Psychometric Properties of the Health Assessment Questionnaire Disability Index (HAQ-DI) and the Modified Health Assessment Questionnaire (MHAQ) in Patients with Knee Osteoarthritis

Sağlık Değerlendirme Anketi Özürlülük İndeksinin (SDA-Öİ) ve Modifiye Sağlık Değerlendirme Anketinin (MSDA) Diz Osteoartritli Hastalarda Psikometrik Özellikleri

Serdal Kenan Köse[1], Derya Öztuna[1], Şehim Kutlay[2], Atilla Halil Elhan[1], Alan Tennant[3], Ayşe Adile Küçükdeveci[2]

[1]Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara, Turkey
[2]Ankara Üniversitesi Tıp Fakültesi, Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı, Ankara, Turkey
[3]University of Leeds, Academic Unit of Musculoskeletal Disease, Leeds, England

Abstract

Objective: To investigate the psychometric properties of the Health Assessment Questionnaire Disability Index (HAQ-DI) and the modified HAQ (MHAQ) in patients with knee osteoarthritis (OA).

Materials and Methods: The internal construct validity of the HAQ-DI and MHAQ were assessed by Rasch analysis and external construct validity by associations with the Western Ontario and McMaster Universities Index of Osteoarthritis Index (WOMAC), the World Health Organization Disability Assessment Schedule (WHODAS-II) and the Nottingham Health Profile (NHP). Reliability was tested by internal consistency and person separation index.

Results: Two hundred and fifteen outpatients with knee OA (mean age±standard deviation (SD) 57.7±10.9 years; 81% female) filled in the assessment scales including HAQ-DI, WOMAC, WHODAS-II and the NHP. MHAQ was not administered as a separate measure but scored by using the HAQ-DI forms. Both the HAQ-DI and the MHAQ data satisfied Rasch model expectations with a mean item fit of 0.096 (SD 1.186) and -0.312 (SD 1.063), and person fit of 0.307 (SD 0.895) and -0.329 (SD 0.879), respectively. Both scales were unidimensional and showed no differential item functioning. The reliabilities of both scales were good with high Cronbach's alpha and PSI levels above 0.85. However neither of them was particularly well targeted to the current population who displayed a level of disability much below the average difficulty level of the scales. External construct validity was confirmed by expected correlations with WOMAC, WHODAS-II and NHP. Although the distribution of both scales was right skewed, the floor effect was more prominent in MHAQ.

Conclusion: Both the HAQ-DI and MHAQ are found to be reliable and valid to assess physical disability in patients with knee OA. However, the possible floor effect in this diagnostic group should be kept in mind.

(Turk J Rheumatol 2010; 25: 147-55)

Key words: Knee osteoarthritis, HAQ-DI, MHAQ, Rasch analysis, validity and reliability

Özet

Amaç: Diz osteoartritli hastalarda, Sağlık Değerlendirme Anketi Özürlülük İndeksinin (SDA-Öİ) ve Modifiye Sağlık Değerlendirme Anketinin (MSDA) psikometrik özelliklerini incelemektir.

Yöntem ve Gereçler: SDA-Öİ ve MSDA'nın iç yapısal geçerliliği Rasch analizi ile dış yapısal geçerliliği ise WOMAC (Western Ontario and McMaster Universities) Osteoartrit İndeksi, Dünya Sağlık Örgütü (DSÖ) Yetiyitimi Değerlendirme Çizelgesi ve Nottingham Sağlık Profili ile ilişkiler incelenerek değerlendirilmiştir. Güvenirlik, içsel tutarlılık ve birey ayırsama indeksi kullanılarak belirlenmiştir.

Bulgular: Diz osteoartritli 215 hasta (yaş ortalaması±standart sapma (SS): 57.7±10.9 yıl; %81 kadın), SDA-Öİ, WOMAC Osteoartrit İndeksi, DSÖ Yetiyitimi Değerlendirme Çizelgesi ve Nottingham Sağlık Profili değerlendirme ölçeklerindeki maddelere yanıt vermişlerdir. MSDA, ayrı bir ölçek olarak uygulanmayıp, SDA-Öİ formlarına verilen yanıtlardan puanlanmıştır. Hem SDA-Öİ hem de MSDA sırasıyla ortalama madde uyumu 0.096 (SS: 1.186) ve -0.312 (SS: 1.063); ortalama birey uyumu 0.307 (SS: 0.895) ve -0.329 (SS: 0.879) değerleri ile Rasch modeline uyum göstermiştir. Her iki ölçeğin de tek boyutlu olduğu ve madde işlev farklılığı göstermediği belirlenmiştir. Ölçeklerin güvenirlikleri 0.85'in üzerindeki Cronbach'ın alfa katsayısı ve birey ayırsama indeksi değerleri ile yüksek bulunmuştur. Mevcut popülasyonun özürlülük düzeyi, SDA-Öİ ve MSDA'nın zorluk düzeylerinin ortalamasından oldukça düşük olduğu için, ölçeklerdeki maddelerin zorluk düzeylerinin dağılımı ile popülasyondaki hastaların özürlülük düzeylerinin dağılımı arasındaki uyum yeterli bulunmamıştır. Her iki ölçeğin de, WOMAC Osteoartrit İndeksi, DSÖ Yetiyitimi Değerlendirme Çizelgesi ve Nottingham Sağlık Profili ile beklenen düzeyde ilişkiler göstermesi, dış yapısal geçerliliklerini doğrulamıştır. Her iki ölçeğin de dağılımlarının sağdan çarpık olmasına rağmen, taban etkisinin MSDA'da daha belirgin olduğu gözlenmiştir.

Sonuç: SDA-Öİ ve MSDA, diz osteoartritli hastalarda fiziksel özürlülüğün değerlendirilmesinde güvenilir ve geçerli olarak bulunmuştur. Ancak, bu tanı grubundaki olası taban etkisinin göz önünde bulundurulması gerekmektedir.

(Turk J Rheumatol 2010; 25: 147-55)

Anahtar sözcükler: Diz osteoartriti, SDA-Öİ, MSDA, Rasch analizi, geçerlik ve güvenirlik

148 Köse et al.
HAQ-DI and MHAQ in Knee OA

Turk J Rheumatol 2010; 25: 147–55

Introduction

The Health Assessment Questionnaire Disability Index (HAQ-DI) is the most widely used self-report questionnaire to assess functional status of patients with arthritis. It was introduced in the 1980s in rheumatoid arthritis (1) and has been applied to other diseases, including osteoarthritis (OA), juvenile rheumatoid arthritis, systemic lupus erythematosus, scleroderma, ankylosing spondylitis, fibromyalgia, and psoriatic arthritis (2). It is a 20-item questionnaire addressing difficulty in eight domains: dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities. It was adapted to various languages and some investigators argue that it can be considered a generic instrument (3). However it has not been validated for all the conditions in which it is applied, for example, the psychometric properties of HAQ-DI in OA have not been extensively investigated.

The modified Health Assessment Questionnaire (MHAQ) was developed by Pincus et. al. from the original HAQ-DI by reducing the questionnaire from 20 to 8 questions retaining one question from each of the eight domains, and supplemented the original questions assessing level of difficulty with additional questions assessing patient satisfaction regarding the same activities of daily living (4). Thus the MHAQ is shorter than the original, and easier to score compared with the HAQ-DI. However it has been reported to be less sensitive to change in rheumatoid arthritis (5, 6). Although the MHAQ has been used in patients with OA (7), its validity and reliability has not yet been reported. Therefore the aim of the current study was to investigate the psychometric properties of both HAQ-DI and MHAQ in patients with knee osteoarthritis.

Materials and Methods

Patients and setting

Data was collected in the Department of Physical Medicine and Rehabilitation at the Medical Faculty of Ankara University, Turkey. A total of 215 outpatients diagnosed as knee OA according to the American College of Rheumatology criteria for the classification and reporting OA of knee were included in the study (8). Patients with concomitant uncontrolled or severe systemic diseases that might affect their health status were excluded. The study was approved by the Ethical Committee of the Faculty of Medicine, Ankara University. All patients gave informed consent and the study was carried out in compliance with Helsinki Declaration.

Outcome measures

The assessment included the administration of the HAQ-DI, MHAQ, the Western Ontario and McMaster Universities Index of Osteoarthritis (WOMAC), the World Health Organization Disability Assessment Schedule II (WHODAS-II) and the Nottingham Health Profile (NHP).

The HAQ-DI contains 20 questions classified into eight domains (items): dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities. There are four possible responses for each question: without any difficulty (0), with some difficulty (1), with much difficulty (2), unable to do (3). The highest score reported by the patient for any component question of each domain determines the score for that domain unless aids or devices are required. In case the need of aids or devices the score is automatically raised to 2 when it is rated as 0 or 1. Then the HAQ-DI score is calculated as the average of 8 domains (items) scores ranging between 0 and 3, higher score showing more disability. The Turkish adaptation was used in the study (9).

The MHAQ is a subset of 8 questions taken from the 8 domains of the original HAQ-DI. It is scored by taking the average of the 8 question scores, with a range of 0-3. Scoring principle of each question is similar to HAQ-DI, except that MHAQ does not consider aids or devices in the scoring process. In the present study, "level of difficulty" was assessed for the MHAQ. The MHAQ was not administered as a separate questionnaire but was scored from the HAQ-DI.

The WOMAC is a disease-specific index developed for OA of the knee or hip (10). It consists of 24 items in three domains: pain (5 items), stiffness (2 items), and physical function (17 items). There are five response options for every question ('0' none, '1' mild, '2' moderate, '3' severe and '4' extreme) in Likert form. The maximum score is 20 for pain, 8 for stiffness, 68 for physical function and 96 for the total WOMAC. Higher scores indicate more or worse symptoms, maximal limitations, and poor health. The Turkish version of the WOMAC (version 3.1) scale was used in this study (11).

The WHODAS-II is a generic, multidimensional disability questionnaire that includes 36 items in six life domains: understanding and communicating (6 items), getting around (5 items), self-care (4 items), getting along with people (5 items), life activities (8 items), and participation in society (8 items) (12). It employs a five-point rating scale on all items in which '1' indicates no difficulty and '5' indicates extreme difficulty or inability to perform the activity. Raw scores are transformed into standardized scores. Total score and subscale scores range between 0 and 100, with higher scores reflecting greater disability. The adapted Turkish version of the WHODAS-II instrument was used (13).

The NHP is a generic health status measure developed to record the perceived distress of patients in physical, emotional, and social domains (14). It comprises 38 statements (answered 'yes' or 'no') that form six sections: physical mobility (8 items), pain (8 items), sleep (5 items), emotional reactions (9 items), social isolation (5 items), and energy level (3 items). The score on each section of the NHP is the percentage of items affirmed by the respondent (i.e., the number of 'yes' responses multiplied by 100 and divided by the number of items in that

Turk J Rheumatol 2010; 25: 147-55

Köse et al.
HAQ-DI and MHAQ in Knee OA 149

section). Possible scores could range from 0 to 100, with a higher score indicating greater distress. The Turkish version of NHP was used (15).

Internal construct validity

Internal construct validity of HAQ-DI and MHAQ was assessed by Rasch Analysis. Rasch analysis is the formal testing of an assessment or an outcome measure against a mathematical measurement model which defines how interval scale measurement can be derived from ordinal questionnaires (16,17). The Rasch model assumes that the probability of a given respondent affirming an item is a logistic function of the relative distance between the item difficulty and the person ability on a linear scale. The model estimates person ability independent of the distribution of the population, and item difficulty independent of the person ability (18). These are requirements for obtaining interval scale estimates (19). Master's partial credit model (PCM) which is an extension of the Rasch dichotomous model for polytomous (more than two response categories) items was used in this study (20).

Common fundamental aspects to the Rasch model were assessed (21). These are 1) the appropriate ordering of response categories and any necessary rescoring for where polytomous items; 2) fit of items and persons to the model; 3) test of the assumption of the local independence of items, including response dependency and unidimensionality; 4) the presence of Differential Item Functioning (DIF).

Before evaluation of item fit, where polytomous items are involved, the response categories should be examined for correct ordering. For an item with an appropriate ordering of thresholds, each response option would demonstrate the highest probability of endorsement at a specific range of the scale, with successive thresholds found at increasing levels of the construct being measured. The respondents' inconsistent use of response options result in disordered thresholds and usually, in these circumstances, the collapsing of categories improves overall fit to the model (22).

A range of fit statistics is used to test if the data conform to Rasch model expectations. Two are item–person interaction statistics transformed to approximate a z score, representing a standardized normal distribution. If the items and persons fit the model, we would expect to see a mean of approximately zero and a standard deviation (SD) of one. The third is a summed chi-square within groups defined by their position on the trait, where the overall chi-square for items is summed to give the item trait interaction statistic, testing the property of invariance across the trait. A significant chi-square indicates that the hierarchical ordering of the items varies across the trait, so compromising the required property of invariance. In addition to these overall summary fit statistics, individual person- and item-fit statistics are presented, as (a) residuals (a summation of individual

person and item deviations), (b) as a chi-square statistic, and (c) as an analysis of variance (ANOVA) with the residuals summed across the main effects of class intervals. Fit residuals between ±2.5 are deemed to be adequate. These are summated within ability groups to provide the basis of the ANOVA analysis.

A formal test of the assumption of unidimensionality is undertaken by performing a, principal component analysis (PCA) of the residuals. Items with the highest positive and negative correlations on the first residual factor are used to construct two smaller scales, anchored to the item difficulties of the main analysis (23). The person estimates derived from these two subsets of items are contrasted for each individual by a t test. A significant difference would be expected to occur by chance in 5% of the cases. Consequently, the percentage of tests outside the range ±1.96 is reported, together with a 95% binomial confidence interval. This interval should overlap 5% for a non-significant finding to confirm unidimensionality.

The assumption of local independence implies that when the 'Rasch factor' has been extracted, that is, the main scale, there should be no leftover patterns in the residuals. This assumption was tested by performing a PCA analysis of the residuals obtained from PCM. If a pair of items had a residual correlation of 0.30 or more, one of the items that showed a higher accumulated residual correlation with the remaining items was eliminated (24).

Items are also tested for DIF. In the framework of Rasch measurement, the scale should be free of item bias or DIF (25). DIF occurs when different groups within the sample (e.g., males and females), despite equal levels of the underlying characteristic being measured, respond in a different manner to an individual item. For example, men and women with equal levels of disability may respond systematically differently to a self-care item such as getting dressed. DIF can be detected both statistically and graphically. In the current analysis, DIF was tested by age, gender and duration of disease.

Reliability

Reliability of HAQ-DI and MHAQ was initially tested by internal consistency which is an estimate of the degree to which its constituent items are interrelated, and is assessed by Cronbach's α (26). Subsequently reliability was further tested by the person separation index (PSI) from the Rasch analysis. This is equivalent to Cronbach's α but has the linear transformation from the Rasch model substituted for the ordinal raw score (27). Usually a reliability of 0.70 is required for analysis at the group level, and values of 0.85 and higher for individual use (28).

External construct validity

External construct validity was determined by testing for expected associations of HAQ-DI and MHAQ with WOMAC, WHODAS-II and NHP through the process of

convergent construct validity (29). In this study, the degree of associations was analyzed by Spearman's correlation coefficient.

Sample size and statistical software

For the Rasch analysis, a sample size of 215 patients will estimate item difficulty, with α of 0.05, to within ±0.27 logits (30). With an operational range of 3 logits for the scale this degree of precision would represent approximately half of a standard deviation, or with a 6 logit range, approximately one quarter of a standard deviation (31). This sample size is also sufficient to test for DIF where, at α of 0.05 a difference of 0.25 within the residuals can be detected for any 2 groups with β of 0.20. Bonferroni correction was applied to both fit and DIF statistics due to the multiple testing (32).

Results

Patient characteristics

The mean age of the 215 patients was 57.7 years (SD: 10.9), 81% were women, and the mean disease duration was 6.07 years (median: 4, range: 1 month-40 years). The scores of patients on HAQ-DI, MHAQ, WOMAC, WHODAS-II and NHP were shown in Table 1. Patients' pain levels were medium to high according to the assessment on WOMAC-Pain and NHP-Pain subscales. They were expressing a medium level of physical functioning rated by a disease-specific measure, WOMAC. Physical mobility of the patient sample presented by both WHODAS-II Getting around subscale and NHP-Physical Mobility section was also at the medium level.

Internal Construct Validity
HAQ-DI

Starting with 8 items, only "grip" item displayed disordered thresholds, necessitating collapsing of response categories. Following this, all items were found to fit the model (given a Bonferroni adjustment fit level of 0.006) (Table 2). Overall mean item fit residual was 0.096 (SD 1.186) and mean person fit residual was -0.307 (SD 0.895). Item trait interaction was non-significant, supporting the invariance of items (chi-square 26.50 (df=16), p=0.047). The PSI (reliability) was good (0.91) indicating the ability of the scale to differentiate more than 4 groups of patients (27). However, with a mean person location of -1.511, the scale was not particularly well targeted to the current population, who displayed a level of disability much below the average difficulty level of the scale (i.e. zero logits) (Figure 1). DIF was tested for age, gender and duration of disease, but all items were free of DIF.

Finally, using the PCA of residuals obtained from PCM, taking the highest positively and negatively correlated items to the first residual factor to make two subsets, no significant difference in person estimates (t=5.6%; CI 2.6%-8.7%) was found between the two subsets, thus supporting the unidimensionality of the 8-item HAQ-DI.

Table 1. Scores of patients on outcome measures		
Scales / subscales (n) (score range)	Mean±SD	Median (Min-Max)
HAQ-DI (n=214) (0-3)	0.94±0.68	0.88 (0-3)
MHAQ (n=214) (0-3)	0.55±0.52	0.38 (0-2.75)
WOMAC-Pain (n=215) (0-20)	10.16±5.02	10 (0-20)
WOMAC-Stiffness (n=215) (0-8)	3.39±2.44	3 (0-8)
WOMAC-Physical function (n=215) (0-68)	32.55±16.20	31 (3-64)
WOMAC-Total (n=215) (0-96)	46.11±22.38	43 (4-92)
WHODAS-II Understanding and communicating (n=212) (0-100)	16.84±18.36	10 (0-70)
WHODAS-II Getting around (n=212) (0-100)	36.26±21.18	40 (0-80)
WHODAS-II Self care (n=212) (0-100)	13.14±15.62	5 (0-65)
WHODAS-II Getting along with people (n=212) (0-100)	7.24±11.09	0 (0-60)
WHODAS-II Life activities (Work items removed) (n=197) (0-100)	33.83±26.26	30 (0-80)
WHODAS-II Participation in society (n=212) (0-100)	20.57±13.81	20 (0-52.5)
WHODAS-II Total (Work items removed (n=212) (0-100)	21.13±13.67	20 (0-58.1)
NHP-Energy (n=212) (0-100)	69.03±35.84	66.67 (0-100)
NHP-Pain (n=212) (0-100)	72.70±27.02	75 (0-100)
NHP-Emotional reactions (n=212) (0-100)	47.90±35.03	44.44 (0-100)
NHP-Sleep (n=212) (0-100)	45.66±33.98	40 (0-100)
NHP-Social isolation (n=212) (0-100)	29.43±32.20	20 (0-100)
NHP-Physical mobility (n=212) (0-100)	50.59±20.21	50 (0-100)

Turk J Rheumatol 2010; 25: 147–55

Köse et al.
HAQ-DI and MHAQ in Knee OA 151

When the assumption of local independence was examined, there was no pair of items which had a residual correlation of 0.15 or more.

MHAQ

Starting with 8 items, only "lift a full cup or glass to your mouth" item displayed disordered thresholds, necessitating collapsing of categories. Following this, all items were found to fit the model (given a Bonferroni adjustment fit level of 0.006) (Table 3). Overall mean item fit residual was -0.312 (SD 1.063) and mean person fit

Figure 1. Person-item threshold map of HAQ-DI

residual was -0.329 (SD 0.879). Item trait interaction was non-significant, supporting the invariance of items (chi-square 42.86 (df=40), p=0.349). The PSI was good (0.88) indicating the ability of the scale to differentiate more than 4 groups of patients (27). Overall, with a mean person score of -3.570, the scale was poorly targeted with patients displaying a significantly lower average level of disability than the average of the scale (Figure 2). DIF was tested for age, gender and duration of disease, but all the items were free of DIF.

Finally, using the PCA of residuals obtained from PCM, taking the highest positively and negatively correlated items to the first residual factor to make two subsets, no significant difference in person estimates (t=4.6%; CI 1.3%-7.8%) was found between the two subsets, thus supporting the unidimensionality of the MHAQ. When the assumption of local independence was examined, there was no pair of items which had a residual correlation of 0.15 or more.

Reliability

Reliabilities of both the HAQ-DI and MHAQ were good, with Cronbach's alpha of 0.95 and 0.87, and PSI of 0.91 and 0.88, respectively.
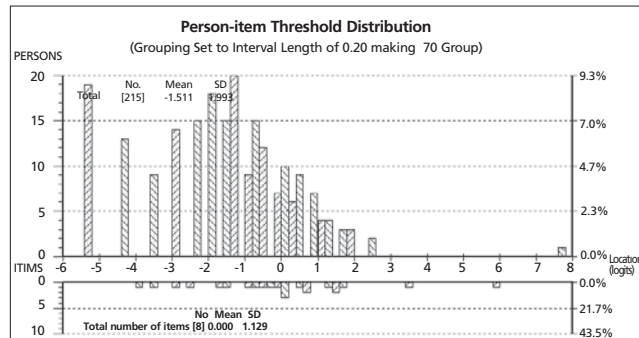
Table 2. Fit of the HAQ-DI item bank to partial credit model

Item	Location	SE	Individual Item Fit Residual	Chi-Square Test Statistics	p
Dressing&Grooming	0.262	0.110	-0.666	1.702	0.427
Arising	1.313	0.135	-1.529	5.169	0.075
Eating	1.124	0.129	1.469	0.752	0.687
Walking	0.083	0.128	-0.279	1.396	0.497
Hygiene	-1.406	0.112	2.123	6.565	0.038
Reach	-0.732	0.099	-0.244	2.454	0.293
Grip	0.931	0.151	-0.458	1.139	0.566
Activities	-1.573	0.109	0.356	7.322	0.026

SE: Standard error

Table 3. Fit of the MHAQ item bank to partial credit model

Item	Location	SE	Individual Item Fit Residual	Chi-Square Test Statistics	p
Dress yourself, including shoelaces and buttons	-1.522	0.122	0.496	2.655	0.753
Get in and out of bed	-0.143	0.143	-2.178	14.953	0.011
Lift a full cup or glass to your mouth	3.264	0.253	-0.154	3.537	0.618
Walk outdoors on flat ground	-0.085	0.138	1.161	4.847	0.435
Wash and dry your body	-0.58	0.143	-1.076	5.318	0.378
Bend down to pick up clothing from the floor	-1.748	0.125	0.399	1.923	0.860
Turn faucets on and off	2.412	0.210	-0.155	1.849	0.870
Get in and out of a car	-1.598	0.135	-0.99	7.782	0.169

SE: Standard error

152 Köse et al.
HAQ-DI and MHAQ in Knee OA

Turk J Rheumatol 2010; 25: 147–55

Distributional characteristics of the HAQ-DI, MHAQ

The floor effect of the HAQ-DI was 9% (score of 0) and 19% for the MHAQ. Although the distribution of both scales was right skewed this was more prominent in the MHAQ (Figure 3a, 3b). The percentages of patients scoring between 0-1, >1-2 and >2-3 were 62%, 30%, 8% in HAQ whereas 83%, 16%, 1% in MHAQ, respectively. To compare with an OA-specific scale, the distribution of WOMAC-Physical function scale was almost normal (Figure 3c).

External construct validity

Correlations of HAQ-DI and MHAQ scores with the WHODAS-II, NHP and WOMAC are presented in Table 4. As only 16 patients responded to the work items of WHODAS-II, the "life activities" subscale score and the

Figure 2. Person-item threshold map of MHAQ

total WHODAS-II score were calculated by excluding the work items. Correlations of both scales with the other 3 measures were similar and, as expected, showed the highest correlation with WOMAC-Physical function scale (Table 4).

Discussion

The HAQ is one of the most widely used measures of physical functioning in arthritis, and is recommended by the American College of Rheumatology for measuring physical functioning (33). Since it was first introduced, various short forms including the MHAQ have followed, and most recently some attempt has been made to provide an exchange rate for scores between the different versions (34). While it is used predominately in patients with rheumatoid arthritis, it is also widely used in other rheumatic conditions such as OA.

The present study investigates the psychometric properties of the HAQ-DI and MHAQ in patients with knee OA. Both scales were found to have high reliability with Cronbach's alpha of 0.95 and 0.87, and PSI of 0.91 and 0.88 for the HAQ-DI and MHAQ, respectively. These values are in concordance with reliability levels reported in RA patients before (9, 35). Internal construct validity of both scales was found to be adequate by fit of the data to the Rasch measurement model. Both scales were strictly unidimensional and showed no DIF. However, there are some concerns about the targeting of both
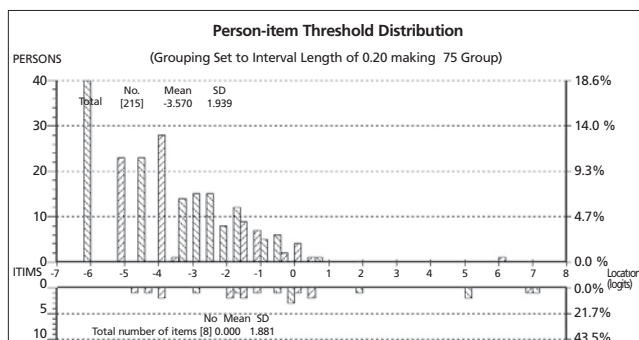
Table 4. Correlations of HAQ-DI and MHAQ with WOMAC, WHODAS-II and NHP		
	HAQ-DI	**MHAQ**
WOMAC-Pain (n=214)	0,556*	0,566*
WOMAC-Stiffness (n=214)	0,563*	0,567*
WOMAC-Physical Function (n=214)	0,720*	0,709*
WOMAC-Total (n=214)	0,709*	0,701*
WHODAS-II Understanding and communicating (n=211)	0,429*	0,452*
WHODA- II Getting around (n=211)	0,604*	0,632*
WHODAS-II Self care (n=211)	0,635*	0,664*
WHODAS-II Getting along with people (n=211)	0,301*	0,318*
WHODAS-II Life activities (Work items removed) (n=196)	0,543*	0,587*
WHODAS-II Participation in society (n=211)	0,597*	0,587*
WHODAS-II Total (Work related items removed-32 items) (n=211)	0,685*	0,703*
NHP-Energy (n=211)	0,470*	0,482*
NHP-Pain (n=211)	0,536*	0,521*
NHP-Emotional Reactions (n=211)	0,446*	0,414*
NHP-Sleep (n=211)	0,326*	0,345*
NHP-Social Isolation (n=211)	0,356*	0,362*
NHP-Physical Mobility (n=211)	0,559*	0,575*
* p<0.001		

Turk J Rheumatol 2010; 25: 147–55

Köse et al.
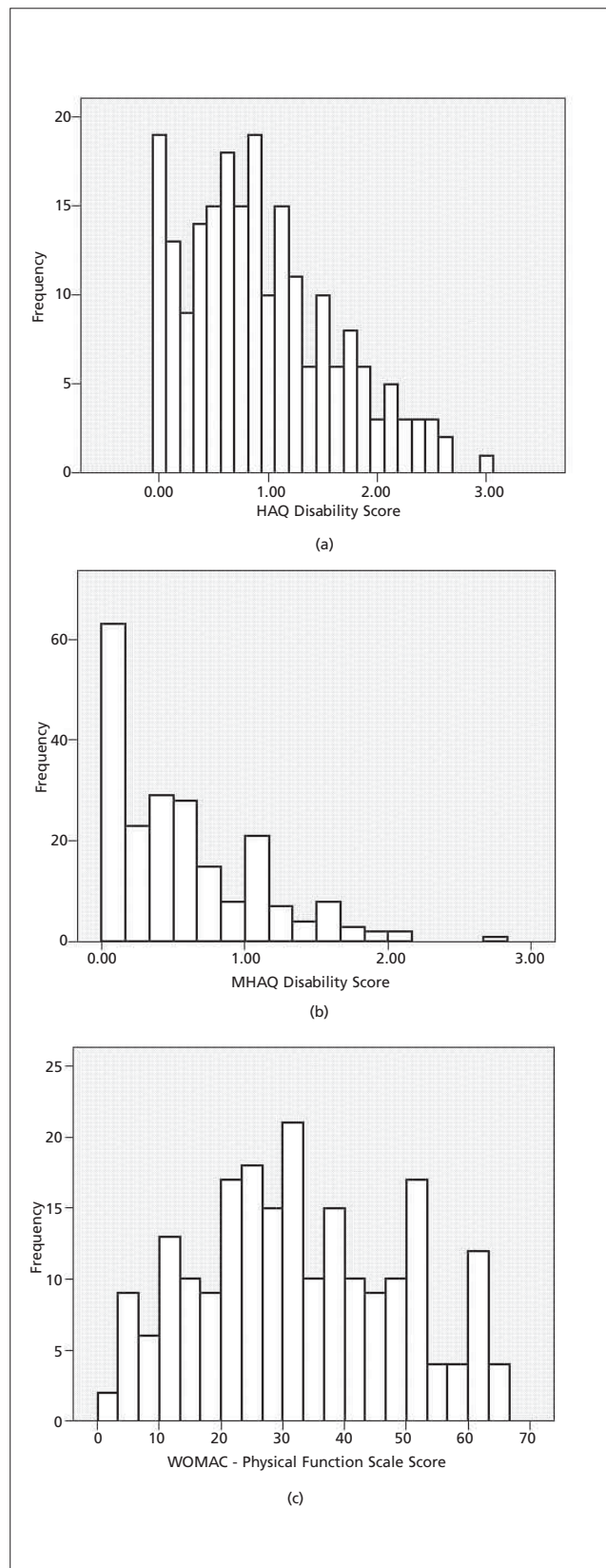HAQ-DI and MHAQ in Knee OA   153

Figure 3. Distributional characteristics of a) HAQ-DI, b) MHAQ and c) WOMAC-Physical function scale

scales for this diagnostic group of patients. The scales were not particularly well targeted to the current population, who displayed a level of disability much below the average difficulty level of the scales. This floor effect was much more prominent in MHAQ. Many patients were found to be at the lower limit for both scales whereas this was not the case for WOMAC-physical function subscale which showed a normal distribution among the patient sample. This distributional difference might be due to the fact that both HAQ-DI and MHAQ contain extra items assessing specifically upper extremity functions (36) whereas assessment of lower extremity function might be more salient in knee OA.

The distributional properties of HAQ and MHAQ were previously demonstrated in RA patients by various authors (5, 6, 37). Stucki et al. reported that 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 sensitivity to change in patients with RA (6). The data of Wolfe's study confirmed the observations of Stucki et al. regarding the floor effect (5). In a recent study which prospectively followed RA patients receiving infliximab treatment, Nagasawa et al showed that the MHAQ inevitably produced lower scores (indicating less disability) than the HAQ-DI, particularly among patients with high disability (37). The floor effect of HAQ-DI has also been demonstrated in patients with psoriatic arthritis (38).

While there has been little work to support the reliability and validity of the HAQ-DI in different diagnostic groups (such conditions), one recent study did report significant differential item functioning between a sample of patients with RA, OA and gout (39). Another study found similar DIF between RA and psoriatic arthritis (38). While this evidence does not preclude the scale working well within each condition, it raises interesting issues about comparability of scores across conditions.

This study has some limitations. First, we did not administer the MHAQ as a separate measure but scored it by using the HAQ-DI forms. Therefore we cannot exclude the possibility of different results if MHAQ had been administered as a separate questionnaire. Secondly, only the level of disability was assessed in the MHAQ whereas the original format also includes an evaluation of patient satisfaction. However most studies omit this second evaluation in MHAQ (5, 6, 37). Thirdly, this was a cross-sectional evaluation and responsiveness was not investigated. It would be good to see whether this abnormal distribution would have a negative effect on the responsiveness of both scales.

In conclusion, both the HAQ-DI and MHAQ are reliable and valid scales for assessing physical disability in patients with knee osteoarthritis. However clinicians and researchers should keep in mind the possible implications of a floor effect within both scales in this diagnostic group. Further evidence of the invariance of the scales

154 Köse et al.
HAQ-DI and MHAQ in Knee OA

Turk J Rheumatol 2010; 25: 147–55

across diagnostic groups, and appropriate score exchange rates across the different HAQ-DI versions will provide further evidence to support the use of the scales across a wide variety of settings.

Conflict of interest
No conflict of interest is declared by the authors.

References

1. Fries JF, Spitz P, Kraines G, Holman H. Measurement of Patient Outcome in Arthritis. Arthritis and Rheum 1980; 23: 137-45.
2. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003; 30: 167-78.
3. Lillegraven S, Kvien TK. Measuring disability and quality of life in established rheumaroid arthritis. Best Pract Res Clin Rheumatol 2007; 21: 827-40.
4. Pincus T, Summey JA, Soracı SA JR, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983; 26: 1346-53.
5. Wolfe F. Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ), and a rescored 20 item HAQ (HAQ20): analyses in 2,491 rheumatoid arthritis patients following leflunomide initiation. J Rheumatol. 2001; 28 (5): 982-9.
6. Stucki G, Stucki S, Briihlmann P, Michel BA. Ceiling effects of the Health Assessment Questionnaire and its modified version in some ambulatory rheumatoid arthritis patients. Ann Rheum Dis 1995; 54: 461-5.
7. Slatkowsky-Christensen B, Mowinckel P, Kvien TK. Health status and perception of pain: a comparative study between female patients with hand osteoarthritis and rheumatoid arthritis. Scand J Rheumatol 2009; 38: 342-8
8. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and therapeutic criteria committee of the American rheumatism association. Arthritis Rheum 1986; 29: 1039-49.
9. Küçükdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis Rheum 2004; 51: 14-9.
10. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.
11. Tüzün EH, Eker L, Aytar A, Daşkapan A, Bayramoğlu M. Acceptability, reliability, validity and responsiveness of the Turkish version of WOMAC osteoarthritis index. Osteoarthr Cartilage 2005; 13: 28-33.
12. World Health Organisation Disability Assessment Schedule II (2001) URL: http://www.who.int/icidh/whodas/
13. Ulug B, Ertugrul A, Gogus A, Kabakcı E. Yetiyitimi degerlendirme cizelgesinin (WHODAS-II) sizofreni

hastalarında gecerlilik ve güvenilirligi. Turk Psikiyatr Derg 2001; 12: 121-30.
14. European Group for Quality of Life Assessment and Health Measurement. European Guide for Nottingham Health Profile. Surrey: Brookwood Medical Publications, 1993.
15. Küçükdeveci AA, McKenna SP, Kutlay S, Gürsel Y, Whalley D, Arasil T. The development and psychometric assessment of the Turkish version of the Nottingham Health Profile. Int J Rehabil Res 2000; 23: 31-8.
16. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institution for Educational Research, 1960.
17. Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. Appl Psych Meas 1979; 3: 237-56.
18. Andrich D. Rasch Models for Measurement. London: SAGE Publications, 1988.
19. Karabatsos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. J Appl Meas 2001; 2: 389-423.
20. Masters G. A Rasch model for partial credit scoring. Psychometrika 1982; 47: 149-74.
21. Tennant A, Conaghan PG. The Rasch Measurement Model in Rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57: 1358-62.
22. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the hospital anxiety and depression scale (HADS). Br J Clin Psychol 2007; 46: 1-18.
23. Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205-231.
24. Wright BD. Local dependency, correlations and principal components. Rasch Meas Trans 1996; 10: 509-11.
25. Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. Stat Med 2000; 19: 1651-83.
26. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297-334.
27. Fisher WP: Reliability statistics. Rasch Measure Trans 1992; 6: 238.
28. Streiner DL, Norman GR. Health measurement scales. Oxford: Oxford University Press, 1995.
29. Nunnally JC. Psychometric theory. New York: McGraw-Hill, 1978.
30. Linacre JM. Sample size and item calibration stability. Rasch Measure Trans 1994; 7: 28.
31. Sloan JA, Symonds T, Vargas-Chanes D, Friedly B. Practical guidelines for assessing the clinical significance of healthrelated quality of life change within clinical trials. Drug Inf J 2003; 37: 23-31.
32. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310: 170.
33. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum. 1993; 36(6): 729-40.

Turk J Rheumatol 2010; 25: 147-55

Köse et al.
HAQ-DI and MHAQ in Knee OA 155

34. Anderson J, Sayles H, Curtis JR, Wolfe F, Michaud K. Converting MHAQ, MDHAQ and HAQII scores into HAQ scores using models developed with a large cohort of RA patients. Arthritis Care Res (Hoboken). 2010 May 23. [Epub ahead of print]

35. Kvien TK, Kaasa S, Smedstad LM. Performance of the Norwegian SF-36 Health Survey in Patients with Rheumatoid Arthritis. II. A Comparison of the SF-36 with Disease-Specific Measures. J Clin Epidemiol 1998; 51: 1077-86.

36. Bruce B, Fries J. Longitudinal Comparison of the Health Assessment Questionnaire (HAQ) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Arthritis & Rheum (Arthritis Care & Research) 2004; 51: 730-7.

37. Nagasawa H, Kameda H, Sekiguchi N, Amano K, Takeuchi T. Normalisation of physical function by infliximab in patients with RA: factors associated with normal physical function. Clin Exp Rheumatol 2010; 28: 365-72.

38. Taylor WJ, McPherson KM. Using Rasch analysis to compare the psychometric properties of the short form 36 physical function score and the health assessment questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis. Arthritis & Rheumatism (Arthritis Care & Research) 2007); 57: 723-9.

39. Van Groen MM, Ten Klooster PM, Taal E, van de Laar MA, Glas CA. Application of the health assessment questionnaire disability index to various rheumatic diseases. Qual Life Res Published [On-line first 2010].