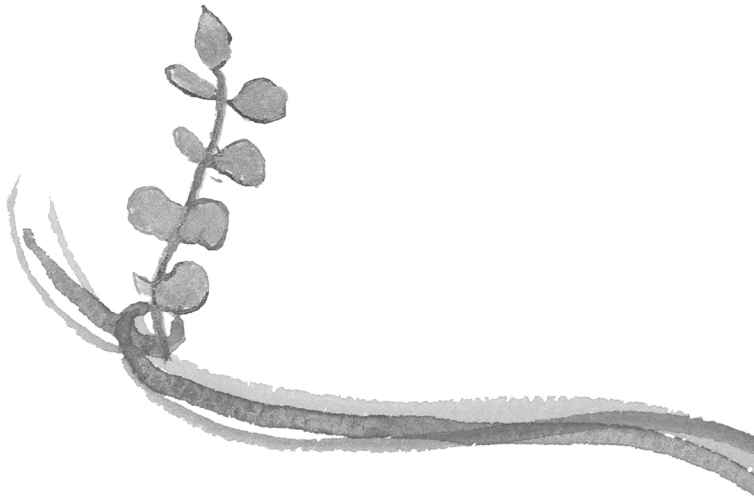
Inflammation as a target for treatment in hand osteoarthritis

Kroon, F.P.B.

CHAPTER 11

Performance of the Michigan Hand Outcomes Questionnaire in hand osteoarthritis



Féline P.B. Kroon
Aukje Boersma
Annelies Boonen
Sjoerd van Beest
Wendy Damman
Désirée van der Heijde
Frits R. Rosendaal
Margreet Kloppenburg

Osteoarthritis and Cartilage 2018;26:1627-35

ABSTRACT

Objective. To investigate the performance of the Michigan Hand Outcomes Questionnaire (MHQ) in hand osteoarthritis (OA) by evaluating truth, discrimination and feasibility.

Design. Symptomatic hand OA patients from the Hand Osteoarthritis in Secondary Care (HOSTAS) cohort completed questionnaires (demographics, MHQ, Australian/Canadian Hand Osteoarthritis Index [AUSCAN], Functional Index for Hand Osteoarthritis [FIHOA] and visual analogue scale [VAS] pain) at baseline (n=383), 1- and 2-year follow-up (n=312, n=293). Anchor questions at follow-up assessed whether pain/function levels were (un)acceptable and had changed compared to baseline. Correlations between MHQ and other pain/function questionnaires were calculated. Validity of unique MHQ domains (work performance, aesthetics, satisfaction), discrimination across disease stages, and responsiveness were assessed by categorizing patients by external anchors (employment, joint deformities, erosions, and anchor questions). Between-group differences were assessed with linear regression, probability plots and comparison of medians.

Results. MHQ pain and function subscales correlated moderately-to-good with other instruments (r_s 0.63-0.81). Work performance scores were worse in patients with reduced working capacity than in employed patients. Aesthetics scores were worse in patients with more deformities. Patients with unacceptable complaints had worse satisfaction scores. All pain/function instruments discriminated between patients with acceptable vs unacceptable pain/function, while only MHQ activities of daily living (ADL), FIHOA, and MHQ aesthetics could discriminate between erosive and non-erosive disease. MHQ and AUSCAN were most responsive.

Conclusions. MHQ has several unique aspects and advantages justifying its use in hand OA, including the unique assessment of work performance, aesthetics, and satisfaction. However, MHQ, AUSCAN and FIHOA appear to measure different aspects of pain and function.

INTRODUCTION

Hand osteoarthritis (OA) is a common disease with steeply increasing prevalence after the age of 50.[1] Patients experience various symptoms, including hand pain, functional limitations, and decreased quality of life.[2] Besides, many patients report dissatisfaction with their hands' appearance.[3-5] Since hand function is essential in many occupational activities, work productivity is likely also affected, though research in that area is scarce for hand OA.[6]

To measure burden of disease and effectiveness of therapeutic interventions, good outcome measures are essential. A recent systematic review showed that the Australian/Canadian Hand Osteoarthritis Index (AUSCAN) and Functional Index for Hand Osteoarthritis (FIHOA) are among the most frequently used instruments to measure pain and function in hand OA.[7] However, these questionnaires have some disadvantages: the AUSCAN is not freely available in the public domain, and the FIHOA includes some outdated and culturally challenging questions (e.g., handwriting more than 10 min, men use a screw driver while women sew, accepting a handshake).[8,9] The domain pain is often measured on a visual analogue scale (VAS), yet a single-item tool may not cover all relevant aspects, while a multi-item questionnaire like the AUSCAN pain scale may provide a more in-depth comprehension of pain.

The Michigan Hand Outcomes Questionnaire (MHQ) was specifically developed to measure outcomes of patients with hand disorders by a panel of patients with various hand conditions, hand therapists, and hand surgeons.[10] It consists of six subscales, measuring overall hand function, activities of daily living (ADL), pain, work performance, aesthetics, and satisfaction. Most subscales assess right and left hand separately. Its validity and reliability have been shown in patients undergoing hand surgery,[10] and in rheumatoid arthritis patients.[11,12] The MHQ has only partly been validated in hand OA.[13,14] This study aimed to examine the performance of the MHQ in hand OA by evaluating truth, discrimination and feasibility, guided by the Outcome Measures in Rheumatology (OMERACT) filter.[15,16] The MHQ was also compared to other widely used pain and function questionnaires in hand OA.

METHODS

Study design and population

Analyses were performed in the Hand OSTeoArthritis in Secondary care (HOSTAS) study, an ongoing observational cohort including patients diagnosed with primary hand OA by their treating rheumatologist at the Leiden University Medical Center (LUMC) outpatient clinic (for details on recruitment and selection[17]). All participants who completed the MHQ at baseline were included in the present study (n=383). Participants who completed the MHQ at the 1- and 2-year follow-up visit were included in longitudinal analyses (n=312 and n=293). The study was approved by the LUMC medical ethics committee. All participants gave written informed consent.

Data collection

Participants completed standardized questionnaires at baseline, 1- and 2-year follow-up visits. Employment status was assessed in eight categories, which were combined into: employed (fulltime/part-time job), reduced working capacity (sick leave, partial/complete disability to work), currently unemployed, and retired. Answers of participants using the option 'different' were reassigned to one of the other categories. Descriptions recorded most frequently were 'volunteering' (categorized as part-time job) and 'housewife' (categorized as currently unemployed). Participants with reduced working capacity could indicate whether this was due to hand OA or other causes.

The MHQ has six subscales: overall hand function (10 items), ADL (17 items), pain (5 items), work performance (5 items), aesthetics (8 items), and satisfaction (12 items). Items were rated on a 5-point Likert scale. Where applicable, right and left hand scores were averaged. Scores of each subscale were summated and subsequently normalised to 0-100, according to the published manual.[10] For each subscale higher scores are better, except for pain (higher scores indicate more pain).

The AUSCAN was completed on a 5-point Likert scale. Scores from the three subscales pain (5 items), stiffness (1 item) and function (9 items) range from 0 to 20, 4 and 36, respectively (higher is worse).[8] The FIHOA was completed on a 4-point Likert scale (10 items, range 0-30, higher is worse).[9] Self-reported hand pain was recorded on a 100 mm VAS.

Additionally, anchor questionnaires for pain and function were collected after 1- and 2-year follow-up in a random patient subgroup. Patients indicated whether their complaints of hand OA were acceptable or unacceptable at this moment (1-year follow-up n=95 [pain] and n=94 [function], 2-year follow-up n=142 [pain] and n=142 [function]), and whether their complaints were worse, unchanged, or improved compared to baseline (1-year follow-up n=114 [pain] and n=115 [function], 2-year follow-up n=164 [pain] and n=165 [function]).

Participants underwent standardized physical examination by trained research nurses at baseline and after 2-year follow-up. The number of hand joints with deformities was assessed in distal and proximal interphalangeal (DIP/PIP), IP-1, first metacarpal, and first carpometacarpal joints (range 0-22). Deformity was defined as >15° lateral deviation from the midline or squaring of the first carpometacarpal joint. Grip strength was measured with a hydraulic hand dynamometer (Saehan Corporation, Masan, South-Korea), and the average of two hands was calculated.

Hand radiographs were obtained at baseline and after 2-year follow-up. Osteophytes were assessed according to the Osteoarthritis Research Society International atlas, and summed to obtain a total score (range 0-58).[18] Presence of erosive disease was scored using the Verbruggen-Veys scoring system, defined as at least one joint in the erosive or remodelling (E/R) phase.[19] Radiographs were scored paired, in known time-order, by a well-trained reader (WD).[20]

Statistical analyses

Assessment of the MHQ was guided by the OMERACT filter, evaluating truth ('Does the MHQ measure what it intends to measure?'), discrimination ('Can the MHQ discriminate across disease stages?' and 'Can the MHQ identify change over time?' [further referred to as responsiveness]), and feasibility.[15,16]

Presence of floor/ceiling effects was explored by calculating the percentage of participants with the lowest/highest possible score at baseline for each questionnaire. A floor/ceiling effect was deemed present when ≥15% of participants scored the lowest/highest value.[21]

Truth

Correlation of the MHQ pain and function subscales with other questionnaires hypothesized to measure the same construct was assessed by calculating Spearman rank correlation coefficients (r_s). At least moderate correlation ($r_s \geq 0.50$) was expected between MHQ-pain and AUSCAN-pain and VAS pain, and between MHQ function subscales and AUSCAN-function and FIHOA. To further explore underlying differences, individual items of the questionnaires were compared. Items of function questionnaires were also grouped into those primarily assessing grip strength (e.g., hold heavy object) vs fine motor skills tasks (e.g., button shirt). Correlations between the four function scales, and grip strength and radiographic damage were tested, and hypothesized to be weak (r_s=0.30-0.50).

Validity of domains not captured in other questionnaires was assessed using external anchors to categorize patients. The work performance subscale was evaluated by comparing scores of participants with a different employment status (employed vs reduced working capacity), and between participants indicating hand OA or other circumstances as the cause of reduced working capacity. MHQ-aesthetics scores were compared between patients with <3, 3-5 and >5 joints with deformities (cut-offs based on tertiles). Scores on the satisfaction subscale at the 1- and 2-year visit were compared cross-sectionally between patients who indicated their complaints to be 'acceptable' vs 'unacceptable' at the same visit. Between-group differences were visualised on cumulative probability plots, in which individual MHQ scores of each patient are plotted in a cumulative order, presenting the cumulative probability per subgroup.[22] Mean between-group differences were calculated with linear regression (independent variable: patient category, dependent variable: MHQ subscale). Patients who were employed, with less deformities, and indicating to have 'acceptable' complaints were hypothesized to have higher scores on the corresponding MHQ subscale compared to the other categories.

Discrimination across disease stages

MHQ scores were compared across subgroups of participants in different disease stages. First, patients with erosive vs non-erosive disease. Since erosive OA is generally viewed as a more severe form of hand OA, scores on all outcome measures were expected to be worse in this group. Second, patients with 'acceptable' vs 'unacceptable' complaints at the 1- and 2-year follow-up visit. MHQ, AUSCAN, FIHOA, and VAS scores at the same follow-up visit were compared cross-sectionally between the groups in cumulative probability plots, and by comparing median scores (Mann-Whitney test).

Responsiveness
Data after 1-year and 2-year follow-up were used to assess responsiveness. Participants were divided into subgroups reporting worse vs stable/improved complaints after 1- and 2-year follow-up compared to baseline. Between-group differences in the change scores (follow-up minus baseline) of MHQ-pain and function subscales, AUSCAN-pain and -function, FIHOA, and VAS pain were compared as before (cumulative probability plots and medians).

Feasibility
To investigate feasibility, the MHQ was compared to other instruments with regard to costs, length of the questionnaire, number of missing items, and interpretability of the scores. Lower costs, lower number of (missing) items, and easily interpretable scores were viewed as advantages of a questionnaire.

Data were analysed using Stata V14, StataCorp LP, Texas.

RESULTS

Of 383 participants, 84% were women, with a median age of 60.3 years, and 90% fulfilled American College of Rheumatology (ACR) hand OA criteria (table 1).[23] At baseline, participants had moderate pain and functional impairment as measured on the different questionnaires. The majority was employed (table 1).

Of the six MHQ subscales, scores on the scales of satisfaction, overall hand function and pain were the worst, while scales of work performance, ADL, and aesthetics were less affected. MHQ scores were evidently worse compared to a healthy population.[24] At baseline, the interquartile range (IQR) of the scores on the MHQ function subscales represented only 17.5% (overall hand function) and 21.4% (ADL) of the maximum possible score, while the IQR of the FIHOA and AUSCAN-function represented 9 (30%) and 12 points (33%) of the maximum possible score, respectively. In contrast, the spread of the IQR of the scores on the different pain questionnaires was comparable. A ceiling effect was present for MHQ subscales work performance and aesthetics, but not for the other questionnaires. There were no floor effects (supplementary table 1).

Assessment of truth of MHQ pain and function domains

Correlations between MHQ-pain and other pain instruments were moderate (table 2). MHQ function subscales overall hand function and ADL both correlated moderately with AUSCAN-function and FIHOA, though correlations with MHQ-ADL were consistently better than with the overall hand function subscale (table 2). Correlations of MHQ function subscales with grip strength and radiographic damage were weak as expected (table 2). AUSCAN-function and FIHOA correlated similarly to these domains (r_s=-0.43/0.28 [grip strength/radiographic damage], and r_s=-0.39/r_s=0.27, respectively).

Table 1. Characteristics of study population at baseline (n=383), 1-year follow-up (n=312), and 2-year follow-up (n=293).

	Baseline		
Women, n (%)	321 (84)		
Age, years	60.3 (55.1-66.6)		
BMI, kg/m^2	26.6 (24.1-30.3)		
Employment status, n (%)			
Contributing to workforce	228 (60)		
Fulltime/part-time employed	179 (79)		
Reduced working capacity	49 (21)		
Due to hand OA	21 (43)		
Due to other causes	22 (45)		
Cause unknown	6 (12)		
Not contributing to workforce	155 (40)		
No job	40 (26)		
Retired	115 (74)		
Fulfilling ACR criteria, n (%)	343 (90)		
Erosive OA*, n (%)	112 (29)		
OARSI osteophyte score, 0-58	5 (12-19)		
Joints with deformity, 0-22	3 (1-6)		
Grip strength, kg	17.5 (22.3-28)		
	Baseline	**1 year**	**2 years**
AUSCAN			
Pain, 0-20	10 (7-12)	9 (6-12)	9 (5-11)
Stiffness, 0-4	2 (1-2)	2 (1-2)	2 (1-2)
Function, 0-36	16 (10-22)	16 (9-22)	16 (9-22)
FIHOA, 0-30	9 (4-13)	9 (4.5-13.5)	10 (5-14)
VAS pain, 0-100	35.5 (19.5-50)	*n/a*	32.5 (16.5-53.5)
MHQ, 0-100			
Overall hand function	57.5 (50-67.5)	57.5 (50-65)	57.5 (47.5-65)
ADL	80.5 (68.2-89.6)	79.9 (66.1-89.1)	77.5 (64.1-87.3)
Work performance	75 (50-95)	75 (55-95)	75 (55-95)
Pain	45 (31.3-60)	42.5 (30-55)	45 (30-55)
Aesthetics	81.3 (68.8-93.8)	84.4 (68.8-93.8)	81.3 (68.8-93.8)
Satisfaction	50 (35.4-66.7)	52.1 (35.4-70.8)	50 (31.3-70.8)

Median (interquartile range) unless otherwise specified. ACR, American College of Rheumatology; ADL, activities of daily living; AUSCAN, Australian/Canadian Hand Osteoarthritis Index; BMI, body mass index; FIHOA, Functional Index of Hand Osteoarthritis; MHQ, Michigan Hand Outcomes Questionnaire; OA, osteoarthritis; OARSI, Osteoarthritis Research Society International; VAS, visual analogue scale. *Erosive OA defined as at least one joint in E/R-phase according to Verbruggen-Veys scoring method.

In the FIHOA 4/10 questions assessed fine motor skills, and one assessed grip strength. Yet in the AUSCAN function more items evaluating grip strength (4/9) and less evaluating fine motor skills were present (2/9). The MHQ-ADL had 4/12 questions on grip strength and 3/12 on fine motor skills (supplementary table 2). The item "opening a jar", assessed in AUSCAN-function and MHQ-ADL, was the most often affected. MHQ overall hand function was unlike the other function questionnaires, with more general questions of hand function, movement, strength and sensation.

Table 2. Spearman correlation coefficients between MHQ and other pain/function instruments.

	MHQ pain	MHQ overall hand function	MHQ ADL
AUSCAN pain	-0.70		
VAS pain	-0.66		
AUSCAN function		-0.65	-0.81
FIHOA		-0.63	-0.81
Grip strength		0.39	0.42
Radiographic damage*		-0.25	-0.26

ADL, activities of daily living; AUSCAN, Australian/Canadian Hand Osteoarthritis Index; CMC-1, first carpometacarpal; DIP, distal interphalangeal; FIHOA, Functional Index of Hand Osteoarthritis; IP-1, first interphalangeal; MHQ, Michigan Hand Outcomes Questionnaire; OARSI, Osteoarthritis Research Society International; PIP, proximal interphalangeal; STT, scaphotrapeziotrapezoid; VAS, visual analogue scale. *OARSI osteophyte sum score of DIP 2-5, PIP 2-5, IP-1, CMC-1, and STT joints.

Assessment of truth of unique MHQ domains

Participants with reduced working capacity had worse scores on the MHQ work performance subscale compared to employed participants as shown in figure 1A, with a mean between-group difference of -25.7 (95% confidence interval [CI] -32.8; -18.6). Moreover, participants who indicated that hand OA was the cause of having a reduced working capacity had worse scores on the MHQ work performance scale than those who indicated there to be another cause (figure 1B, mean between-group difference -21.4 [-37.1; -5.8]).

To evaluate the aesthetics subscale, participants were divided into groups with increasing number of hand joints with OA deformities. Although the between-group difference was not large, participants with more deformities scored worse on the aesthetics subscale (figure 1C). Per additional deformity, the aesthetics score decreased with a mean of -1.03 (-1.60; -0.45) points.

The satisfaction subscale was tested using the anchor question of acceptability of the complaints of pain and function to categorize participants. Those who rated their pain at the 2-year follow-up visit as being unacceptable (n=19) had significantly worse scores on the satisfaction subscale at that visit than those who thought their level of pain was acceptable (n=123) (figure 1D), with a mean between-group difference of -27.2 (-37.1; -17.3). A similar result was found for participants rating their function at the 2-year follow-up visit as unacceptable (n=16) vs acceptable (n=126) (mean difference -28.5 [-39.1; -17.8]). Results at 1 year were comparable (data not shown).

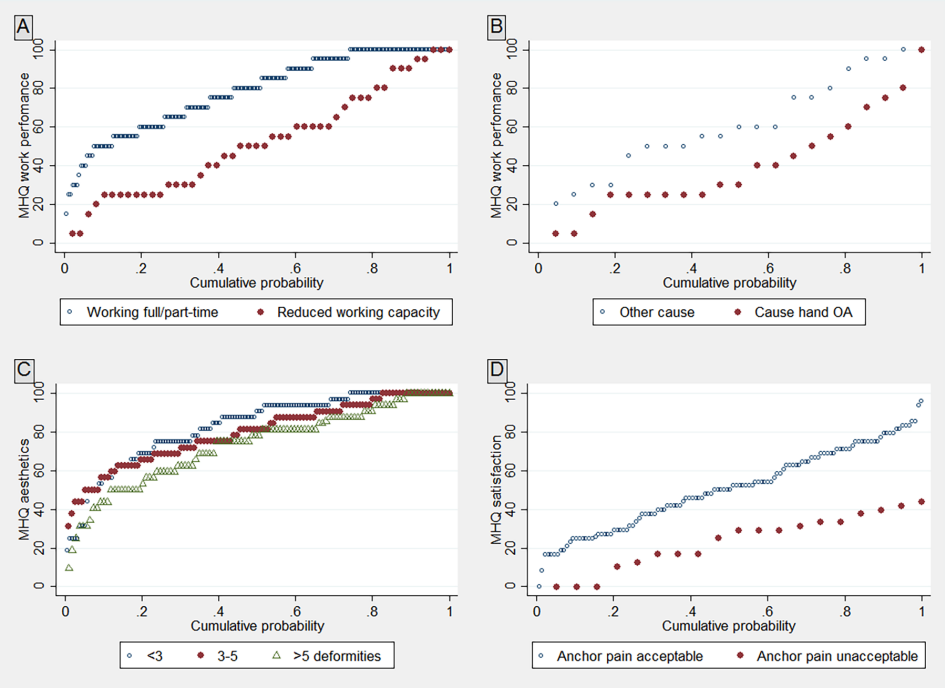


Figure 1. Cumulative probability plots comparing scores of MHQ work performance (A, B), aesthetics (C), and satisfaction (D) subscales across applicable external anchors.

Discrimination across disease stages

The scores of participants with erosive OA were compared to those with non-erosive OA (table 3). Although statistically significant between-group differences were found for MHQ-ADL and FIHOA, differences were small. None of the other function (MHQ overall function, AUSCAN-function) or pain scales (MHQ, AUSCAN or VAS pain) was able to clearly distinguish between the groups. Further exploration of the function questionnaires revealed that scores on the grip strength tasks were not different between patients with erosive and non-erosive OA, while a significant difference was apparent in the fine motor skills tasks. This difference was largest for fine motor skills tasks from MHQ-ADL (median [IQR] 20 [15; 30] vs. 15 [5; 25] for erosive vs. non-erosive OA, p=0.003), though tasks from AUSCAN-function and FIHOA showed the same trend (both $p<0.05$).

MHQ work performance and satisfaction subscales did not differ between those with erosive and non-erosive OA, while the scores on the MHQ aesthetics subscale of participants with erosive OA were evidently lower compared to the reference group (table 3).

Using the questions on acceptability of pain or function levels as the anchor to distinguish groups with more or less severe disease, all instruments that measure pain and function were similarly able to discriminate between those with acceptable vs unacceptable pain or function (table 3). None of the instruments performed better than another.

Responsiveness

To assess responsiveness, the anchor question assessing change in complaints of pain and function was used to divide participants into groups who experienced worse complaints compared to baseline, and those who indicated they remained stable or had improved (table 3). At the 2-year follow-up assessment, n=110 out of 164 who completed the anchor question reported worse pain, while n=54 reported stability (n=41) or improvement (n=9). For function, n=115 out of 165 had worsened, while n=50 stayed stable (n=36) or improved (n=18). Cumulative probability plots provide additional information to better assess differences in responsiveness between the questionnaires, including an assessment of the amount of change, minimum/maximum change scores, proportions of patients with a certain change score, and whether a questionnaire can discriminate over the whole spectrum of change or only in a certain range. The plots showed that both MHQ function subscales and AUSCAN-function could discriminate between these groups, while FIHOA could not (figure 2A, table 3). All pain scales discriminated between the two groups, though MHQ performed best, based on the largest between-group difference (figure 2B, table 3). Change scores in those reporting worse complaints were small for all questionnaires, while change scores in those reporting improvement were generally larger (not shown). Analyses were repeated using one-year follow-up data, achieving similar results (not shown).

Feasibility

Use of the MHQ requires online application for a license, though is free for academic/non-profit use. FIHOA and VAS are also freely available, while the AUSCAN is copyrighted, requiring a paid license to use. The MHQ has the most questions (n=58) and subscales. It is the only questionnaire with separately assessing right and left hand. The MHQ-ADL subscale is the longest function scale (17 items); the length of the others is comparable (MHQ overall hand function 10, FIHOA 10 and AUSCAN-function 9 items). The MHQ- and AUSCAN-pain subscales both contain 5 items, whereas the VAS is a single question. While the MHQ, FIHOA and AUSCAN are standardized questionnaires, wording of the VAS may differ among studies. Number of missing items was low (<1%) and not different for the questionnaires (supplementary table 3). Scores on the MHQ-pain subscale have to be interpreted in the opposite direction as compared to the questionnaire's other subscales, which may negatively impact interpretability.

Table 3. Discrimination across disease stages (baseline scores) and responsiveness (change scores over 2 years) of MHQ and other questionnaires.

I. Discrimination across disease stages			
Domain	**Non-erosive OA (n=268)**	**Erosive OA (n=112)**	**p-value**
Function			
MHQ overall hand function, 0-100	60 (50; 67.5)	56.3 (50; 65)	0.37
MHQ ADL, 0-100	81.3 (69.8; 90.2)	76.8 (64.5; 87.9)	0.04
AUSCAN function, 0-36	15 (10; 21)	17 (10; 22)	0.22
FIHOA, 0-30	8 (4; 12)	10 (5; 14)	0.01
Pain			
MHQ pain, 0-100	45 (30; 60)	47.5 (35; 60)	0.23
AUSCAN pain, 0-20	9 (6; 12)	10 (7; 12.5)	0.11
VAS pain, 0-100	34 (19.5; 49)	38.3 (20.3; 54)	0.05
Additional domains			
MHQ work performance, 0-100*	75 (55; 95)	80 (60; 100)	0.09
MHQ aesthetics, 0-100	87.5 (71.9; 96.9)	75 (59.4; 87.5)	<0.001
MHQ satisfaction, 0-100	52.1 (37.5; 68.8)	50 (33.3; 65.6)	0.21
Function	**Function acceptable (n=126)†**	**Function unacceptable (n=16)†**	
MHQ overall hand function, 0-100	56.3 (47.5-62.5)	42.5 (32.5-47.5)	<0.001
MHQ ADL, 0-100	76.4 (65.3-84.5)	51.1 (45.2-71.0)	<0.001
AUSCAN function, 0-36	16 (10-21)	22.5 (17-25)	0.01
FIHOA, 0-30	10 (6-14)	14.5 (11.5-17)	0.02
Pain	**Pain acceptable (n=123)†**	**Pain unacceptable (n=19)†**	
MHQ pain, 0-100	45 (30-55)	65 (50-75)	<0.001
AUSCAN pain, 0-20	9 (5-11)	14 (11-15)	<0.001
VAS pain, 0-100	33 (16-53)	63 (53.5-72)	<0.001
II. Responsiveness			
Function	**Function worse (n=115)†**	**Function stable/improved (n=50)†**	
MHQ overall hand function, 0-100‡	-5 (-12.5; 5)	1.3 (-2.5; 7.5)	0.001
MHQ ADL, 0-100‡	-4.4 (-11.4; 3.2)	0 (-5.9; 7.3)	0.01
AUSCAN function, 0-36§	1 (-3; 4)	-1 (-5; 2)	0.01
FIHOA, 0-30§	1 (-2; 3)	0 (-2; 2)	0.16
Pain	**Pain worse (n=110)†**	**Pain stable/improved (n=54)†**	
MHQ pain, 0-100§	0 (-10; 15)	-10 (-20; 5)	<0.001
AUSCAN pain, 0-20§	0 (-2; 2)	-1 (-5; 0.5)	0.004
VAS pain, 0-100§	2.5 (-10.5; 14)	-2 (-14; 4)	0.04

Median (interquartile range). P-value for between-group difference, derived from Mann-Whitney test. ADL, activities of daily living; AUSCAN, Australian/Canadian Hand Osteoarthritis Index; FIHOA, Functional Index of Hand Osteoarthritis; MHQ, Michigan Hand Outcomes Questionnaire; OA, osteoarthritis; VAS, visual analogue scale. *Calculated for participants in the workforce at baseline (n=174 with non-erosive OA and n=53 with erosive OA). †Two-year follow-up data. ‡Positive values indicate improvement compared to baseline. §Negative values indicate improvement compared to baseline.

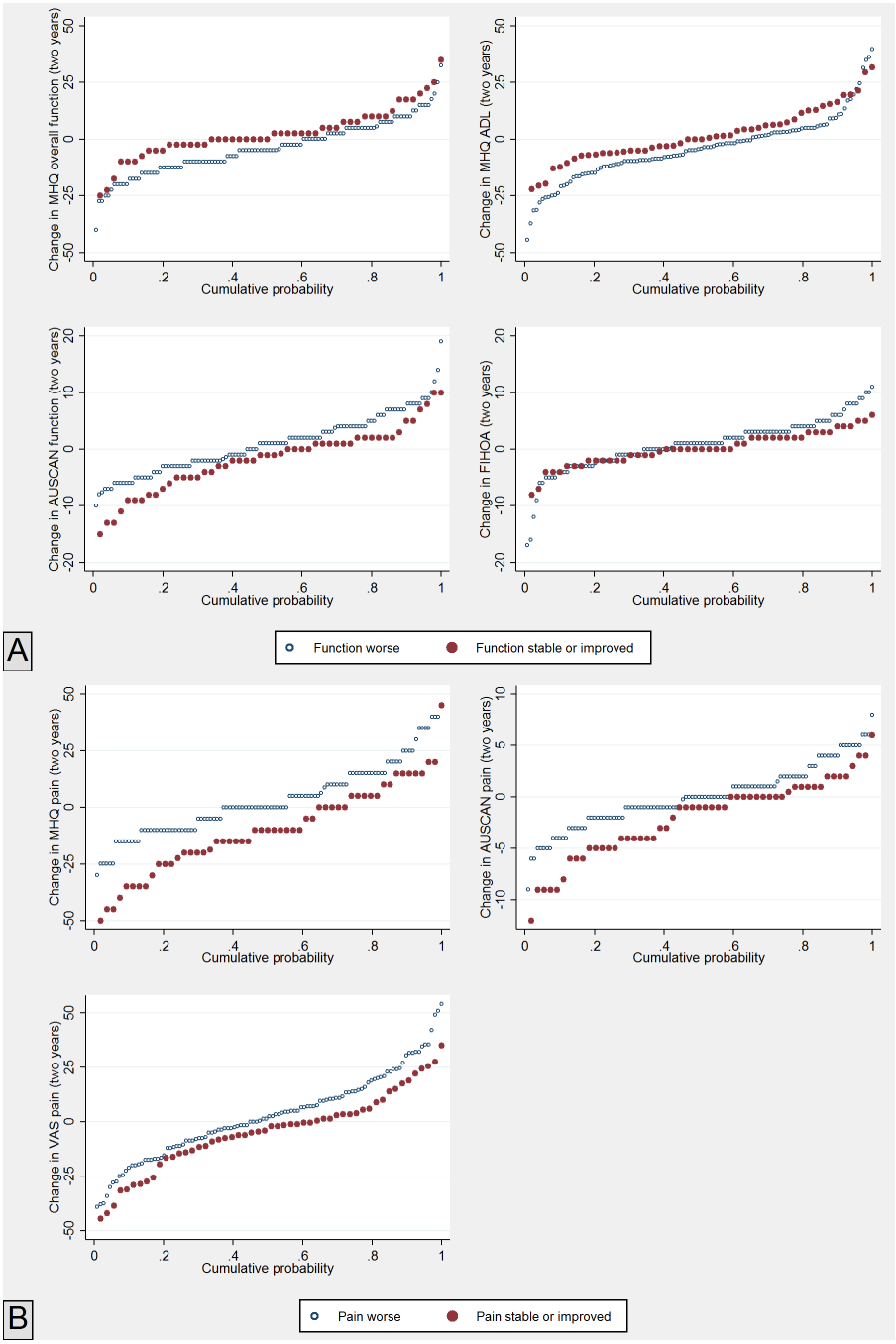


Figure 2. Cumulative probability plots comparing change scores of questionnaires assessing function (A) and pain (B) of patients who stated after 2 years that their complaints had worsened vs stayed stable or improved.

DISCUSSION

We studied performance of the MHQ in hand OA patients regarding the aspects truth, discrimination, and feasibility according to the OMERACT filter.[15,16] The three domains unique for the MHQ (work, aesthetics and satisfaction) were demonstrated to be valid, compared to relevant external anchors. These subscales can provide useful insights in domains that are likely of importance for hand OA patients but currently understudied in absence of valid measurement instruments. Furthermore, MHQ-pain and function subscales measured partly the same, but partly also different aspects of pain and function as other widely used pain and function questionnaires in hand OA (AUSCAN, FIHOA, VAS). Discrimination across disease stages and responsiveness of MHQ was similar to the reference instruments and on some aspects even better. Regarding feasibility, MHQ has the advantage of being freely available and possibility to assess left and right hand separately, though it was the most lengthy (also reflected in previously published longer time to complete[25]) and different direction of effect of different subscales may negatively impact its interpretability.

A core set of domains to be measured in hand OA studies, as well as a preliminary set of instruments to be used, was developed by OMERACT.[26] Before endorsing the MHQ as a core instrument to measure outcomes in hand OA, investigation of its metric properties should be performed.[26] Two previous studies assessed measurement properties of the MHQ in hand OA patients.[13,14] Both concluded it to be valid and reliable, yet neither included specific hand pain questionnaires or the often used AUSCAN as a reference. A small study by Poole *et al* compared several hand function tests, including the Cochin hand function scale, FIHOA, MHQ, and three performance-based instruments, in a sample of 40 volunteers with self-reported hand OA.[13] They found a correlation of -0.86 between the MHQ total score (including all subscales) and the FIHOA, which is higher than the correlation we found between the MHQ function subscales and FIHOA. However, the results are not directly comparable since their study did not look at the MHQ function subscales specifically. A study by Marks *et al* investigated the MHQ in 177 thumb base OA patients receiving conservative or surgical treatment.[14] They assessed correlations between MHQ subscales and the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire and Short-Form 12, and found lower correlations than we did. Our use of specific hand pain and function (subscales of) questionnaires as a comparator, while Marks *et al* used the more generic (total) DASH, which reflects functioning of the total arm, may explain the difference.

The MHQ was compared to the most often used hand pain and function instruments, but a golden standard is not available. Although the moderate correlations indicate less than perfect concordance between the different questionnaires, one cannot conclude from this which is more truthful.

In-depth comparison of the individual items of the questionnaires revealed that each assessed rather different aspects of pain and function. While the AUSCAN-pain asks about pain severity during rest and several tasks (lifting, squeezing, turning, gripping), the MHQ-pain focusses on the frequency of experiencing pain in several situations (in general, while sleeping, during ADL) and whether it affects the respondent's happiness. The questionnaires

may therefore answer different research questions, and reflect different dimensions of disease impact. Although the MHQ-ADL, AUSCAN-function and FIHOA appeared to be more alike at first glance, all assessing the ability to perform certain tasks with one's hands, the type of tasks that were assessed differed. In accordance with a previous study, we found that fine motor skills tasks were only affected in more severe disease stages.[27] Furthermore, in agreement with previous studies, we found that "opening a jar" (assessed in AUSCAN and MHQ, but not FIHOA) was the most often affected item.[27,28] Since the different function questionnaires place more or less emphasis on certain types of tasks (e.g., grip strength or fine motor skills) by including relatively more or less items assessing it, the selection of a function questionnaire should also depend on the targeted population.

Another difference between the studied questionnaires, was the use of different recall periods, ranging from none specified (FIHOA) to 48 h (AUSCAN) or a week (MHQ). This may also explain the moderate correlations. While it is undoubtedly better to define a time period, shorter and longer time periods both have their advantages. A shorter time period reduces the amount of recall bias, yet may be more subject to fluctuations in symptom severity due to short term circumstances like daily activities or work.

In this study, correlations between self-reported pain and function on the one side, and hand strength or radiographic damage on the other, were only weak. The limited association between clinical symptoms and radiographic damage is known from many previous studies.[29] In this respect, the MHQ was comparable to other pain and function questionnaires for hand OA.

While discrimination between patients who indicated their complaints to be acceptable or non-acceptable was clear for all questionnaires, discrimination between erosive and non-erosive disease was not. Only MHQ-ADL, FIHOA, and MHQ-aesthetics were able to detect a difference between these two groups, though differences on the MHQ-ADL and FIHOA were small. It is possible that inclusion of relatively more fine motor skills tasks increased the ability of these questionnaires to discriminate between these hand OA subtypes. A previous study comparing clinical signs and symptoms of erosive and non-erosive hand OA patients, generally found larger between-group differences than the current study. This may be due to differences in patient population, as that study assessed patients with polyarticular OA whose hands may not have been primarily affected.[30]

Despite a 2-year follow-up, not much change occurred in this cohort. Since inclusion occurred often after their first visit at our outpatient clinic, patients were likely included at the height of their complaints, resulting in the phenomenon 'regression to the mean'. Nevertheless, in absence of disease-modifying interventions, disease progression will have occurred in a number of patients. These two trends may have balanced each other out, resulting in average little change over time. It also reflects the previously described heterogeneous character of the disease course, for about half of the patients deteriorates, while a quarter improves.[31] Moreover, assessment of pain and function at patient level using questionnaires may not capture changes on joint level. Complaints in one joint may improve, while complaints in another joint may arise or worsen, so on patient level we measure no change over time. Nevertheless, participants who self-reported worse or stable/improved complaints over time were compared to assess responsiveness. In this respect, the MHQ and AUSCAN performed best. However, for all

questionnaires change scores of patients reporting worse complaints were small. This may indicate that these instruments are not so sensitive with regard to worsening. In trial settings though, assessment of improvement is more important. Change scores of those reporting improvement were indeed relatively larger. The instruments may thus be more sensitive to measure improvement, or it could be indicative of a threshold phenomenon (i.e., improvement has to be above a certain threshold before a patient reports to be improved).

Strengths of this study are its large sample size, the inclusion of the most often used hand pain and function instruments in hand OA for comparison, as well as external anchor questions. The most important limitation is the lack of an (effective) treatment group to be able to assess sensitivity-to-change.

In conclusion, this study shows that the MHQ has several advantages that justify its use in future hand OA studies. Most important benefits are the three unique domains, that the questionnaire is free to use, and on some aspects better discriminative abilities than other questionnaires. However, MHQ, AUSCAN and FIHOA do appear to measure different aspects of pain and function, and it cannot be concluded from this study which questionnaire is more valid in the setting of hand OA, which may further depend on the specific study question. Assessment of the MHQ's sensitivity-to-change is warranted, preferably in future (positive) clinical trials.

REFERENCES

1. Kloppenburg M, Kwok WY. Hand osteoarthritis--a heterogeneous disorder. *Nat Rev Rheumatol* 2012;8:22-31.
2. Michon M, Maheu E, Berenbaum F. Assessing health-related quality of life in hand osteoarthritis: a literature review. *Ann Rheum Dis* 2011;70:921-8.
3. Hodkinson B, Maheu E, Michon M, et al. Assessment and determinants of aesthetic discomfort in hand osteoarthritis. *Ann Rheum Dis* 2012;71:45-9.
4. Liu R, Damman W, Beaart-van de Voorde L, et al. Aesthetic dissatisfaction in patients with hand osteoarthritis and its impact on daily life. *Scand J Rheumatol* 2016;45:219-23.
5. Neuprez A, Bruyère O, Maheu E, et al. Aesthetic discomfort in hand osteoarthritis: results from the LIège Hand Osteoarthritis Cohort (LIHOC). *Arthritis Res Ther* 2015;17:346.
6. Marks M, Vliet Vlieland T, Audigé L, et al. Healthcare costs and loss of productivity in patients with trapeziometacarpal osteoarthritis. *J Hand Surg Eur Vol* 2015;40:927-34.
7. Visser AW, Bøyesen P, Haugen IK, et al. Instruments measuring pain, physical function, or patient's global assessment in hand osteoarthritis: a systematic literature search. *J Rheumatol* 2015;42:2118-34.
8. Bellamy N, Campbell J, Haraoui B, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis Cartilage* 2002;10:855-62.
9. Dreiser R, Maheu E, Guillou G, et al. Validation of an algofunctional index for osteoarthritis of the hand. *Rev Rhum Engl Ed* 1995;62:43S-53S.
10. Chung K, Pillsbury M, Walters M, et al. Reliability and valdity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg* 1998;23A:575-87.
11. Massy-Westrop N, Krishnan J, Ahern M. Comparing the AUSCAN Osteoarthritis Hand Index, Michigan Hand Outcomes Questionnaire, and Sequential Occupational Dexterity Assessment for patients with rheumatoid arthritis. *J Rheumatol* 2004;31:1996-2001.
12. Waljee J, Chung K, Kim H, et al. Validity and responsiveness of the Michigan Hand Questionnaire in patients with rheumatoid arthritis: a multicenter, international study. *Arthritis Care Res* 2010;62:1569-77.
13. Poole JL, Lucero SL, Mynatt R. Self-reports and performance-based tests of hand function in persons with osteoarthritis. *Phys Occup Ther Geriatr* 2010;28:249-58.
14. Marks M, Audigé L, Herren D, et al. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. *Arthritis Care Res* 2014;66:245-52.
15. Boers M, Brooks P, Strand CV, et al. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198-9.
16. Boers M, Kirwan JR, Wells G, Beaton D, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol* 2014;67:745-53.
17. Damman W, Liu R, Kroon F, et al. Do comorbidities play a role in hand osteoarthritis disease burden? Data from the Hand OSTeoArthritis in Secondary care cohort. *J Rheumatol* 2017;44:1659-66.
18. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage* 2007;15:A1-56.
19. Verbruggen G, Veys EM. Numerical scoring systems for the anatomic evolution of osteoarthritis of the finger joints. *Arthritis Rheum* 1996;39:308-20.
20. Damman W, Liu R, Bloem JL, et al. Bone marrow lesions and synovitis on MRI associate with radiographic progression after 2 years in hand osteoarthritis. *Ann Rheum Dis* 2017;76:214-7.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
22. Landewé R, van der Heijde D. Radiographic progression depicted by probability plots: presenting data with optimal use of individual values. *Arthritis Rheum* 2004;50:699-706.
23. Altman R, Alarcón G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis Rheum* 1990;33:1601-10.
24. Nolte MT, Shauver MJ, Chung KC. Normative values of the Michigan Hand Outcomes Questionnaire for patients with and without hand conditions. *Plast Reconstr Surg* 2017;140:425e-33e.
25. Poole JL. Measures of Hand Function. Arthritis Hand Function Test (AHFT), Australian Canadian Osteoarthritis Hand Index (AUSCAN), Cochin Hand Function Scale, Functional Index for Hand Osteoarthritis (FIHOA), Grip Ability Test (GAT), Jebsen Hand Function Test (JHFT), and Michigan Hand Outcomes Questionnaire (MHQ). *Arthritis Care Res* 2011;63:S189-99.
26. Kloppenburg M, Bøyesen P, Visser AW, et al. Report from the OMERACT hand osteoarthritis working group: set of core domains and preliminary set of

instruments for use in clinical trials and observational studies. *J Rheumatol* 2015;42:2190-7.
27. Haugen IK, Moe RH, Slatkowsky-Christensen B, et al. The AUSCAN subscales, AIMS-2 hand/finger subscale, and FIHOA were not unidimensional scales. *J Clin Epidemiol* 2011;64:1039-46.
28. Kroon F, Ramiro S, Royston P, et al. Reference curves for the Australian/Canadian Hand Osteoarthritis Index in the middle-aged Dutch population. *Rheumatology (Oxford)* 2017;56:745-52.
29. Visser AW, Bøyesen P, Haugen IK, et al. Radiographic scoring methods in hand osteoarthritis - a systematic literature search and descriptive review. *Osteoarthritis Cartilage* 2014;22:1710-23.
30. Bijsterbosch J, Watt I, Meulenbelt I, et al. Clinical burden of erosive hand osteoarthritis and its relationship to nodes. *Ann Rheum Dis* 2010;69:1784-88.
31. Bijsterbosch J, Watt I, Meulenbelt I, et al. Clinical and radiographic disease course of hand osteoarthritis and determinants of outcome after 6 years. *Ann Rheum Dis* 2011;70:68-73.