived overall health, as well as psychological, emotional, and social well-being is measured by 5 questions (Likert-type rating scale); when a question obtains a less than optimal score, a followup question probes whether this is due to OA.

At the followup interview (Week 13), changes in physical function are investigated. Patients evaluate progress on their 5 most important activities as indicated in the transitional part of the baseline interview, by evaluating each activity as "less of a problem" (3 points), "the same" (2 points), or "more of a problem" (1 point). Patients also rate the perceived change in their OA on a 7-point Likert scale. The status part reassesses patients' health status.

It is quite difficult to allocate a total score to the MACTAR tool, because each part measures different domains. Moreover, the transitional part and the status part employ different scoring methods. While the transitional part measures change in physical function between baseline and followup, the status part investigates current health status, at both baseline and followup. The scoring method is presented in Table 1. Because of the differences between the transitional and status parts, scores were not added together, but presented separately. The MACTAR was translated into Dutch by Verhoeven, *et al* and validated in a population with RA[21].

Dutch WOMAC: The physical function subscale of the WOMAC contains 17 items that represent common activities in daily living[14,28]. Patients are asked how much difficulty they have had performing the activities mentioned. Each item is scored on a categorical scale, from "no difficulty" (score 0) to "extreme difficulty" (score 4). The total score varies from 0 (no difficulties) to 68 (extreme difficulties). Change scores on the WOMAC physical function subscale can vary between –68 (maximum improvement) to +68 (maximum deterioration). The WOMAC has been shown to be reliable and valid in patients with OA of the hip or knee[14,28] and the Dutch WOMAC permits valid international Dutch-English comparisons after differential item functioning[28].

Dutch SF-36: The SF-36 investigates quality of life[15,16,17,29]. It comprises 8 subscales, 3 of which were used in this validation study: physical functioning, role-physical, and general health. Scores on each subscale range from 0 to 100; higher scores reflect better health status. The SF-36 has been validated for patients with various diagnoses, including OA[30]; the Dutch language version has proved to be practical, reliable, and valid for use in general population surveys[29].

Self-perceived change: At the followup evaluation, self-perceived change in physical function was assessed by a patient global assessment (PGA) score. Patients were asked to rate their overall perception of improvement since the start of the intervention on a scale ranging from 1 (vastly deteriorated) to 8 (completely recovered). PGA scores provide reliable assessments of health transition in people with musculoskeletal disorders[31].

Statistical analyses. Descriptive statistics were applied to describe the study population. PGA ratings were dichotomized as "improved" (PGA score 5, 6, 7, or 8) versus "not improved" (PGA score 1, 2, 3, or 4). For continuous data, independent t tests were used to calculate differences at baseline between those patients who improved and those who did not. For categorical data, Mann-Whitney U tests were used to compare between groups.

Content validity. Content validity examines the extent to which the domain in question is comprehensively represented by the items in the questionnaire[32,33]. To determine whether the items in the MACTAR refer to relevant aspects of the construct and are relevant to the purpose of the instrument, the impaired activities mentioned by patients were compared with items on the WOMAC and the SF-36 physical functioning subscale[34].

Construct validity. There is currently no "gold standard" for attributes such as disability and functional status[35,36,37,38]. Therefore, construct validity rather than criterion validity was assessed. Construct validity refers to the extent to which scores on a particular instrument relate to other assessment tools in a manner that is consistent with theoretically derived hypotheses[39].

To investigate the construct validity of the MACTAR in patients with OA,

Table 1. Scoring method of the McMaster Toronto Arthritis Patient Preference Questionnaire.

	Transitional Part	Status Part
Baseline	—	Min: 5 (poor health status) Max: 25 (good health status)
Followup	Min: 6 (maximum deterioration) Max: 22 (maximum improvement)	Min: 5 (poor health status) Max: 25 (good health status)
Change score	The same as the followup score	Min: –20 (maximum deterioration) Max: +20 (maximum improvement)

change scores on the transitional part of the MACTAR were correlated with change scores on both the WOMAC and the SF-36 physical function subscales, as well as the SF-36 role-physical subscale. Further, followup scores on the status part of the MACTAR were correlated with followup scores on the SF-36 general health subscale. For normally distributed data, Pearson correlation coefficients (r) were used to express these correlations[40]. Spearman's rank correlation coefficients (r_s) were applied when data were not distributed normally. The following hypotheses were tested: (1) The change score on the physical function subscale of the WOMAC is negatively correlated ($r_s \leq -0.5$; $p < 0.05$)[41] with the change score on the transitional part of the MACTAR; the correlation was expected to be negative, because the WOMAC and the MACTAR use reverse scales. (2) Change scores on the physical functioning and role-physical subscales of the SF-36 are positively correlated ($r_s \geq 0.5$; $p < 0.05$)[41] with the change scores on the transitional part of the MACTAR. (3) Followup scores on the general health subscale of the SF-36 are positively correlated ($r_s \geq 0.4$; $p < 0.05$)[41] with the followup scores on the status part of the MACTAR.

Responsiveness. Responsiveness can be assessed in many different ways. However, one can distinguish 2 definition groups[42]. The first describes responsiveness as "the ability to detect clinically important change"[42,43]. In this group, an instrument is indicated as high-responsive if it is able to distinguish real change from measurement error. Responsiveness is calculated as the magnitude of a treatment effect in which the standardized response mean (SRM) and effect size could be very useful[42,43]. The second group defines responsiveness as "the ability to detect changes over time in the construct to be measured"[34,42]. In this case, responsiveness is independent from any treatment effect and is interpreted as longitudinal validity. It should be assessed in analogy to construct validity[34]. Therefore, predefined hypotheses concerning change scores on the MACTAR, WOMAC, and SF-36 in relation to PGA scores were tested. In the case of normally distributed change scores, parametric statistics were applied; nonparametric variants were applied for data that were not distributed normally. It was hypothesized that (1) the correlation between change scores on the MACTAR (transitional part) and the PGA will be better than that between change scores on the PGA and the WOMAC physical function subscale, the SF-36 physical functioning subscale, and the role-physical subscale, respectively ($p < 0.05$). (2) Change scores on the MACTAR (transitional part) for patients who have improved according to PGA will differ significantly ($p < 0.05$) from change scores for those who have not improved according to PGA.

Second, responsiveness was determined by plotting a receiver-operating characteristics (ROC) curve. The first step in this construction was to calculate sensitivity and specificity statistics for MACTAR change scores in patients identified as improved (PGA score > 4) and patients identified as nonimproved (PGA score ≤ 4). Next, the true-positive rate (sensitivity) was plotted in functions of the false-positive rate (1 – specificity) for different cutoff points. The best possible cutoff point would yield a point in the upper left corner of the ROC space, representing 100% sensitivity, 100% specificity, and an area under the curve (AUC) of 1.0. An instrument is indicated as highly responsive if AUC is > 0.90, moderately responsive where the AUC is between 0.70 and 0.90, and lowly responsive if the AUC is between 0.50 and 0.70[44].

All analyses were performed using PASW Statistics 18.0. If patients were unable to identify at least 5 impaired activities on the transitional part of the MACTAR, missing activity scores were filled with a score indicating a "no-change situation" (2 points); data from patients who mentioned fewer than 3 impaired activities were excluded from the responsiveness analyses. Further, in cases of just 1 missing followup item for the status part of the MACTAR, the score obtained on the equivalent question in the baseline interview was also used for the followup.

Following the initial analyses, a sensitivity analysis was performed on various cutoff points of the dichotomized PGA score, the aim of which was to determine whether the chosen cutoff point was the optimal point to dichotomize.

RESULTS

Study population. A total of 192 patients participated in both the baseline and the first followup assessment and were included for content and construct validity analyses. The median PGA score of these 148 women and 44 men was 5, representing "slightly improved." Baseline characteristics of the study population are presented in Table 2.

Outcomes. Table 3 shows absolute scores on the MACTAR, WOMAC, SF-36, and PGA at baseline and followup for both the total population and improved/nonimproved patients. At baseline, there were no differences on any of the outcome measures between patients who indicated that they had improved and patients who indicated that they had not improved. At followup, MACTAR scores (both transitional and status parts), WOMAC physical function scores, and SF-36 physical functioning scores differed significantly between improved and nonimproved patients. The measurement variation was higher in the WOMAC and SF-36 compared with the MACTAR, at both baseline and followup (Table 3).

Content validity. The study population (n = 192) identified a total of 894 impaired activities, a mean of 4.6 impaired activities per patient. Seventy-one patients (37%) were unable to identify at least 5 impaired activities; 1 patient could name only 1 impaired activity; 10 patients identified 2 impaired activities; and 33 patients were able to name a maximum of 3 impaired activities. Walking was most frequently mentioned as the most impaired activity (43%). Overall, 72% of the impaired activities that were identified comprised activities in the category of mobility. Table 4 summarizes all the activities mentioned, ranked by category.

All items from both the WOMAC and the SF-36 physical function subscales were represented in the impaired activities list based on the MACTAR questionnaire. However, 27% of the activities mentioned by patients during the MACTAR interview were not represented in the WOMAC, and 41%

Table 2. Baseline characteristics of total population and subgroups. Values are number (%), unless otherwise indicated.

Characteristic	Total Population, n = 192	Improved Group, n = 144†	Nonimproved Group, n = 48†	p
Female	148 (77)	118 (82)	30 (63)	0.01
Age, mean (± SD), yrs	64.7 (7.9)	65.5 (7.9)	62.3 (7.6)	0.02
Location of OA				0.58
Knee	126 (66)	96 (67)	30 (63)	
Hip	49 (26)	36 (25)	13 (27)	
Both	17 (9)	12 (8)	5 (10)	
Radiologic severity of OA*				0.62
No features/doubtful	44 (23)	35 (24)	9 (19)	
Minimal/moderate	91 (47)	67 (47)	24 (50)	
Severe or prosthesis	6 (3)	5 (4)	1 (2)	
Unknown	51 (27)	37 (26)	14 (29)	
Duration of symptoms, yrs				0.55
< 1	44 (23)	36 (25)	8 (17)	
1–5	71 (37)	52 (36)	19 (40)	
> 5	75 (39)	55 (38)	20 (42)	

* Kellgren-Lawrence score[27,47]. † Improved group: patient global assessment score (PGA) > 4; nonimproved group PGA ≤ 4. OA: osteoarthritis.

Table 3. Baseline, followup, and change scores on the outcome measures (mean ± SD).

Outcome Measure	Total Population, n = 192	Improved Group, n = 144	Nonimproved Group, n = 48	p
MACTAR transitional				
Followup	16.6 (2.8)	17.6 (2.0)	13.6 (2.6)	< 0.01
MACTAR status				
Baseline	19.3 (4.0)	19.3 (4.1)	19.1 (3.9)	0.76
Followup	20.5 (3.6)	20.8 (3.4)	19.4 (3.8)	0.03
Change score	1.2 (4.3)	1.5 (4.2)	0.1 (4.5)	0.05
WOMAC physical function				
Baseline	28.6 (11.0)	28.3 (10.9)	29.8 (11.3)	0.42
Followup	23.0 (11.4)	21.3 (10.6)	27.9 (12.1)	< 0.01
Change score	–5.6 (8.9)	–6.8 (9.0)	–1.9 (7.7)	< 0.01
SF-36 physical function				
Baseline	48.7 (20.1)	49.1 (19.5)	47.6 (22.1)	0.68
Followup	56.0 (21.7)	58.0 (21.2)	50.1 (22.1)	0.03
Change score	7.6 (17.6)	9.4 (18.5)	2.0 (12.9)	0.01
SF-36 role-physical subscale				
Baseline	42.5 (41.1)	41.2 (41.0)	46.1 (41.5)	0.48
Followup	55.9 (42.0)	59.1 (41.5)	46.1 (42.5)	0.07
Change score	14.7 (46.0)	19.2 (48.3)	1.1 (35.3)	0.02
SF-36 general health				
Baseline	50.8 (19.4)	51.8 (19.7)	47.9 (18.8)	0.24
Followup	48.9 (16.5)	50.0 (17.2)	45.7 (14.1)	0.13
Change score	–2.1 (17.3)	–2.1 (17.1)	–2.1 (17.9)	> 0.99
PGA, median (range)	5 (6)	6 (3)	4 (2)	—

MACTAR: McMaster Toronto Arthritis Patient Preference Questionnaire; WOMAC: Western Ontario McMaster Universities Osteoarthritis Index; SF-36: Medical Outcomes Study Short-Form 36; PGA: patient global assessment.

were not represented in the SF-36. Eleven percent of the impaired activities mentioned were not covered by items of the WOMAC or the SF-36: examples of these include gardening and activities related to professional life.

Construct validity. Correlations (r_s) between change scores on the transitional part of the MACTAR and change scores on the physical function subscales of the WOMAC and the SF-36 were –0.40 ($p < 0.01$) and 0.27 ($p < 0.01$), respectively.

Table 4. Patient-mentioned impaired activities on the MACTAR questionnaire.

Category Activity	Mentioned as Most Impaired Activity, n (%)	Mentioned in Total, n (%)
Housekeeping activities†*	4 (2.3)	70 (8.5)
Vacuum cleaning, mopping, washing windows or dishes, lifting buckets, etc.		
Leisure activities	19 (11.0)	102 (12.3)
Gardening	5 (2.9)	30 (3.6)
Remaining leisure activities (cultural activities, shopping†)	4 (2.3)	17 (2.2)
Sports (jogging, tennis, swimming, fitness, riding a horse, dancing)*	10 (5.8)	54 (6.5)
Mobility	134 (77.9)	591 (71.5)
Bicycling (including getting up/off)*	19 (11.0)	81 (9.8)
Climbing stairs†*	13 (7.6)	106 (12.8)
Driving (including getting in/out of car)†	4 (2.3)	52 (6.3)
Getting up from the floor/a chair, getting out of bed†	10 (5.8)	53 (6.4)
Inability to stand for long†	1 (0.6)	54 (6.5)
Kneeling down, bending over, reaching down†*	12 (7.0)	77 (9.3)
Remaining mobility activities	1 (0.6)	39 (4.7)
Walking†*	74 (43.0)	129 (15.6)
Professional activities	1 (0.6)	13 (1.6)
Remaining activities	1 (0.6)	2 (0.2)
Self-care activities	7 (4.1)	82 (9.9)
Dressing (socks, underwear, trousers)†	6 (3.5)	74 (9)
Remaining self-care activities*	1 (0.6)	8 (1.0)
Sexuality	1 (0.6)	5 (0.6)
Sleeping and resting, including turning around in bed†	5 (2.9)	18 (2.2)
Social roles	—	11 (1.3)
Total	172 (100)	894 (100)

† Item is represented in Western Ontario McMaster Universities Osteoarthritis Index. * Item is represented in Medical Outcomes Study Short-Form 36. MACTAR: McMaster Toronto Arthritis Patient Preference Questionnaire.

Change scores on the transitional part of the MACTAR and the role-physical subscale of the SF-36 were also moderately correlated ($r_s = 0.27$, $p < 0.01$). Spearman's r_s between followup score of the MACTAR status part and the general health subscale of the SF-36 was 0.44 ($p < 0.01$; Table 5).

Responsiveness. Data from 133 patients (82% women, mean age 64.0 ± 8.1 yrs) were used in the responsiveness analyses. Seventy-seven percent of these patients indicated that they had improved following treatment (PGA score > 4), while 23% reported that they had not improved (PGA score ≤ 4). With the exception of age, the improved and nonimproved groups had similar baseline characteristics. Absolute change scores on physical function outcomes are presented in Table 6. Change scores for patients who indicated that they had improved differed significantly from patients who indicated that they had not improved on all outcome measures (Table 6).

Correlations between change scores on the physical function outcomes and the PGA score are also presented in Table

Table 5. Correlation of change scores/followup score outcome measures (n = 189) (Spearman's rank correlation coefficient). All correlations were significant at the 0.01 level.

	MACTAR Transitional	WOMAC Physical Function	SF-36 Physical Functioning	SF-36 Role-physical	MACTAR Status	SF-36 General Health
Change scores						
MACTAR transitional	1.00	–0.40	0.27	0.27		
WOMAC physical function subscale		1.00	–0.36	–0.30		
SF-36 physical functioning subscale	1.00	0.32				
SF-36 role-physical subscale		1.00				
Followup scores						
MACTAR status					1.00	0.44
SF-36 general health subscale						1.00

MACTAR: McMaster Toronto Arthritis Patient Preference Questionnaire; WOMAC: Western Ontario McMaster Universities Osteoarthritis Index; SF-36: Medical Outcomes Study Short-Form 36.

Table 6. Change scores on physical function measures and correlation coefficients (r_s) with patient global assessment (PGA).

	Absolute Change Score, mean (95% CI)	Correlation with PGA Score**
MACTAR transitional*		
Total population, n = 133		0.69
Improved group, n = 102	17.7 (17.3 to 18.2)	
Nonimproved group, n = 31	13.7 (12.7 to 14.7)	
WOMAC physical function*		
Total population	–6.1 (–7.7 to –4.5)	–0.39
Improved group	–7.2 (–9.0 to –5.3)	
Nonimproved group	–2.7 (–5.8 to 0.4)	
SF-36 physical functioning*		
Total population	8.3 (5.2 to 11.3)	0.26
Improved group	9.7 (6.0 to 13.5)	
Nonimproved group	3.5 (–0.6 to 7.6)	
SF-36 role-physical*		
Total population	14.9 (6.8 to 22.9)	0.25
Improved group	19.7 (10.1 to 29.2)	
Nonimproved group	–0.9 (–14.2 to 12.4)	

* Significant differences between improved and nonimproved group ($p < 0.05$). ** Significant at $p < 0.05$, unless otherwise indicated. MACTAR: McMaster Toronto Arthritis Patient Preference Questionnaire; WOMAC: Western Ontario McMaster Universities Osteoarthritis Index; SF-36: Medical Outcomes Study Short-Form 36.

6. As hypothesized, change scores on the MACTAR correlate better with PGA (r_s = 0.69) than change scores on the WOMAC (r_s = –0.39) and SF-36 (r_s = 0.26 and 0.25, respectively; Table 6).

Figure 1 presents an ROC curve of the change scores for the MACTAR (transitional part), in which the sensitivity of the MACTAR amounted to its 1 – specificity. The AUC was 0.90 (95% CI 0.89–0.96) with a standard error of 0.03.

The sensitivity analysis showed that the cutoff point for the PGA score dichotomization (> 4) was chosen correctly. Higher and lower cutoff points resulted in less optimal responsiveness values.

DISCUSSION

Our aim was to investigate the content validity, construct validity, and responsiveness of the MACTAR in patients with OA of the hip or knee.

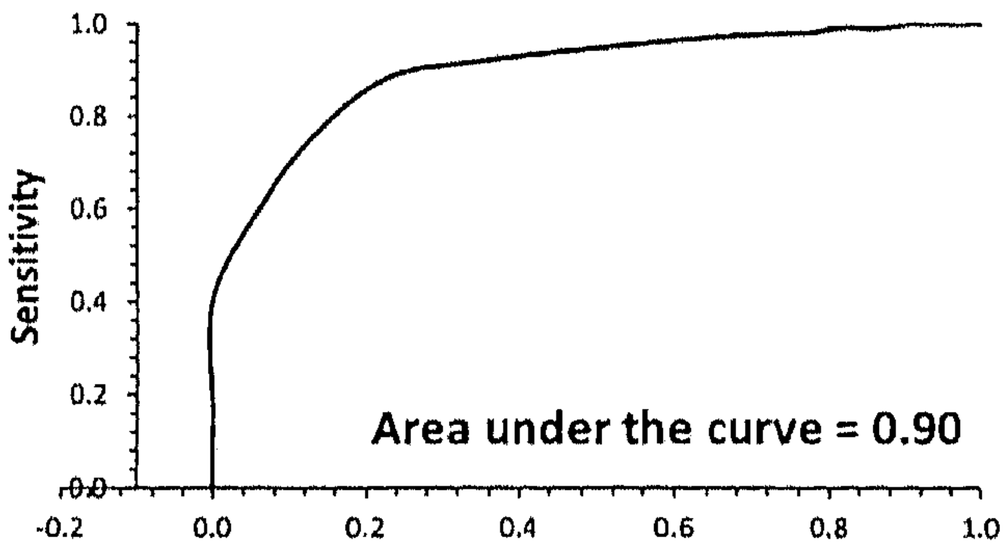

Figure 1. Receiver-operating characteristics curve of the change score on the McMaster Toronto Arthritis Patient Preference questionnaire (MACTAR; transitional part).

The content validity of the MACTAR seems to be good. Specifically, the majority of the impaired activities identified correlate with items on the WOMAC and/or SF-36, which also aim to assess physical function. However, the MACTAR fits better with the WOMAC questionnaire than with the SF-36. This is not surprising, since the WOMAC is aimed specifically at patients with OA, whereas the SF-36 has a more general purpose. Data for 11% of the activities are gathered only by the MACTAR and are not represented in either the WOMAC or the SF-36. These comprised activities in the areas of leisure, professional life, and social interaction. Indeed, participation in these fields varies widely among individuals. Disease-specific and general instruments do not take account of individual limitations, but patient-specific measures such as the MACTAR allow clinicians to evaluate physical functioning at the individual level.

The majority of the activities identified by the MACTAR questionnaire comprised activities in the mobility domain, which corresponds with the majority of activities in daily life. Recent validation studies on the MACTAR questionnaire in patients with chronic low back pain and RA showed comparable results[21,23]. The most frequently mentioned impaired activity in patients with chronic low back pain was taking part in sports activities[23]; in patients with hip/knee OA, walking was the most commonly cited impaired activity.

Although the content validity of the MACTAR seems to be good in patients with OA, the construct validity is less convincing. Moderate associations between the transitional part of the MACTAR and presumed comparable outcomes ($r_s \leq$

0.40) might be explained by an unbalanced distribution of impaired activities across the various activity categories. Specifically, the mobility category comprised almost 72% of all reported impaired activities, whereas the mobility domain in the WOMAC contains only 58% and in the SF-36, 60% of the total questionnaire. Thus, the transitional part of the MACTAR covers one specific part of the physical function domain extensively, whereas disease-specific and general tools account for a broader spectrum of this domain. Another explanation for the moderate construct validity could be the narrow variance around the mean on the MACTAR, compared with a wide variance in WOMAC and SF-36 scores. The variance is caused by patients who tend to assign the same disability score to very different impaired activities. The difference in variance impedes a comparison between a patient-specific instrument on the one hand and a disease-specific/generic instrument on the other.

As hypothesized, the status part of the MACTAR was moderately correlated with the general health subscale of the SF-36 ($r_s = 0.44$). Previous studies identified comparable correlation coefficients between the MACTAR and other physical function measures. Sanchez, *et al*[23] found a correlation (r_s) of 0.40 between the MACTAR and the Quebec Back Pain Disability Scale[37] in patients with chronic low back pain, and a correlation (r_s) of 0.38 ($p = 0.002$) was found between the MACTAR and the Health Assessment Questionnaire (HAQ)[45] in patients with SSc[22]. Verhoeven, *et al*[21] showed a correlation coefficient (r) of 0.73 ($p < 0.0003$) between the MACTAR and the HAQ in patients with RA.

The MACTAR was developed to evaluate patient-specific physical function over time. With this goal in mind, responsiveness is the most important psychometric property. For that reason, we evaluated the responsiveness of the questionnaire. As hypothesized, change scores for the MACTAR correlated better with the PGA score than change scores on the WOMAC and SF-36 do, leading to the conclusion that the MACTAR is better able to detect changes over time in patients with hip/knee OA than the WOMAC or SF-36. It has also been demonstrated that the MACTAR is capable of distinguishing patients who reported an improvement from those patients who reported no improvement. An AUC of 0.90 confirms the high responsiveness of the MACTAR in patients with hip/knee OA. Verhoeven, *et al*[21] also investigated the responsiveness of the MACTAR, concluding that it showed a high degree of responsiveness, based on an SRM of 3.5. However, an SRM is not an appropriate measure of assessing responsiveness as the ability to detect changes over time in the construct to be measured, but can be used to detect clinically important change[34,42].

One limitation of our study is the use of PGA scores as an external criterion to distinguish patients who improved from those who did not. Guyatt, *et al*[46] showed that patients are unduly influenced by their current health status when they complete transition ratings such as PGA scores. Moreover, the reproducibility of a single-item transitional scale is probably lower than that for a more extended measurement tool[47]. Finally, "a little better" is not, as a matter of course, equivalent to an important change[48]. However, better external criteria to discriminate between improved and nonimproved patients have not yet been elaborated.

Although the MACTAR appears to have some advantages over the WOMAC and the SF-36 in assessing physical function in individual patients, it also has some limitations. The need for a trained interviewer to apply the MACTAR, as well as its complicated scoring method, may reduce the likelihood that the MACTAR will become the instrument of first-choice in clinical practice. Further, patient-specific measures, including the MACTAR, do not take account of shifts in patient priorities that can occur over time in cases of change in disease status. Therefore, further studies should take account of the application of patient-specific measures at longterm followup.

Our results suggest that the MACTAR exhibits moderate construct validity and good responsiveness in a population of patients with OA of the hip or knee. Further, the MACTAR is potentially better able to detect changes over time in activities that are important to individual patients compared to other tools measuring physical function (WOMAC and SF-36). Therefore, clinicians could use the MACTAR to evaluate clinically relevant changes over time in patient-specific physical functioning.

REFERENCES

1. Poos M, Gommer A. Hoe vaak komt artrose voor en hoeveel mensen sterven eraan? In: Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid. Bilthoven, Netherlands: Dutch National Institute for Public Health and the Environment; 2009.
2. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799-802.
3. Dekker J, Boot B, van der Woude L, Bijlsma J. Pain and disability in osteoarthritis: A review of biobehavioural mechanisms. J Behav Med 1992;15:189-214.
4. Felson D, Lawrence R, Dieppe P, Hirsch R, Helmick C, Jordan J, et al. Osteoarthritis: new insights. Part 1. The disease and its risk factors. Ann Intern Med 2000;133:635-46.
5. McAlindon T, Cooper C, Kirwan J, Dieppe P. Determinants of disability in osteoarthritis of the knee. Ann Rheum Dis 1993;52:258-62.
6. Jordan K, Arden N, Doherty M, Bannwarth B, Bijlsma J. EULAR recommendations 2003: An evidence based approach to the management of knee osteoarthritis: Report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCICIT). Ann Rheum Dis 2003;62:1145-55.
7. Vogels E, Hendriks H, van Baar M, Dekker J, Hopman-Rock M, Oostendorp R, et al. KNGF-richtlijn artrose heup-knie. Amersfoort, Netherlands: Royal Dutch Society for Physical Therapy; 2001.
8. Brandt K. The importance of nonpharmacologic approaches in management of osteoarthritis. Am J Med 1998;105:39S–44s.
9. Fransen M, McConnell S. Exercise for osteoarthritis of the knee. Cochrane Database Sys Rev 2008;8:CD004376.
10. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Exercise for osteoarthritis of the hip. Cochrane Database Sys Rev 2009;8:CD007912.

11. Tan J, Horn S. Practical manual of physical medicine and rehabilitation: Diagnostics, therapeutics, and basis problems. 2nd ed. St. Louis: Elsevier; 2006.
12. Terwee C, Mokkink L, Steultjens M, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: A systematic review of measurement properties. Rheumatology 2006;45:890-902.
13. Veenhof C, Bijlsma J, van den Ende C, van Dijk G, Pisters M, Dekker J. Psychometric evaluation of osteoarthritis questionnaires: A systematic review of the literature. Arthritis Rheum 2006;55:480-92.
14. Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stitt L. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
15. McHorney C, Ware J, Lu R, Sherbourne C. The MOS 36-item short-form health survey (SF-36): III. Tests of data-quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40-66.
16. McHorney C, Tarlov A. Individual-patient monitoring in clinical practice: Are available health status surveys adequate? Qual Life Res 1995;4:293-307.
17. Ware J, Sherbourne C. The MOS 36-Item Short-Form Health Survey (SF-36). I: Conceptual framework and item selection. Med Care 1992;30:473-83.
18. Seror R, Tubach F, Baron G, Guillemin F, Ravaud P. Measure of function in rheumatoid arthritis: Individualized or classical scales? Ann Rheum Dis 2010;69:97-101.
19. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. J Clin Epidemiol 2005;58:791-801.
20. Tugwell P, Bombardier C, Buchanan W, Goldsmith C, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire — An individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol 1987;14:446-51.
21. Verhoeven A, Boers M, van der Linden S. Validity of the MACTAR Questionnaire as functional index in a rheumatoid arthritis clinical trial. J Rheumatol 2000;27:2801-9.
22. Mouthon L, Rannou F, Bérezné A, Pagnoux C, Guilpain P, Goldwasser F, et al. Patient preference disability questionnaire in systemic sclerosis: A cross-sectional survey. Arthritis Rheum 2008;59:968-73.
23. Sanchez K, Papelard A, Nguyen C, Jousse M, Rannou F, Revel M, et al. Patient-preference disability assessment for disabling chronic low back pain. Spine 2009;34:1052-9.
24. Veenhof C, Koke J, Dekker J, Oostendorp R, Bijlsma J, van Tulder M, et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee: A randomized clinical trial. Arthritis Rheum 2006;55:925-34.
25. Altman R, Asch E, Bloch DA, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039-49.
26. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505-14.
27. Kellgren J, Lawrence J. Radiological assessment of osteo-arthritis. Ann Rheum Dis 1957;15:494-502.
28. Roorda L, Jones C, Waltz M, Lankhorst G, Bouter L, van der Eijken J, et al. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. Ann Rheum Dis 2004;63:36-42.
29. Aaronson N, Muller M, Cohen P, Essink-Bot M, Fekkes M, Sandermann R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 2011;51:1055-68.
30. Kosinski M, Keller S, Hatoum H, Kong S, Ware JE Jr. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: Tests of data quality, scaling assumptions and score reliability. Med Care 1999;37:10-22.
31. Kamper S, Ostelo R, Knol DC, Maher C, de Vet H, Hancock M. Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 2010;63:760-6.
32. Guyatt G, Feeny D, Patrick D. Measuring health related quality of life. Ann Intern Med 1993;118:622-9.
33. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: Results of the COSMIN study. J Clin Epidemiol 2010;63:737-45.
34. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Med Res Methodol 2010;10:22.
35. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996;65:71-6.
36. Chatman AB, Hyams SP, Neel JM, Binkley JM, Stratford PW, Schomberg A, et al. The patient-specific functional scale: Measurement properties in patients with knee dysfunction. Phys Ther 1997;77:820-9.
37. Kopec J, Esdaille J, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement properties. Spine 1995;20:341-52.
38. Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: A report of a patient specific measure. Physiotherapy Canada 1995;47:258-63.
39. Kirshner B, Guyatt B. A methodological framework for assessing health indices. J Chronic Dis 1985;38:27-36.
40. Nunnaly JC. Introduction to psychological measurement. New York: McGraw-Hill; 1972.
41. Portney LG, Watkins MP. Foundations of clinical research. 2nd ed. Upper Saddle River, NJ: Prentice-Hall Inc.; 2000.
42. Terwee C, Dekker F, Wiersinga W, Prummel M, Bossuyt P. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. Qual Life Res 2003;12:349-62.
43. Streiner D, Norman G. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.
44. Hanley J, McNeil B. The meaning and use of the area under receiver operating characteristic (ROC) curve. Radiology 1982;143:29-39.
45. Fries J, Spitz P, Young D. Dimensions of health outcome: The Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:727-35.
46. Guyatt G, Norman G, Juniper E, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002;55:900-8.
47. Norman G, Stratford P, Regehr G. Methodological problems in retrospective computation of responsiveness to change: The lessons of Cronbach. J Clin Epidemiol 1997;50:869-79.
48. Terwee C, Roorda L, Dekker J, Bierma-Zeinstra S, Peat G, Jordan K, et al. Mind the MIC: Large variations among populations and methods. J Clin Epidemiol 2010;63:524-34.

APPENDIX. McMaster Toronto Arthritis Patient Preference Questionnaire (MACTAR). From J Rheumatol 1987;14:446-50, with permission.

Baseline Interview

Interviewers note:
Read the questions and response categories provided. Tick the response given.

1a. *How would you say your overall health has been during the last 2 weeks?*
You think of it as
- ☐3 *very good*
- ☐2 *pretty good*
- ☐1 *not too good*

2. *Osteoarthritis may cause restrictions in several areas of your daily life. For different people the impact of osteoarthritis is also different. We will ask you to name activities in which you experience difficulties because of your osteoarthritis. What matters here, is what your personal experience has been. Please, think of what became a problem, now that you have osteoarthritis.*

Interviewers note:
In order to elicit a comprehensive list of activities: First, give the patient opportunity to react spontaneously. Then read the probes. Record the exact phrases of the patient on the line hereunder.
To support you in naming any problems caused by osteoarthritis, I will read you a number of areas of your daily life that might be affected.
Does your osteoarthritis limit:
- *any (other) activities around the house such as cooking, housework etc.?*
- *any activity related to dressing such as buttoning, pulling a sweater over your head etc.?*
- *any (other) activities at your work (outside the home), drive a car or other transportation?*
- *any (other) leisure activities. Either sports such as bowling, swimming, golf, etc. or non-sports such as needlework, woodwork etc.?*
- *any (other) social activities. Such as visiting, playing cards, going to church, etc.?*
- *sexuality?*

Are there changes in the relationship with your family?
If you live together with a husband / wife / partner, are there changes in the relation with him / her?
If you have children living at home, are there changes in your relationship with them?

.....

[The line above is printed 10 times]

Interviewers note:
To rank the activities in order of importance to the patient, ask the following questions:

Which of these activities would you most like to be able to do without pain or discomfort of your osteoarthritis?
Show and read the list to the patient. Write '1' next to the activity the patient chose.

After (read activity 1), *which activity would you next most like to be able to do without pain or discomfort of your osteoarthritis?*
Show and read the list to the patient, with exception of the activity with priority 1.
Write '2' next to the activity the patient chose.
After (read activity 1 and 2), *which activity would you next most like to be able to do without pain or discomfort of your osteoarthritis?*
Show and read the list to the patient, with exception of the activity with priority 1 and 2.
Write '3' next to the activity the patient chose.
Continue like this, until all activities are ranked. The 5 with the highest priority will return in the follow-up interview.

3a. *In general, how satisfying do you find the way you spend your life?*
Over the last week you think of it as:
- ☐3 *completely satisfying →go to Q4a*
- ☐2 *pretty satisfying*
- ☐1 *not very satisfying*

3b. *Is your life not completely satisfying because of your osteoarthritis?*
☐0 yes ☐1 *no*

4a. *How would you say your overall physical functioning has been?*
Over the last week you think of it as:
- ☐5 *good* →go to Q5a
- ☐4 *fair to good*
- ☐3 *fair*
- ☐2 *fair to poor*
- ☐1 *poor*

4b. *Is your physical function not as good as it might be because of your osteoarthritis?* ☐0 yes ☐1 *no*

5a. *How would you say your overall social functioning has been? Over the last week you think of it as:*
☐5 *good* →*go to Q6a*
☐4 *fair to good*
☐3 *fair*
☐2 *fair to poor*
☐1 *poor*

5b. *Is your social functioning not as good as it might be because of your osteoarthritis?* ☐0 yes ☐1 *no*

6a. *How would you say your overall emotional functioning has been? Over the last week you think of it as:*
☐5 *good* →*stop here*
☐4 *fair to good*
☐3 *fair*
☐2 *fair to poor*
☐1 *poor*

6b. *Is your emotional functioning not as good as it might be because of your osteoarthritis?* ☐0 yes ☐1 *no*

Follow-up interview

1a. *How would you say your overall health has been during the last 2 weeks? You think of it as*
☐3 *very good*
☐2 *pretty good*
☐1 *not too good*

1b. *Have you noticed any change in your osteoarthritis since we talked during the first interview?*
☐ yes ☐ *no → 'no change' at Q1d, go to Q2a*

1c. *Please describe how your osteoarthritis has changed?*
.....

1d. *When you think of your osteoarthritis during the two weeks before the first interview, how much better or worse overall has your osteoarthritis become?*
☐7 *a great deal better*
☐6 *moderately better*
☐5 *slightly better*
☐4 *no change*
☐3 *slightly worse*
☐2 *moderately worse*
☐1 *a great deal worse*

You may remember the first time we spoke. You told me which activities were at that time problems due to your arthritis. I will ask you again about the five most important.

2a. *Since the first interview, have you noticed any change in your ability to* (activity 1)?
☐3 *less of a problem*
☐2 *the same*
☐1 *more of a problem*

2b. *Since the first interview, have you noticed any change in your ability to* (activity 2)?
☐3 *less of a problem*
☐2 *the same*
☐1 *more of a problem*

2c. *Since the first interview, have you noticed any change in your ability to* (activity 3)?
☐3 *less of a problem*
☐2 *the same*
☐1 *more of a problem*

2d. *Since the first interview, have you noticed any change in your ability to* (activity 4)?
☐3 *less of a problem*
☐2 *the same*
☐1 *more of a problem*

2e. *Since the first interview, have you noticed any change in your ability to* (activity 5)?
☐3 *less of a problem*
☐2 *the same*
☐1 *more of a problem*

Questions 3a to 6b are equal to the equally numbered questions in the baseline interview.