Universiteit Leiden
The Netherlands

Treatment of patients with hand osteoarthritis : outcome measures, patient satisfaction, and economic evaluation

Marks, M.

Cover Page

Universiteit Leiden

Leiden University Repository

The handle http://hdl.handle.net/1887/28732 holds various files of this Leiden University dissertation.

Author: Marks, Miriam
Title: Treatment of patients with hand osteoarthritis : outcome measures, patient satisfaction, and economic evaluation
Issue Date: 2014-09-11

CHAPTER **FOUR**

MEASUREMENT PROPERTIES OF THE GERMAN MICHIGAN HAND OUTCOMES QUESTIONNAIRE IN PATIENTS WITH TRAPEZIOMETACARPAL OSTEOARTHRITIS

Miriam Marks
Laurent Audigé
Daniel B Herren
Stephan Schindele
Rob GHH Nelissen
Thea PM Vliet Vlieland

ABSTRACT

Objective To investigate the reliability, validity, and responsiveness of the Michigan Hand Outcomes Questionnaire (MHQ) in patients with trapeziometacarpal (TMC) joint osteoarthritis (OA).

Methods In this prospective observational study, patients diagnosed with TMC joint OA who received either conservative or surgical treatment were included. At baseline and at 1 year following the beginning of treatment, we measured key pinch strength and the patients filled out the MHQ, the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire, and the Short Form 12 health survey. Patients also completed these questionnaires 2 – 11 days after the last study visit. In order to analyze the measurement properties of the MHQ, we calculated test–retest reliability (intraclass correlation coefficient [ICC]), internal consistency (Cronbach's alpha for the 6 subscales), construct validity (Pearson's correlation coefficient [r]), responsiveness (effect sizes), and the minimum important change (MIC).

Results We included 177 patients, of whom 109 were scheduled for surgery. The mean ± SD MHQ total score for surgical patients increased from 48 ± 14 at baseline to 75 ± 18 at 1 year (p ≤ 0.001). In contrast, no treatment effect was observed in the conservative group (p = 0.74). The MHQ total score showed excellent test–retest reliability (ICC = 0.95) and correlated strongly with the DASH (r = -0.77). Internal consistency of the MHQ subscales ranged between 0.77 and 0.89. A large effect size of 1.7 was found for the surgical patients, with an MIC of 17 points.

Conclusion The MHQ demonstrated good reliability, validity, and responsiveness in patients with TMC joint OA and can be recommended as a suitable assessment instrument in this population.

FOUR

Significance & Innovations

- The Michigan Hand Outcomes Questionnaire (MHQ) is widely used in clinical trials to assess the outcome of interventions in patients with various hand disorders.
- Although the MHQ has already been used in studies that included patients with trapeziometacarpal (TMC) joint osteoarthritis (OA), its measurement properties have not yet been investigated in this population.
- Based on good results regarding reliability, validity, and responsiveness, we can recommend the MHQ as a suitable assessment instrument for patients with TMC joint OA.

INTRODUCTION

Among the joints of the hand, the trapeziometacarpal (TMC) joint is, after the distal interphalangeal joints, the joint most frequently affected by osteoarthritis (OA). The prevalence is 14.2% in the 50-59 years age group and increases with higher age[1–3]. TMC joint OA causes symptoms such as pain and loss of grip strength, as well as limiting daily activities and social participation[1, 4]. Given this high prevalence, it is essential to have a standardized assessment tool that allows comparison of interventions and provides evidence of best practice. In recent years, subjective evaluations based on the patient's self-assessment of function, activities of daily living (ADL), and quality of life, as well as on patient satisfaction, have emerged as increasingly important outcome measures for musculoskeletal conditions in general. Various questionnaires are available to assess subjective aspects in patients experiencing TMC joint OA, with the Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire[5] being the one used most frequently[6]. However, the validity and responsiveness of this tool in these particular patients remain questionable as the score is also influenced by function/dysfunction of the elbow and shoulder. For this reason, it might be more appropriate to administer a hand-specific questionnaire[6]. The Michigan Hand Questionnaire (MHQ), developed by Chung et al.[7], is one such hand-specific questionnaire. In contrast to other commonly used function questionnaires, the MHQ has some unique features. First, it yields results for each hand separately. Second, it consists of a multidimensional construct, including a section on aesthetics, which is especially important in patients with rheumatoid arthritis[8, 9]. The MHQ consists of 37 items categorized into 6 subscales as follows: hand function, ADL, pain, work performance, aesthetics, and satisfaction with hand function. The MHQ has been translated and culturally adapted into several languages[10–15]. Furthermore, a short version of the MHQ (the BriefMHQ), including only 12 items, has recently been developed[16, 17]. However, from the brief version it is not possible to derive subscale scores or to distinguish between the right and left hand.

The measurement properties of the original MHQ have been assessed in patients with rheumatoid arthritis[9, 18–20], carpal tunnel syndrome[19, 21, 22], and distal radius fractures[19, 23], as well as in patients with various other hand problems[10, 12, 21, 24, 25], with overall good reliability, validity, and responsiveness. Furthermore, the MHQ compares favorably with other hand outcomes instruments[8]. Although it has already been used in several studies that included patients with TMC joint OA, the measurement properties of the MHQ have not yet been demonstrated in this population[26, 27]. The aim of the present study was to investigate the reliability, validity, and responsiveness of the MHQ in patients with TMC joint OA.

PATIENTS AND METHODS

Study design

The MHQ study was part of a prospective observational study on the effects of conservative and surgical treatment for TMC joint OA. The study was carried out in accordance with the ethical principles of the Declaration of Helsinki and approved by the local ethics committee (Kantonale Ethikkommission Zurich, Switzerland).

FOUR

Patients and interventions

Patients were eligible for the study if they had radiologic-proven TMC joint OA diagnosed by an experienced hand surgeon and had undergone either conservative or surgical treatment for that condition between September 2011 and November 2012. All eligible patients were asked to participate by their treating hand surgeon, and were consecutively enrolled in the study once they had given written informed consent. Exclusion criteria were as follows: TMC joint OA was not the main problem at the time of consultation, rheumatoid arthritis or other diseases interfering with hand function, concomitant surgery on other finger joints, legal incompetence, poor general condition precluding study participation, previous inclusion in the study for the other hand, and insufficient knowledge of the German language to complete the questionnaires.

Treatment consisted of conservative management (injection, analgesics, or occupational therapy) or surgery (resection/suspension/interposition arthroplasty or arthrodesis) as chosen by the surgeon in discussion with the patient in each case.

Outcome measures

Patients in the main study were assessed before treatment and at 3, 6, and 12 months after the start of treatment. For this substudy on the measurement properties of the MHQ, we used data from baseline and the 1-year followup. At baseline, sociodemographic and disease-related data were gathered. At each study visit, patients were assessed clinically and completed a questionnaire set consisting of the MHQ, the DASH, and the Short Form 12 (SF-12) health survey, version 2.0. Two to 11 days after the 1-year followup, patients filled out the questionnaire set again.

Key pinch strength was assessed using a digital pinch gauge (ELINK, Biometrics) in a standardized sitting position. The average of 3 measurements on the affected hand was retained for further analysis.

The MHQ has been translated into German[11]. The 6 subscales were calculated using the algorithm published by Chung et al[7]. The raw figures were converted to a score ranging from 0 to 100. Higher scores indicate better performance, except for the pain subscale, where a higher score denotes more pain. The MHQ total score was obtained by summing the scores for all 6 subscales (after reversing the pain scale) and then dividing the sum score by 6[28]. For the present analysis, only the data for the affected hand were retained.

The DASH is a questionnaire commonly used to evaluate pain and function of the upper extremity and does not distinguish between affected and nonaffected upper extremities[5, 29]. It shows sound measurement properties for patients with TMC joint OA, although the items are not purely hand-specific and are partly influenced by function/dysfunction of the elbow and shoulder joints[6]. Like the MHQ, the DASH total score ranges from 0 to 100, where higher scores indicate greater disability.

The SF-12 is a short version of the SF-36, which assesses quality of life[30]. Its 12 questions cover the 8 subscales of physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health, leading to the 2 component summary measures of physical health and mental health. The SF-12 itself has not been investigated in patients with TMC joint OA, although its original version, the SF-36, has[31–34]. At the 1-year followup, we asked about the perceived change in the thumb condition related to baseline and patients answered on a 5-point Likert scale. This scale was transformed into a dichotomous scale, with patients who had answered "much better" or "slightly better" being allocated into

the improved group. Patients who answered "unchanged," "slightly worse," or "worse" were allocated into the comparison group of unimproved subjects.

Statistical analysis

Sociodemographic and disease-related characteristics were analyzed descriptively. We determined the items initially missing from the questionnaires returned at baseline and contacted the patients by telephone to ask them to complete their answers in order to have as few missing items as possible. The Wilcoxon signed rank test was carried out in each subgroup for the MHQ total score, the DASH, the SF-12, and the key pinch to see whether there were significant treatment effects in patients treated surgically or conservatively.

Evaluation of the measurement properties of the MHQ was based on the definitions and recommendations of the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) Group[35–37], which are outlined below.

Reliability is defined as the degree to which the measurement is free from measurement error and is usually established by test–retest reliability, internal consistency, and estimated measurement error. Test–retest reliability was estimated by the intraclass correlation coefficient (ICC) using the data from the 1-year followup and those collected 2 - 11 days later. No change in the thumb condition was expected within this short period. An ICC ≥ 0.7 is considered acceptable, but values ≥ 0.8 are much better[36]. Using baseline data, we calculated Cronbach's alpha for each subscale to evaluate internal consistency. Values between 0.7 and 0.9 are regarded as good internal consistency, higher values indicate redundancy[36]. To obtain the measurement error, the standard error of measurement (SEM) was calculated by dividing the SD of the difference between test and retest by $\sqrt{2}$. Not every change in a measurement instrument can be considered as a true change; a change might occur due to measurement errors. The smallest detectable change (SDC) represents the change beyond measurement error, and any change smaller than the SDC can be regarded as measurement error. The SDC was calculated as $1.96 \times \sqrt{2} \times$ SEM[36].

Construct validity is the degree to which an instrument measures the construct(s) it is intended to measure and can be further divided into convergent and discriminant construct validity. Convergent construct validity means that the instrument under investigation highly correlates with another instrument that reflects a similar construct[36]. In the case of the MHQ, we chose the DASH as a comparator, as it intends to measure function and pain of the upper extremities, including the hand. Discriminant construct validity means that instruments that measure different constructs show only slight or no correlations[36]. For this purpose, we chose key pinch as a comparator for the MHQ function subscale, because hand function includes more aspects than only key pinch strength. Moreover, we selected the SF-12 mental health score, which intends to measure a completely different construct than the MHQ.

According to the recommendations of the COSMIN Group[35–37], we tested predefined specific hypotheses to investigate the construct validity. The number of hypotheses to be tested has not been defined by this group[38], so we assumed that 5 would be sufficient to prove or reject the construct validity of the MHQ. Using the baseline data, the following 3 hypotheses for convergent construct validity were tested with Pearson's correlation coefficients: 1) the MHQ ADL subscale correlates strongly with the DASH, with $r \leq -0.7$, 2) the MHQ pain subscale correlates strongly with the DASH, with $r \geq 0.7$, and 3) the MHQ total score correlates strongly with the DASH, with

FOUR

r ≤ -0.7. For discriminant construct validity, the following hypotheses were tested: 1) the MHQ hand function subscale correlates mildly with key pinch strength, with 0.5 ≥ r ≥ 0.3, and 2) the MHQ hand function subscale does not correlate with the SF-12 mental health score, with r ≤ 0.3.

Responsiveness is defined as the ability of an instrument to detect change over time in the construct to be measured[35]. Although not recommended by de Vet et al.[36], we calculated measures of responsiveness because this is common in many publications on measurement properties of hand function instruments[39]. For this purpose, we used the data of the subgroup(s) of patients (surgical and/or conservative) in whom, on the group level, a statistically significant change over time was seen for the majority of all outcome measures used. In this or these subgroup(s), effect sizes (Cohen's d) and standardized response means (SRMs) were calculated. An effect size of 0.2 is regarded as small, of 0.5 as medium, and of 0.8 as large[40]. In accordance with the recommendations of the COSMIN Group[35–37], we tested predefined hypotheses, similar to the approach we used for validity, i.e., 1) the effect size of the MHQ total score in a subgroup of improved patients is ≥ 0.8, and 2) the effect size of the MHQ total score is higher than the effect size of the DASH.

For interpretability, which is defined as the degree to which qualitative meaning can be ascribed to quantitative scores, we calculated the minimum important change (MIC). The MIC was defined as the smallest change that patients consider important and was calculated using an anchor-based method. For the anchor, we used the question about perceived change in the thumb condition at 1 year related to baseline. The MIC was calculated with receiver operating characteristic (ROC) curves and the optimal cut point, which reflects the MIC, was chosen for which was smallest ([1 - sensitivity] + [1 - specificity])[41]. The MIC should be higher than the SDC[36]. Furthermore, the area under the ROC curve (AUC) shows the ability of the MHQ to discriminate between improved and unimproved patients. A value of 0.5 indicates no discriminative ability, while an AUC ≥ 0.75 is regarded as appropriate[42].

Floor/ceiling effects were calculated from the percentage of patients showing the highest (100) or lowest (0) value in each subscale at baseline. If > 15% of the patients achieve the lowest/highest values, a floor/ceiling effect is present[43].

RESULTS

After screening 260 patients, we included 177 patients in our study (Figure 1). After inclusion, 3 patients scheduled for surgery cancelled their treatment. Nevertheless, their baseline data were analyzed. For the 1-year followup, we used data from 60 patients, 48 of whom completed the questionnaires twice (at the final visit and a few days later) for test–retest analysis. The mean age was 63.5 years and patients had been experiencing their symptoms for 2 years (median; range 0.2–40 years) (Table 1). Considering the returned baseline questionnaires, 2% of the MHQ items was initially missing (Table 2). The mean ± SD MHQ total score for surgical patients increased from 48 ± 14 at baseline to 75 ± 18 at 1 year (p ≤ 0.001). These patients also showed significant improvements in the DASH (p ≤ 0.001) and in the SF-12 physical health scores (p ≤ 0.001), whereas no significant improvements were seen in the SF-12 mental health scores (p = 0.71) and in key pinch (p = 0.64). In the conservative group, no treatment effect was observed since neither the MHQ, the MHQ subscales, the DASH, the SF-12, nor the key pinch showed statistically significant changes (p > 0.3 for all measures).
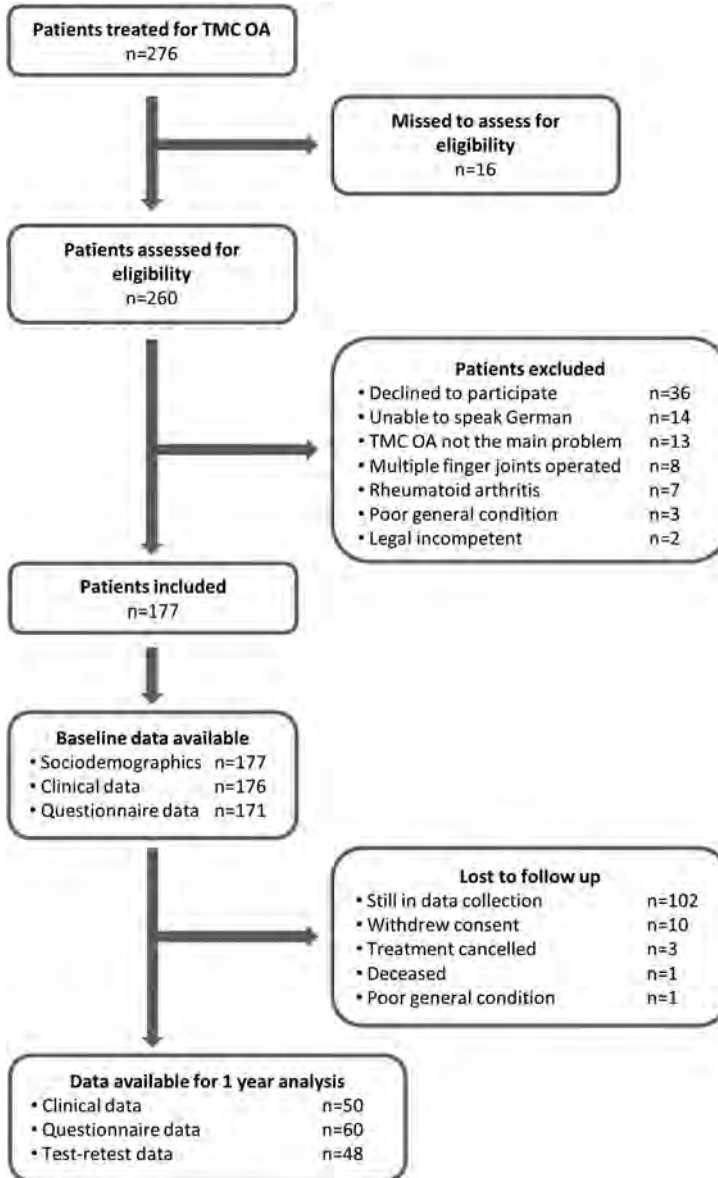
Figure 1 Patient selection diagram. TMC = trapeziometacarpal joint; OA = osteoarthritis

Reliability

Test–retest reliability was high for the MHQ and its subscales, with the ICC ranging between 0.85 (hand function and aesthetics) and 0.95 (total score) (Table 2). Internal consistency for the MHQ subscales showed a Cronbach's alpha range of 0.77 - 0.89. The measurement error of the MHQ total score (SEM) was 3.9 (Table 2), resulting in an SDC of 11 points (Table 3).

Table 1 Baseline characteristics of 177 patients with TMC OA*

Characteristic	Value
Female sex, no. (%)	145 (82)
Age, years	63.5 (9.2)
Symptom duration, median (range) years	2.0 (0.2-40)
Treatment, no. (%)	
Scheduled for surgery	109 (62)
Conservative	68 (38)
Drug intake; no. (%)	63 (37)
MHQ total score	53 (16)
MHQ hand function	55 (18)
MHQ ADL	56 (22)
MHQ work	56 (21)
MHQ pain	59 (18)
MHQ aesthetics	74 (26)
MHQ satisfaction	34 (21)
DASH score	43 (18)
SF-12 physical health score	39 (8)
SF-12 mental health score	50 (12)
Key pinch, kg	3.6 (2.1)

* Values are the mean ± SD unless indicated otherwise
TMC = trapeziometacarpal; OA = osteoarthritis; MHQ = Michigan Hand Outcomes Questionnaire; ADL = activities of daily living; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; SF-12 = Short Form 12 health survey

Table 2 Reliability, measurement error, and floor/ceiling effects of the MHQ and its subscales, the DASH and the SF-12*

	ICC	Cronbach's α	SEM	Floor effect, %	Ceiling effect, %	Missing items, %[†]
MHQ hand function	0.85	0.81	6.8	0.6	0.6	
MHQ ADL	0.89	0.89	7.0	0	1.8	
MHQ work	0.94	0.87	6.0	0	3.5	
MHQ pain	0.92	0.77	6.0	0.6	0	
MHQ aesthetics	0.85	0.86	8.5	0.6	19	
MHQ satisfaction	0.88	0.84	8.2	3.5	0.6	
MHQ total score	0.95		3.9	0	0	2
DASH	0.93	0.95	4.4	0	0	3.1
SF-12 physical	0.91		3.0	0	0	1.5[‡]
SF-12 mental	0.89		3.3	0	0	1.5[‡]

* MHQ = Michigan Hand Outcomes Questionnaire; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; SF-12 = Short Form 12 health survey; ICC = intraclass correlation coefficient; SEM = standard error of measurement; ADL = activities of daily living
[†] Missing items are given for each entire questionnaire
[‡] Based on all 12 items of the SF-12

Table 3 Responsiveness, minimum important change (MIC), and smallest detectable change for the MHQ subscales, DASH, SF-12, and key pinch strength for 35 patients with TMC joint OA who underwent surgery and attended followup after one year*

	Baseline score, mean ± SD	1-year score, mean ± SD	p	ES	SRM	MIC	SDC[†]
MHQ hand function	50 ± 19	73 ± 17	≤ 0.001	1.2	1.0	16	19
MHQ ADL	47 ± 20	77 ± 22	≤ 0.001	1.4	1.2	25	19
MHQ work	54 ± 17	70 ± 28	≤ 0.01	0.7	0.6	24	17
MHQ pain	64 ± 16	26 ± 23	≤ 0.001	1.9	1.8	19	17
MHQ aesthetics	71 ± 28	84 ± 22	≤ 0.01	0.5	0.4	1	24
MHQ satisfaction	30 ± 19	70 ± 23	≤ 0.001	1.9	1.5	30	23
MHQ total score	48 ± 14	75 ± 18	≤ 0.001	1.7	1.7	17	11
DASH	46 ± 15	26 ± 20	≤ 0.001	1.1	1.1	22	12
SF-12 physical	37 ± 9.0	45 ± 12	≤ 0.001	0.7	0.7	1	8
SF-12 mental	49 ± 14	50 ± 10	0.71	0.1	0.1	4	9
Key pinch, kg[‡]	3.5 ± 2.2	3.7 ± 2.0	0.64	0.1	0.1	1.5	

* MHQ = Michigan Hand Outcomes Questionnaire; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; SF-12 = Short Form 12 health survey; TMC = trapeziometacarpal; OA = osteoarthritis; ES = effect size; SRM = standardized response mean; ADL = activities of daily living
[†] Based on test–retest data from 48 patients [‡] N = 31

Validity

Two convergent construct validity hypotheses, i.e., correlation of MHQ ADL with the DASH ≤-0.7 and correlation of MHQ total score with the DASH ≤ -0.7, were verified, as the correlations of the MHQ ADL subscale and MHQ total score with the DASH were r=-0.76 and r = -0.77, respectively (Table 4). The pain subscale correlated only moderately well with the DASH (r = 0.67), which leads to the rejection of the other convergent construct validity hypothesis (correlation of MHQ pain with the DASH ≥ 0.7). The 2 discriminant construct validity hypotheses, i.e., correlation of MHQ hand function with key pinch strength between 0.5 ≥ r ≥ 0.3 and correlation of MHQ hand function with SF-12 mental health ≤ 0.3 were confirmed by the mild correlation between the hand function subscale and key pinch (r = 0.36) and the poor correlation between hand function and the SF-12 mental health score (r = 0.21), respectively.

Responsiveness

As there was no significant effect of conservative treatment for the MHQ or for any of the other outcome measures, effect size and SRM were only calculated for the surgical group (n = 35). The effect size of the MHQ total score was 1.7 (Table 3). The two hypotheses regarding responsiveness (effect size MHQ total score ≥ 0.8 and effect size MHQ total score greater than the effect size of the DASH) were therefore verified.

FOUR

Table 4 Construct validity: correlations of the MHQ and its subscales with the DASH, SF-12, and key pinch strength using baseline data*

	MHQ							DASH	SF-12 physical	SF-12 mental	Key pinch (kg)
	Hand function	ADL	Work	Pain	Aesthetics	Satisfaction	Total score				
MHQ hand function	1										
MHQ ADL	0.54†	1									
MHQ work	0.33†	0.57†	1								
MHQ pain	-0.41†	-0.67†	-0.63†	1							
MHQ aesthetics	0.32†	0.27†	0.28†	-0.23†	1						
MHQ satisfaction	0.62†	0.64†	0.49†	-0.69†	0.40†	1					
MHQ total score	0.70†	0.81†	0.73†	-0.79†	0.59†	0.85†	1				
DASH	-0.44†	-0.76§	-0.65†	0.67§	-0.30†	-0.64†	-0.77§	1			
SF-12 physical	0.23*	0.45†	0.56†	-0.46†	0.12	0.38†	0.48†	-0.63†	1		
SF-12 mental	0.21¶	0.30†	0.39†	-0.30†	0.32†	0.35†	0.43†	-0.36†	0.03	1	
Key pinch (kg)	0.36§	0.44†	0.25†	-0.32†	0.31†	0.41†	0.46†	-0.36†	0.17#	0.08	1

* MHQ = Michigan Hand Outcomes Questionnaire; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; SF-12 = Short Form 12 health survey; ADL = activities of daily living

† p ≤ 0.001

‡ p ≤ 0.01

§ p ≤ 0.001; correlations relevant for hypotheses testing

¶ p ≤ 0.01; correlations relevant for hypotheses testing

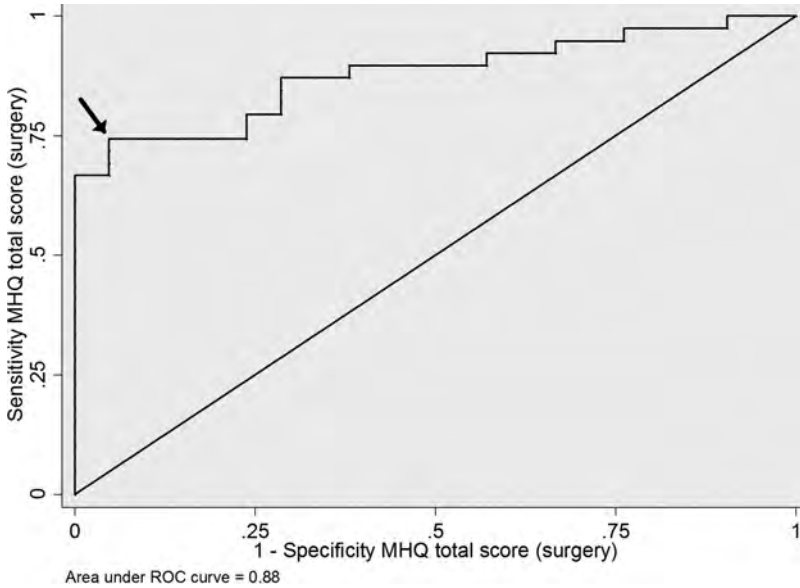# p ≤ 0.05

Area under ROC curve = 0.88

Figure 2 Receiver operating characteristic (ROC) curve for the Michigan Hand Outcomes Questionnaire (MHQ) total score at 1 year for 35 surgical patients

Interpretability

The AUC for the MHQ total score was 0.88 for surgical patients and the resulting MIC was 17 points (Figure 2 and Table 3), which is larger than the SDC of 11 points. We found a ceiling effect for the aesthetics subscale but no floor/ceiling effects were present for the other subscales and the MHQ total score (Table 2).

DISCUSSION

The results of this study provide evidence that the MHQ demonstrates good reliability, validity, and responsiveness in the assessment of patients with TMC joint OA. Regarding reliability and validity, our data support the excellent test–retest reliability of the MHQ that has already been shown in other studies[10, 20]. In our study, internal consistency was satisfactory, whereas item redundancy was apparent in other studies[18].

According to Terwee et al.[43], construct validity can be rated positively if predefined hypotheses are tested and if at least 75% of the results are in correspondence with the hypotheses. As we were able to support 4 out of the 5 hypotheses, we concluded that the MHQ demonstrates good validity for the assessment of patients with TMC joint OA. However, our hypothesis that the MHQ pain subscale correlates highly with the DASH had to be rejected, even though the correlation coefficient of 0.67 was quite strong. This slightly weaker correlation could be due to the fact that only 3 items out of 30 in the DASH are about pain, while the other 27 items concern ADL. Other studies[10, 12] investigating patients with various hand disorders found similar, but somewhat poorer, correlations between the MHQ and the DASH.

In terms of responsiveness, large effect sizes in the surgical group were shown for the MHQ total score, as well as for the pain and satisfaction subscales. The lowest effect sizes seen in our group were related to the MHQ aesthetics subscale. This fact, combined with the relatively high baseline scores and the ceiling effect, indicates that the appearance of the hand may not be as important to patients with TMC joint OA as it is to patients with rheumatoid arthritis[18]. On average, patients who underwent metacarpophalangeal joint arthroplasty for rheumatoid arthritis had baseline values in the MHQ aesthetics subscale 40 points lower than our patients, and in that group the SRM of 1.2 was very high[18].

Our data show a higher effect size and SRM of the MHQ total score than the DASH. Better responsiveness of the MHQ compared with the DASH has also been shown in other studies investigating patients with finger injuries[24], carpal tunnel syndrome, and wrist pain[21]. The poorer responsiveness of the DASH might be because the score is influenced by function/dysfunction of the elbow and shoulder joints.

The medium effect size of the SF-12 physical health score indicates that patients who had undergone surgery for their TMC joint OA also experienced a moderate improvement in their quality of life. The SF-12 should not be used as a single outcome measure in patients with thumb or hand OA, but rather it is recommended as an additional tool to investigate the impact of treatment on the patient's quality of life perception[31, 44].

Regarding interpretability, the present study showed that the MHQ allows an appropriate distinction between improved and unimproved patients. The large AUC attests to the discriminative ability of the MHQ. Similar AUCs were found for patients with rheumatoid arthritis following silicone metacarpophalangeal joint arthroplasty and carpal tunnel syndrome[19]. However, we found MIC values in our population different from those reported by Shauver and Chung in their patients mentioned previously[19]. Possible reasons for this are the different conditions in the patient groups and the disparate methods used to calculate the MIC. Shauver and Chung[19] used the satisfaction subscale of the MHQ as an anchor for the ROC curve, whereas we used an additional question regarding perceived change of the thumb condition as the external criterion, since this is recommended in the literature[36, 45].

Beside measurement properties, other aspects such as the administration mode and associated costs have to be considered when choosing a questionnaire. The time to complete the MHQ is between 8 and 20 minutes[20, 46] and the questionnaire with the scoring algorithm as well as an Excel scoring sheet is freely available[28]. Patients perceived the MHQ to be more complex to understand and complete than, for example, the DASH[46]. In order to avoid these issues, the BriefMHQ has recently been developed[16, 17]. The BriefMHQ shows similar measurement properties to the original version in a population including patients with rheumatoid arthritis, TMC joint OA, carpal tunnel syndrome, and distal radius fracture[16]. However, the BriefMHQ is not able to produce subscale scores or distinguish between the 2 hands. It is intended as a more efficient tool for clinical settings but not for research[47]. Despite indicating item redundancy, use of the original MHQ is still advocated, as it provides a more comprehensive analysis of the patient's condition[16]. In addition, the full MHQ can assess the 2 hands separately, so that stratification for hand dominance or the affected hand is possible[18]. Overall, the advantages regarding measurement properties, multidimensionality, and hand differentiation of the original MHQ may predominate over its brief version in scientific settings.

FOUR

This study has some limitations. For the test–retest analysis, data from only 48 patients were available. For responsiveness and MIC, only the data from the 35 surgical patients were used because there was no statistical treatment effect in the conservative group. This approach reduced the sample size and the transferability of the results to patients treated conservatively. As we only intended to study reliability of the MHQ and not of key pinch, we were not able to show data for the ICC, SEM, and SDC for the latter. However, previous studies have indicated high test– retest reliability of key pinch with r being > 0.8[48, 49]. Furthermore, our patients had different surgical and conservative treatment. For that reason, we cannot draw any conclusions about the effect of a specific treatment option, which was, however, beyond the scope of this study. Further comparisons with other hand-specific questionnaires such as the Australian/Canadian Hand Osteoarthritis Index[50], the Patient-Rated Wrist Evaluation[51], and the Patient Evaluation Measure[52] are indicated in order to find the best questionnaire for each purpose and target population.

In conclusion, this study evaluated the measurement properties of the MHQ with the help of the DASH and SF-12. For patients with TMC joint OA who underwent surgery or who were conservatively treated for their condition, our results indicate good reliability and validity. Additionally, the MHQ demonstrated high responsiveness for the patients who underwent surgery. Based on these results, we can recommend the MHQ as a suitable assessment instrument for patients with TMC joint OA.

FOUR

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Ms Marks had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Marks, Audigé, Herren, Schindele, Nelissen, Vliet Vlieland.

Acquisition of data. Marks, Herren, Schindele.

Analysis and interpretation of data. Marks, Audigé, Vliet Vlieland.

REFERENCES

1. Kalichman L, Hernandez-Molina G. Hand osteoarthritis: an epidemiological perspective. Semin Arthritis Rheum 2010;39:465-76.

2. Wilder FV, Barrett JP, Farina EJ. Joint-specific prevalence of osteoarthritis of the hand. Osteoarthritis Cartilage 2006;14:953-57.

3. Sodha S, Ring D, Zurakowski D, Jupiter JB. Prevalence of osteoarthrosis of the trapeziometacarpal joint. J Bone Joint Surg Am 2005;87:2614-18.

4. Kjeken I, Dagfinrud H, Slatkowsky-Christensen B, Mowinckel P, Uhlig T, Kvien TK, et al. Activity limitations and participation restrictions in women with hand osteoarthritis: patients' descriptions and associations between dimensions of functioning. Ann Rheum Dis 2005;64:1633-8.

5. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996;29:602-48.

6. Marks M, Schoones J, Kolling C, Herren D, Goldhahn J, Vliet Vlieland T. Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis – A systematic literature review. J Hand Surg Eur Vol 2013;38:822 – 38

7. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am 1998;23:575-87.

8. Shauver MJ, Chung KC. The michigan hand outcomes questionnaire after 15 years of field trial. Plast Reconstr Surg 2013;131:779e-87e.

9. Van der Giesen FJ, Nelissen RG, Arendzen JH, de Jong Z, Wolterbeek R, Vliet Vlieland TP. Responsiveness of the Michigan Hand Outcomes Questionnaire--Dutch language version in patients with rheumatoid arthritis. Arch Phys Med Rehabil 2008;89:1121-26.

10. Öksüz C, Akel BS, Oskay D, Leblebicioglu G, Hayran KM. Cross-cultural adaptation, validation, and reliability process of the Michigan Hand Outcomes Questionnaire in a Turkish population. J Hand Surg Am 2011;36:486-92.

11. Knobloch K, Kuehn M, Papst S, Kraemer R, Vogt PM. German standardized translation of the michigan hand outcomes questionnaire for patient-related outcome measurement in dupuytren disease. Plast Reconstr Surg 2011;128:39e-40e.

12. Roh YH, Yang BK, Noh JH, Baek GH, Song CH, Gong HS. Cross-cultural adaptation and validation of the Korean version of the Michigan hand questionnaire. J Hand Surg Am 2011;36:1497-503.

13. Pinho N, Rafael A, Cardoso A, Nardi A. Translation and cross-cultural adaptation of the Michigan Hand Outcomes Questionnaire. J Bras Psiquiatr 2011;60:99-110.

14. Zyluk A, Piotuch B. [Use of questionnaires in outcome measurement in hand surgery] Chirurgia Narzadów Ruchu I Ortopedia Polska 2009;74:193-201.

15. Huijsmans R, Sluiter H, Aufdemkampe G. [Michigan Hand Outcomes Questionnaire-Dutch Language Version; een vragenlijst voor patienten met handfunctieproblemen.] Fysiopraxis 2001;9:38-41.

16. Waljee JF, Kim HM, Burns PB, Chung KC. Development of a brief, 12-item version of the michigan hand questionnaire. Plast Reconstr Surg 2011;128:208-20.

17. Knobloch K, Kraemer R, Papst S, Sorg H, Vogt PM. German version of the brief Michigan Hand Outcomes Questionnaire: implications for early quality of life following collagenase injection in dupuytren contracture. Plast Reconstr Surg 2012;129:886e-87e.

18. Waljee JF, Chung KC, Kim HM, Burns PB, Burke FD, Wilgis EF, et al. Validity and responsiveness of the Michigan Hand Questionnaire in patients with rheumatoid arthritis: a multicenter, international study. Arthritis Care Res (Hoboken) 2010;62:1569-77.

19. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. J Hand Surg Am 2009;34:509-14.

20. Massy-Westropp 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.

21. McMillan CR, Binhammer PA. Which Outcome Measure is the Best? Evaluating Responsiveness of the Disabilities of the Arm, Shoulder, and Hand Questionnaire, the Michigan Hand Questionnaire and the Patient-Specific Functional Scale Following Hand and Wrist Surgery. Hand (N Y) 2009;4:311-18.

22. Kotsis SV, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder and Hand questionnaire in carpal tunnel surgery. J Hand Surg Am 2005;30:81-86.

23. Kotsis SV, Lau FH, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and physical measurements in outcome studies

FOUR

of distal radius fracture treatment. J Hand Surg Am 2007;32:84-90.

24. Horng YS, Lin MC, Feng CT, Huang CH, Wu HC, Wang JD. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder, and Hand questionnaire in patients with hand injury. J Hand Surg Am 2010;35:430-6.

25. Chung KC, Hamill JB, Walters MR, Hayward RA. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. Ann Plast Surg 1999;42:619-22.

26. Van Rijn J, Gosens T. A cemented surface replacement prosthesis in the basal thumb joint. J Hand Surg Am 2010;35:572-79.

27. Burke NG, Walsh J, Moran CJ, Cousins G, Molony D, Kelly EP. Patient-reported outcomes after Silastic replacement of the trapezium for osteoarthritis. J Hand Surg Eur Vol 2012;37:263-68.

28. University of Michigan Medical School, Department of Surgery. 2013. Scoring the MHQ. http://sitemaker.umich.edu/mhq/scoring.

29. Offenbaecher M, Ewert T, Sangha O, Stucki G. Validation of a German version of the disabilities of arm, shoulder, and hand questionnaire (DASH-G). Z Rheumatol 2003;62:168-77.

30. Ware JE, Kosinski M, Gandek B, Sundaram M, Bjorner JB, Turner-Bowker DM, et al. User's manual for the SF-12v2 Health Survey 2nd edition. Lincoln, RI, QualityMetric Incorporated 2010.

31. Angst F, John M, Goldhahn J, Herren DB, Pap G, Aeschlimann A, et al. Comprehensive assessment of clinical outcome and quality of life after resection interposition arthroplasty of the thumb saddle joint. Arthritis Rheum 2005;53:205-13.

32. Angst F, Goldhahn J, Drerup S, Flury M, Schwyzer HK, Simmen BR. How sharp is the short QuickDASH? A refined content and validity analysis of the short form of the disabilities of the shoulder, arm and hand questionnaire in the strata of symptoms and function and specific joint conditions. Qual Life Res 2009;18:1043-51.

33. John M, Angst F, Awiszus F, Pap G, Macdermid JC, Simmen BR. The patient-rated wrist evaluation (PRWE): cross-cultural adaptation into German and evaluation of its psychometric properties. Clin Exp Rheumatol 2008;26:1047-58.

34. MacDermid JC, Wessel J, Humphrey R, Ross D, Roth JH. Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. Osteoarthritis Cartilage 2007;15:524-30.

35. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737-45.

36. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge: Cambridge University Press 2011.

37. 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.

38. Mokkink LB. Cosmin Checklist Manual. 2012. http://www.cosmin.nl/the-cosmin-checklist_8_5.html.

39. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol 2011;11:152

40. Cohen J. A power primer. Psychol Bull 1992;112:155-59.

41. De Vet HC, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RW, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. J Clin Epidemiol 2010;63:37-45.

42. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. CJEM 2006;8:19-20.

43. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.

44. Michon M, Maheu E, Berenbaum F. Assessing health-related quality of life in hand osteoarthritis: a literature review. Ann Rheum Dis 2011;70:921-38.

45. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-09.

46. Dias JJ, Rajan RA, Thompson JR. Which questionnaire is best? The reliability, validity and ease of use of the Patient Evaluation Measure, the Disabilities of the Arm, Shoulder and Hand and the Michigan Hand Outcome Measure. J Hand Surg Eur Vol 2008;33:9-17.

47. University of Michigan Medical School, Department of Surgery. BriefMHQ. http://sitemaker.umich.edu/mhq/brief_mhq.

48. Myers HL, Thomas E, Hay EM, Dziedzic KS. Hand assessment in older adults with musculoskeletal

FOUR

hand problems: a reliability study. BMC Musculoskelet Disord 2011;12:3.

49. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. J Hand Surg Am 1984;9:222-26.

50. Bellamy N, Campbell J, Haraoui B, Buchbinder R, Hobby K, Roth JH, 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.

51. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther 1996;9:178-83.

52. Macey AC, Burke FD, Abbott K, Barton NJ, Bradbury E, Bradley A, et al. Outcomes of hand surgery. British Society for Surgery of the Hand. J Hand Surg Br 1995;20:841-55.

FOUR