Reliability and Validity Testing of the Michigan Hand Outcomes Questionnaire

Kevin C. Chung, MD, Matthew S. Pillsbury, BS, Madonna R. Walters, BSN, Rodney A. Hayward, MD, Ann Arbor, MI

In this study, psychometric principles were used to develop an outcomes questionnaire capable of measuring health state domains important to patients with hand disorders. These domains were hypothesized to include (1) overall hand function, (2) activities of daily living (ADL), (3) pain, (4) work performance, (5) aesthetics, and (6) patient satisfaction with hand function. An initial pool of 100 questions was pilot-tested for clarity in 20 patients; following factor analysis, the number of questions was reduced to a 37-item Michigan Hand Outcomes Questionnaire (MHQ). The MHQ, along with the Short Form-12, a generic health status outcomes questionnaire, was then administered to 200 consecutive patients at a universitybased hand surgery clinic and was subjected to reliability and validity testing. The mean time required to complete the questionnaire was 10 minutes (range, 7-20 minutes). Factor analysis supported the 6 hypothesized scales. Test-retest reliability using Spearman's correlation demonstrated substantial agreement, ranging from 0.81 for the aesthetics scale to 0.97 for the ADL scale. In testing for internal consistency, Cronbach's alphas ranged from 0.86 for the pain scale to 0.97 for the ADL scale (values >0.7 for Cronbach's alpha are considered a good internal consistency). Correlation between scales gave evidence of construct validity. In comparing similar scales in the MHQ and the Short Form-12, a moderate correlation (range, 0.54-0.79) for the ADL, work performance, and pain scales was found. In evaluating the discriminate validity of the aesthetics scale, a significant difference (p = .0012) was found between the aesthetics scores for patients with carpal tunnel syndrome and patients with rheumatoid arthritis. The MHQ is a reliable and valid instrument for measuring hand outcomes. It can be used in a clinic setting with minimal burden to patients. The questions in the MHQ have undergone rigorous psychometric testing, and the MHQ is a promising instrument for evaluation of outcomes following hand surgery. (J Hand Surg 1998;23A:575-587. Copyright © 1998 by the American Society for Surgery of the Hand.)

Upper extremity injuries are common in the United States. The National Center for Health Sta-

Supported by the Robert Wood Johnson Foundation and the Department of Veterans Affairs Health Services Research and Development. Dr Chung was a Robert Wood Johnson Clinical Scholar at the

University of Michigan at the time this study was conducted.

tistics Health Interview Survey (1975–1976) estimated that 16 million upper extremity injuries occur each year, resulting in 16 million days lost from work.¹ In 1980, the estimated total cost of upper extremity disorders, including both direct and indirect costs, exceeded \$10 billion.² Because of the

From the Section of Plastic and Reconstructive Surgery, Department of Surgery, University of Michigan Medical Center, Ann Arbor, MI; Health Services Research and Development, Department of Veterans Affairs, Ann Arbor, MI; the Trauma Program, St. Joseph Mercy Hospital, Ann Arbor, MI; and the Departments of Internal Medicine and Health Management and Policy. VA Center for Practice Management and Outcomes Research, Veterans Administration Medical Center, Ann Arbor, MI.

Received for publication June 9, 1997; accepted in revised form April 10, 1998.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Reprint requests: Kevin C. Chung, MD, Section of Plastic and Reconstructive Surgery, University of Michigan Medical Center, 1500 E Medical Center Dr, 2130 TC, Ann Arbor, MI 48109-0340.

Copyright © 1998 by the American Society for Surgery of the Hand 0363-5023/98/23A04-0018\$3.00/0

576 Chung e al. / Michigan Hand Outcomes Questionnaire

huge social and economic costs of treating upper extremity disorders, evaluating treatment outcomes of hand disorders should be an important component in the current government and private-sector initiatives to assess quality in the American health care system. In contrast to well-defined treatment outcomes such as mortality, outcomes in hand surgery are more difficult to assess because the major area of interest is improvement in quality of life and function, variables that are hard to quantify. In 1983, Robert A. Chase,³ on behalf of the Socioeconomic Committee of the American Society for Surgery of the Hand, recommended the development of "a universally available system for measurement of outcome of disorders treatable by surgery." He also stated that "patient questionnaires and measurement checklists need to be developed and validated."

To measure outcomes pertinent to patients with hand disorders, we developed a hand-specific outcomes instrument, the Michigan Hand Outcomes Questionnaire (MHQ), to assess outcomes that were considered important by a panel of patients with hand disorders, hand therapists, and hand surgeons. This report describes the development of the MHQ and its initial psychometric testing for reliability and validity. Our goal was to develop a questionnaire to be used as a standardized instrument capable of measuring outcomes for patients with all types of hand disorders.

Materials and Methods

To construct a preliminary version of the MHQ, we used a MEDLINE search to evaluate existing questionnaires containing items related to the measurement of upper extremity function. Of the questionnaires evaluated, the Short Form-36 (SF-36)⁴ and the Arthritis Impact Measurement Scale⁵ contained hand-related work performance and physical function questions; we incorporated these questions into the preliminary MHQ. The McGill Pain Scale⁶ and the Carpal Tunnel Questionnaire⁷ were used as guides for the development of a list of questions that quantified pain. In addition, a group of patients with hand disorders generated additional items that they considered important to hand function. One hundred items were generated for the early versions of the MHQ. A panel comprising patients with hand disorders, hand therapists, and hand surgeons evaluated this preliminary list of 100 questions and hypothesized that the questions would fit into 6 scales: (1) overall hand function, (2) activities of daily living (ADLs), (3) pain, (4) work performance, (5) aesthetics, and (6) patient satisfaction with hand function.

Two psychometricians, both experienced in questionnaire design, reviewed the preliminary version of the MHQ for question structure and clarity. Redundant questions were eliminated and ambiguous wording in some questions was modified. The question response categories in each scale were derived from a study that assessed the magnitude of response categories, which reflected the numerical values given by a group of subjects during psychometric testing.⁸ For example, in the pain scale, the response category (always, often, sometimes, rarely, and never) represented nearly equal numerical intervals between the responses.

We used factor analysis to further decrease the number of questions in each scale. Factor analysis is a statistical technique used in the social sciences to reduce the number of items in a questionnaire, but still allows the reduced questionnaire to accurately measure the domains or factors of interest.9-11 For example, in the pain scale, factor analysis revealed that a question regarding duration of pain had a response pattern that was not congruent with other questions in measuring pain severity. Therefore, that question could be eliminated without affecting the overall measurement of pain. Likewise, a question regarding whether pain interfered with social activities was removed because it measured a factor other than pain. Using the factor analytic technique, we further reduced the questionnaire to 37 distinct items that would measure the 6 domains of interest. Within each scale, factor analysis showed that all the remaining questions focused on only 1 major factor, which indicated that all the questions in each scale measured only the factor of interest.

Of the 6 scales in the MHQ, 4 (overall hand function, ADLs, aesthetics, and satisfaction with hand function) contain items that measure both rightand left-hand outcomes. Because items relating to hand function are confounded by hand dominance, measuring the function of both right and left hands will control the confounding effect of hand dominance and preserve the richness of the data. A previous study has shown that hand dominance can significantly affect responses to various tasks.¹² Right-hand dominant persons with right-hand disorders obviously have more difficulty performing tasks than if their left hands are injured. Therefore, by evaluating the function in each hand, the MHQ can accurately measure the performance of the affected hand. In the ADL scale, an item regarding writing

was eliminated because of hand-dominance issues. In the aesthetics scale, measuring both hands is particularly pertinent to rheumatoid patients who have varying degrees of deformities in their hands.

After questionnaire revisions and pilot testing, the MHQ was administered to 200 consecutive patients at a university-based hand clinic. While waiting for their appointments, patients were asked to fill out the MHQ and the Short Form-12 (SF-12),¹³ a shorter version of the SF-36. Because the MHQ and the SF-12 contain several comparable questions for physical function, items in the SF-12 were used to validate scales in the MHQ.

Patients were included in the study if they were new patients to the clinic and if they had hand symptoms of at least 3 months' duration. Recruiting only new patients for this study eliminated potential selection bias using existing patients who were undergoing treatment. Patients were excluded if they were (1) younger than 18 years (because of consenting issues in minors), (2) if their hands were currently immobilized in splints or casts, or (3) if they were unable to complete the questionnaire in English.

To test for reliability and validity, the responses from the 200 patients were analyzed using the Stata statistical package.¹⁴ For test-retest reliability, 25 patients completed the initial questionnaire in the clinic and a second identical questionnaire approximately 1 week later. The 1-week time interval was chosen to minimize recall bias. Because all the patients did not return for a visit the following week, they were asked to complete the second questionnaire at home. As a result, we received 3 questionnaires that were completed just 1 day after the completion of the initial questionnaires. These 3 questionnaires were excluded from the analysis. The scores for each of the 6 scales were correlated for the first and the second administrations using interclass correlation.15

Scoring Mechanism

Multiple linear regression was initially used to derive weights for each of the questions in a scale. We also scored the MHQ using the unweighted method, by adding the responses (ranging from 1 to 5) in each scale, and normalizing the scores to a scale from 0 to 100 (Appendix 1). Correlations between these 2 scoring mechanisms showed correlations over 0.99 in all the scales. Because the weighted method is cumbersome and does not present a clear advantage, we used the unweighted method for scoring the MHQ. In 5 of the scales (overall hand function, ADLs, work performance, aesthetics, and satisfaction with hand function), higher scores indicated better hand performance while lower scores indicated poorer hand performance. In the pain score, higher scores indicated greater pain while lower scores indicated less pain.

Reliability

Two types of reliability (test-retest and internal consistency) were evaluated for the MHQ. Testretest reliability means repeatability, the extent to which an experiment, test, or any measuring procedure yields the same result in repeated trials.¹⁶ For example, a Jamar dynamometer is considered reliable if grip strength testing on the same patient during 2 successive trials yields the same value; therefore, the dynamometer provides consistent results across repeated measures. Internal consistency reliability measures the homogeneity of the items comprising a scale; that is, whether the items in a scale are highly correlated with each other. High inter-item correlations suggest that the items in a scale are all measuring the same thing.¹⁷ We also calculated the mean difference in the scores between the 2 administrations as a test of agreement. A mean difference of 0 indicates perfect agreement in the test-retest.

Validity

Validity assesses whether the instrument actually measures the phenomena of interest. In other words, if the MHQ is designed to measure various dimensions of hand function, patients who score poorly on this questionnaire (worse function) will have greater difficulty in performing various hand-related tasks. Although many different types of validity have been proposed in the social science literature, 3 types of validity (content, criterion, and construct validity) are important general categories.

Content Validity

Content validity is sometimes referred to as face validity. It indicates whether the questionnaire instrument appears logical to a group of experts. For example, a hand outcomes questionnaire will not have content validity if it does not contain questions that measure performance of ADLs. A panel of hand surgeons, hand therapists, and patients with hand disorders evaluated the MHQ for content validity. It also was subjected to extensive pilot testing to derive the 6 scales, which the panel regarded as important and pertinent in measuring hand outcomes.

Criterion Validity

Criterion validity assesses how an instrument measures up against a well-accepted "gold standard." For example, in evaluating the validity of ventilation/ perfusion scans in the diagnosis of pulmonary embolism, criterion validity can be assessed by comparing ventilation/perfusion scan findings to pulmonary angiograms, which are considered the "gold standard" for the diagnosis of pulmonary embolism. However, in outcomes questionnaire testing, a "gold standard" generally is not available for comparison because outcomes questionnaires are designed to measure patients' perceptions of their health and function, concepts that cannot be quantified by any other means. Therefore, criterion validity cannot be used to establish the validity of the MHQ.

Construct Validity

In contrast to criterion validity, construct validity is a major component in the testing of all outcomes instruments. Construct validity means that the scales in the questionnaire behave as expected. For example, in the MHQ, patients with poor overall hand function are expected to have poor performance in their work. Similarly, patients who perform well in their ADLs should be more satisfied with their hand function. By establishing the theoretical relationships between scales, we were able to test the validity of the MHQ. We established these theoretical relationships *a priori* and included here all construct validity tests performed.

We also evaluated items that were similar in both the MHQ and the SF-12. The SF-12 contained generic questions that related to pain and upper extremity function. We hypothesized that the generic items in the SF-12 would have moderate correlation to the scales in the MHQ. Using an existing validated questionnaire, such as the SF-12, helped to establish construct validity in the MHQ.

Results

Two hundred consecutive eligible patients completed the MHQ (Appendix 2). The response rate was 99%. Table 1 lists the demographic data.

Reliability Testing

Test-retest reliability. Twenty-two patients completed the initial questionnaire in the clinic and a

Table 1. Demographic Data for the 200 Patients
Who Completed the Michigan Hand Outcomes
Questionnaire

2 aconomiano	
Mean age, yr (mean ± SD)	45 ± 14.4
Male sex (%)	53.0
Right-hand dominant (%)	85.5
Injured hand (%)	
Right	45.0
Left	45.5
Both	9.5
Ethnic background (% white)	82.5
Education (%)	
High school	30.0
<college< td=""><td>47.5</td></college<>	47.5
>College	22.5
Income (%)	
<\$20,000	28.5
\$20,000-\$49,000	35.5
>\$50,000	36.0
Worker's Compensation (%)	67.0

second identical questionnaire approximately 1 week later. Table 2 presents the results of the intraclass correlation for the test–retest reliability.¹⁵

Test-retest showed excellent correlation in the 6 hypothesized scales. Scores of 1.0 indicate perfect correlation and scores of 0 indicate no correlation. Five of the 6 scales had correlation scores over 0.85. The aesthetics scores had slightly lower correlation, which might reflect fluctuations in a patient's attitude about body image and sense of beauty.

To measure whether the scores between the 2 MHQ administrations agreed with each other, we

Table 2. Test-Retest Correlation for the Six Scales in the Michigan Hand Outcomes Questionnaire (N = 22)

		Intraclass Correlation ¹⁵
1.	Overall hand function	
	Right hand	0.89
	Left hand	0.95
2.	Activities of daily living	
	Right hand	0.95
	Left hand	0.97
	Both hands	0.94
3.	Work performance	0.93
4.	Pain	0.91
5.	Aesthetics	
	Right hand	0.90
	Left hand	0.81
6.	Satisfaction with hand function	
	Right hand	0.96
	Left hand	0.96

calculated the mean difference in the scores between the 2 administrations (Table 3). Mean differences in the scores that are closer to 0 indicate better agreement.¹⁸ As shown in Table 3, the mean difference in the scores between the 2 administrations were less than 5 points in 10 of 11 scales (based on a scoring scheme of 0 to 100). The 95% confidence intervals were close to 0 in all the scales. The excellent agreement in the scores gave further support to the test– retest reliability of the MHQ.

Internal consistency. Internal consistency measures how well all the questions in a scale are correlated with each other, and high interitem correlations may suggest that all the questions measure the factor of interest. For example, in the aesthetics scale, all the questions should contribute to measuring the aesthetics factor. If each of the questions in the scale measure a factor other than aesthetics, the internal consistency for the aesthetic scale will be poor. Internal consistency is expressed by Cronbach's alpha, which is a measure of the reliability of the summative rating scale.¹⁹ Cronbach's alphas can range from 0 to 1.0, where 1.0 indicates perfect internal consistency. Generally, Cronbach's alphas of greater than 0.80 in a scale are considered acceptable.²⁰ Table 4 summarizes the alphas for the 6 scales.

Table 3. Limits of Agreement in the Test-Retest
for the Michigan Hand Outcomes Questionnaire
(N = 22)

_					
			95% Confidence Interval		
		Mean			
		Difference*			
1.	Overall hand function				
	Right hand	2.42	-3.23	8.07	
	Left hand	3.80	-0.30	7.90	
2.	Activities of daily living				
	Right hand	-0.92	-5.81	3.96	
	Left hand	2.27	-1.22	5.76	
	Both hands	6.03	3.45	8.71	
3.	Work performance	-1.00	-5.83	3.83	
4.	Pain	-0.50	-5.77	4.76	
5.	Aesthetics				
	Right hand	1.80	-8.65	5.04	
	Left hand	4.50	-6.01	15.00	
6.	Satisfaction with hand function				
	Right hand	-2.75	-7.69	2.19	
	Left hand	-0.82	-5.00	3.35	

* The mean difference represents the difference in the secres between the first and second administration of the Michigan Hand Outcomes Questionnaire (based on a scoring scheme from 0 to 100).

Table 4.	Cronbach's	Alpha for	the Six Scales in
the Mic	higan Hand	Outcomes	Questionnaire

		Cronbach's Alphas
Ι.	Overall hand function	
	Right hand	0.93
	Left hand	0.93
2.	Activities of daily living	
	Right hand	0.96
	Left hand	0.97
	Both hands	0.95
3.	Work performance	0.94
4.	Pain	0.86
5.	Aesthetics	
	Right hand	0.90
	Left hand	0.87
6.	Satisfaction with hand function	
	Right hand	0.93
	Left hand	0.94

All the scales in the MHQ had Cronbach's alphas greater than 0.85, and 9 of the 11 scales had Cronbach's alphas greater than 0.90. The MHQ had excellent internal consistency in all the scales.

Construct validity. We hypothesized that the functional scales in the MHQ (overall hand function, ADLs, work performance, pain, and satisfaction with hand function) would be significantly correlated with each other. For example, patients with poor overall hand function would have poor work performance. Likewise, patients with more pain in their hands would have difficulties performing ADLs. Table 5 lists the Spearman's rank correlation for the 6 scales in the MHQ.

There is high correlation among the 5 scales that measure functional outcomes: overall hand function, ADLs, work performance, pain, and satisfaction with hand function. The overall hand function scale was highly correlated with the other functional scales. Patients who perceive that their hands are functioning well perform well in the ADLs and in their work, and also have less pain and are more satisfied with their hand performance. Similarly, patients with increased pain in their hands did poorly in all the other scales. Therefore, the 5 functional scales had correlation in the expected direction. The aesthetics scale showed weaker correlation with the other scales because the aesthetics scale measured a factor that was distinct from the other functional scales.

Three of the scales in the MHQ (ADLs, work performance, and pain) were compared with similar questions in the SF-12, which asked about physical

	Overall Hand Function	Activities of Daily Living	Work Performance	Paint	Aesthetics	Satisfaction With Hand Function
Overall hand function	1.00					
Activities of daily living	0.78	1.00				
Work performance	0.69	0.74	1.00			
Pain†	-0.65	-0.69	-0.72	1.00		
Aesthetics	0.36	0.46	0.35	-0.29	1.00	
Satisfaction with hand function	0.75	0.73	0.63	-0.74	0.32	1.00

Table 5. Spearman's Correlation for the Six Scales in the Michigan Hand Outcomes Questionnaire*

* Four scales (overall hand function, activities of daily living, aesthetics, and satisfaction with hand function) represent the scores for the injured hand. All the correlations are statistically significant at p < .0001, with adjustment for multiple comparison using Bonferroni's method.²⁹

[†] The negative sign in the pain scale indicates reverse scoring: patients with a higher pain score (more pain) have worse hand performance.

limitations due to health. Because questions in the SF-12 inquired about health and not specifically about hand performance, we hypothesized that there would only be moderate correlation between similar items in these 2 questionnaires. Table 6 lists the correlation. The ADL and work performance scales had moderate correlation with items in the SF-12. The pain scale had substantial correlation (0.79) with the pain question in the SF-12.

In the validity testing for the aesthetics scale, we compared patients with carpal tunnel syndrome and patients with rheumatoid arthritis who had severe hand deformities. Carpal tunnel syndrome patients were selected if they had no physical changes, such as thenar atrophy, or if they had not had prior hand operations. Rheumatoid hand patients were selected if they had greater than 30° of ulnar deviation to the fingers or if they had severe destruction of their wrists. We hypothesized that rheumatoid hand patients with hand deformities would have significantly

lower aesthetic scores than the carpal tunnel syndrome patients. We used the unpaired Student's *t*-test to compare the mean aesthetics scores between the 2 groups. The mean aesthetics score for the carpal tunnel syndrome patients (83.7) differed significantly from that of the rheumatoid arthritis patients (50.4) (p = .0012).

To determine which scales in the MHQ were significant predictors of physical function, we regressed the scales in the MHQ against the physical function component of the SF-12 (Table 7). Of the 6 scales in the MHQ, the pain scale was the strongest predictor of physical function (p = .001) and the aesthetics scale was the weakest (p = .933). Therefore, patients with less hand pain reported better physical function. When we regressed the 6 scales in the MHQ against the mental function component of the SF-12, only the aesthetics scale was a significant predictor (p =.007). Patients with poor aesthetic appearance in their hands generally reported poor mental function.

Table 6.SpearmQuestions in Questionna	an's Corre the Michig aire and the	elations B an Hand e Short Fo	etween Si Outcomes orm-12	milar			
Questions in the SF-12							
МНО	ADLs	W. Perfor	ork mance*	Pain			
ADLs Work performance Pain	0.64	0.58	0.54	0.79			

MHQ, Michigan Hand Outcomes Questionnaire; SF-12, Short Form-12; ADLs, activities of daily living.

* These are the 2 questions in the SF-12 that relate to work performance.

Table 7. Independent Predictors of Short Form-12Physical Function Scale Using the Six Scales in the
Michigan Hand Outcomes Questionnaire

MHQ Scale	Beta Coefficient*	p Value	
Pain	-0.36	.001	
Satisfaction with hand function	0.24	.020	
Work performance	0.20	.033	
Activities of daily living	0.17	.087	
Overall hand function	-0.14	.147	
Aesthetics	0.005	.933	

MHQ, Michigan Hand Outcomes Questionnaire.

* The beta coefficient is a standardized coefficient that uses standard deviation as the unit of interest.

Discussion

In recent years, the United States has undergone a fundamental change in how it perceives health. Instead of emphasizing only disease processes, the new paradigm focuses on other outcomes relating to health, function, and well-being.²¹ This new paradigm acknowledges that the patients' points of view should be central¹ and it values patients' perceptions as to whether their medical treatments have improved their quality of life. Recognizing the importance of measuring outcomes in hand surgery, E. F. Shaw Wilgis,²² in his presidential address at the Forty-third Annual Meeting of the American Society for Surgery of the Hand, called for more outcomes studies to critically evaluate hand surgery procedures.

Traditionally, outcomes in hand surgery are based on "objective" measures, such as grip strength and range of motion. Although these variables are useful in assessing physical changes in the hand, they do not measure outcomes that directly affect patients' daily lives. Hand-related outcomes include (1) the ability to carry out ADLs, (2) work performance, and (3) pain control. In recent years, health services researchers have developed questionnaires that can measure and quantify these health status variables. Two of the most widely used questionnaires are the SF-36 and the Sickness Impact Profile,²³ which are generic instruments for measuring overall health. Using psychometric principles in developing these instruments, researchers have shown that these questionnaires are reliable and valid tools for the assessment of outcomes involving a broad range of diseases.24-26

To evaluate outcomes, hand surgeons need reliable and valid instruments capable of measuring factors that affect patients' everyday lives. In structuring the MHQ, we undertook several pilot studies to derive the 6 factors considered central to the measurement of outcomes for patients with hand disorders. Using the factor analysis technique, we reduced the MHQ to 37 distinct items. The self-administered MHQ could be completed in 10 minutes and patients indicated that it was an acceptable length. The ease of administration allowed the MHQ to be easily given in clinic settings, which markedly increased the response rate.

In reliability testing, the MHQ was found to have high test-retest reliability and internal consistency. For test-retest reliability, Spearman's correlation exceeded 0.9 in 9 of the 11 scales, indicating that the scales were highly reliable in repeated testing and were stable over time. The Cronbach's alphas were over 0.9 in 8 of the 11 scales in testing for internal consistency. The high Cronbach's alphas verified that all the questions in the scales were highly correlated with each other and provided assurance that random errors were minimized in the scales. However, high Cronbach's alphas could also indicate redundancy in the scales, and efforts to develop an even shorter version of the MHQ should be considered. For example, the single pain question in the SF-12 correlated highly with the pain scale in the MHQ (Spearman's correlation = 0.79); therefore, severity of pain might be measured with 1 or 2 questions.

In validity testing for the MHQ, 2 types of construct validity were examined: convergent or predictive validity and divergent or discriminate validity. Convergent validity indicated whether the MHQ displayed patterns as predicted. In testing for convergent validity, predictive relationships between the scales were proposed *a priori*.

We hypothesized that the functional scales in the MHQ would have high correlation to each other. As expected, the scales correlated significantly with each other in predictable directions (Table 5). Patients having difficulties with ADLs reported poorer overall hand function (correlation = 0.78), impaired work performance (correlation = 0.74), and less satisfaction with their hand function (correlation = 0.75). Furthermore, pain was a significant variable in predicting hand function. Patients experiencing more pain in their hands had greater functional difficulties.

Although aesthetics was considered an important outcomes variable by patients, it had much lower correlation to the other functional scales. Divergent validity was applied in testing the validity of the aesthetics scale. Divergent validity evaluated a scale's ability to discriminate between conditions. In comparing 2 groups of patients (carpal tunnel syndrome and rheumatoid arthritis) who had marked differences in their hand appearance, there also was a significant difference in their aesthetic scores. Among the 6 scales in the MHQ, the aesthetics scale was the only predictor of mental function in the SF-12.

Concurrent validity assesses the degree to which an instrument correlates with another well-established measure. We used similar items in the SF-12 to establish concurrent validity for the MHQ. As shown in Table 6, the ADL and work performance scales had moderate correlation to similar items in the SF-12. Because SF-12 is a generic measure, we expect only moderate correlation. The pain scale had a high correlation between the 2 instruments, which might suggest that the measurement of pain could be achieved using a smaller number of questions.

Among the 6 scales in the MHQ that predicted physical function, the pain scale was the strongest predictor, as patients with more pain in their hands generally reported worse physical function (Table 7). Therefore, if we direct our treatment strategies to control our patients' pain, their physical performance may improve.

There are several limitations in this study. First, this was a single-center study with patients having different types of hand conditions. The data presented might not be generalizable to other hand practices with different patient populations. Second, although the sample size was adequate for instrument testing, stratification based on types of hand conditions for validity testing was not possible; stratification would have diluted the sample size and prevented valid statistical analyses.

However, the MHQ has undergone strict psychometric testing and has performed well both in reliability and validity testing. It is easy to use and well-accepted by patients. Future studies will examine the generalizability of this outcomes instrument across different hand practices and for different hand conditions. Further validity testing will compare the MHQ to traditional objective measures, such as grip strength and hand dexterity. Responsiveness of the MHQ for different hand procedures will need to be assessed in longitudinal studies to better define its utility. Furthermore, comparison of the MHQ to currently available hand outcomes instruments^{27,28} will enhance our understanding of the strengths and weaknesses of each of these instruments.

The authors appreciate the help of Nancy K. Janz, PhD, and David L. Ronis, PhD, in reviewing the psychometric properties of the MHQ, and acknowledge the assistance of Karen Selden, MD, in collecting part of the data.

References

- Kelsey JL, Pastides H, Kreiger N, Harris C, Chernow RA. Upper extremity disorders: a survey of their frequency and cost in the United States. St. Louis, MO: CV Mosby, 1980.
- Kelsey JL, White AA, Pastides H, Bisbee GE. The impact of musculoskeletal disorders on the population of the United States. J Bone Joint Surg 1979;61A:959–964.
- 3. Chase RA. Costs, risks, and benefits of hand surgery. J Hand Surg 1983;8:644-648.

- Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–483.
- Meenan RF, Mason JH, Anderson JJ, Guecione AA, Kazis LE. AIMS2: the content and properties of a revised and expanded arthritis impact measurement scales health status questionnaire. Arthritis Rheum 1992;35:1–10.
- 6. Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30:191–197.
- Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Joint Surg 1993;75A:1585–1615.
- Bass BM, Cascio WF, O'Connor EJ. Magnitude estimations of expression of frequency and amount. J Appl Psychol 1974;59:313–320.
- Kim J, Mueller CW. Factor analysis: statistical methods and practical issues. Newbury Park, CA: SAGE Publications, 1978.
- Kim J, Mueller CW. Introduction to factor analysis: what it is and how to do it. Newbury Park, CA: SAGE Publications, 1978.
- Long JS. Confirmatory factor analysis. Newbury Park, CA: SAGE Publications, 1983.
- Peters M, Murphy K. Factor analyses of pooled hand questionnaire data are of questionable value. Cortex 1993; 29:305–314.
- Ware JE, Kosinski M, Keller SD, A 12-item short-form health survey, construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:3:220-233.
- Stata Corp. Stata Statistical Software: Release 5.0. College Station, TX: Stata Corporation, 1997.
- Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. Clin Pharmacol Ther 1972; 29:111–123.
- Carmines EG, Zeller RA. Reliability and validity assessment. Newbury Park, CA: SAGE Publications, 1979.
- DeVellis RF. Reliability. In: Scale development. Newbury Park, CA: Sage Publications, 1991:24-42.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420--428.
- Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical settings. Med Care 1992;30:MS23–MS41.
- 22. Wilgis EFS. Medical communication: its uses and abuses. J Hand Surg 1989;14A:425-428.
- 23. Bergner M, Bobbitt RA, Pollard WE, Martin DP, Gilson BS. The Sickness Impact Profile: validation of a health status measure. Med Care 1976;14:57-67.
- 24. Bell MJ, Bombardier C, Tugwell P. Measurement of functional status, quality of life, and utility in rheumatoid arthritis. Arthritis Rheum 1990;33:591-601.
- 25. Greenfield L, Cohen BD, Cleary PD, Greenfield S. Evaluating the quality of health care: what research offers

decision makers. New York, NY: Milbank Memorial Fund, 1996.

- Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical settings. Med Care 1992;30:MS23–MS41.
- Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The Upper Extremity Collaborative Group (UECG). Am J Ind Med. 29:602–

608, 1996 [published erratum appears in Am J Ind Med 1996;30:372].

- Stock SR, Cole DC, Tugwell P, Streiner D. Review of applicability of existing functional status measures to the study of workers with musculoskeletal disorders of the neck and upper limb. Am J Ind Med 1996;29:679-688.
- 29. Woolson RF. Analysis for variance for company groups. In: Statistical methods for the analysis of biomedical data. New York, NY: John Wiley, 1987:337.

Appendix 1: Scoring Mechanism

The Michigan Hand Outcomes Questionnaire (MHQ) contains 6 scales: (1) overall hand function, (2) activities of daily living. (3) work performance, (4) pain, (5) aesthetics, and (6) satisfaction with hand function. In the pain scale, high scores indicate greater pain; in the other 5 scales, high scores denote better hand performance.

The raw scale score for each of the 6 scales is the sum of the responses of each scale item. The raw score is converted to a score range from 0 to 100. The scoring equation for each of the scales is listed below.

The score for the affected hand is obtained by

selecting either the right- or the left-hand score. If both hands are affected (eg, rheumatoid patients), the right- and left-hand scale scores are averaged to get the score.

Missing values in each scale may affect the validity of the scores. If 50% or more of the items in a scale are missing, then that particular scale cannot be scored. For scales with less than 50% missing, the average of the existing scale items may be imputed for the missing items. An overall MHQ score can be obtained by summing the scores for all 6 scales and divide by 6. If scores for more than 2 scales are missing, an overall MHQ score cannot be computed.⁸

		2 Scoring mgorin	
Scale	Recodet	Raw Score Range+	Normalization§
Overall hand function	None	5 to 25	-(raw score -25)/20*100
Activities of daily living	None	5 to 25 I-handed	= -(raw score -25)/20*100
		7 to 35 2-handed	= -(raw score -35)/28*100
		Overall ADL	= (1-handed+2-handed)/2
Work	None	5 to 25	(raw score -5)/20*100
Pain	Question 2: $(1 = 5) (2 = 4)$ (4 = 2) (5 = 1)	5 to 25	If question $1 = 5$, then pain score = 0; if question $1 \neq 5$, then $-(\text{raw score } -25)/20*100$
Aesthetics	Question I: $(1 = 5) (2 = 4)$ (4 = 2) (5 = 1)	4 to 16	(raw score -4)/16*100
Satisfaction	None	6 to 30	-(raw score -30)/24*100

MHQ Scoring Algorithm^{*}

* The scoring algorithm is available from the authors in SAS program.

The response categories for some of the questions are reversed and are recoded.

Sum of the responses for each scale.

§ For the pain scale, higher scores indicate more pain. For the other 5 scales, higher scores indicate better hand performance. The scores are normalized to a range of 0 to 100.

Appendix 2: Michigan Hand Outcomes Questionnaire

Instructions: This survey asks for your views about your hands and your health. This information will help keep track of how you feel and how well you are able to do your usual activities. Answer *every* question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

- I. The following questions refer to the function of your hand(s)/wrist(s) *during the past week*. (Please circle 1 answer for each question.)
- A. The following questions refer to your *right* hand/wrist.

	Very Good	Good	Fair	Poor	Very Poor
1. Overall, how well did your <i>right</i> hand work?	1	2	3	4	5
2. How well did your <i>right</i> fingers move?	1	2	3	4	5
3. How well did your <i>right</i> wrist move?	1	2	3	4	5
4. How was the strength in your <i>right</i> hand?	1	2	3	4	5
5. How was the sensation (feeling) in your <i>right</i> hand?	1	2	3	4	5

B. The following questions refer to your *left* hand/wrist.

	Very Good	Good	Fair	Poor	Very Poor
1. Overall, how well did your <i>left</i> hand work?	I	2	3	4	5
2. How well did your <i>left</i> fingers move?	1	2	3	4	5
3. How well did your <i>left</i> wrist move?	1	2	3	4	5
4. How was the strength in your <i>left</i> hand?	1	2	3	4	5
5. How was the sensation (feeling) in your <i>left</i> hand?	1	2	3	4	5

II. The following questions refer to the ability of your hand(s) to do certain tasks *during the past week*. (Please circle 1 answer for each question.)

A. How difficult was it for you to perform the following activities using your *right hand?*

	Not at All Difficult	A Little Difficult	Somewhat Difficult	Moderately Difficult	Very Difficult
1. Turn a door knob	1	2	3	4	5
2. Pick up a coin	1	2	3	4	5
3. Hold a glass of water	1	2	3	4	5
4. Turn a key in a lock	1	2	3	4	5
5. Hold a frying pan	I	2	3	4	5

B. How difficult was it for you to perform the following activities using your *left hand?*

	Not at All Difficult	A Little Difficult	Somewhat Difficult	Moderately Difficult	Very Difficult
1. Turn a door knob	1	2	3	4	5
2. Pick up a coin	1	2	3	4	5
3. Hold a glass of water	1	2	3	4	5
4. Turn a key in a lock	1	2	3	4	5
5. Hold a frying pan	1	2	3	4	5

C. How difficult was it for you to perform the following activities using both of your hands?

	Not at All Difficult	A Little Difficult	Somewhat Difficult	Moderately Difficult	Very Difficult
I. Open a jar	1	2	3	4	5
2. Button a shirt/blouse	1	2	3	4	5
3. Eat with a knife/fork	1	2	3	4	5
4. Carry a grocery bag	1	2	3	4	5
5. Wash dishes	1	2	3	4	5
6. Wash your hair	1	2	3	4	5
7. Tie shoelaces/knots	1	2	3	4	5

	Always	Often	Sometimes	Rarely	Never
1. How often were you unable to do your work					
because of problems with your					
hand(s)/wrist(s)?	1	2	.3	4	5
2. How often did you have to shorten your work	κ.				
day because of problems with your hand(s)/					
wrist(s)?	1	2	3	4	5
3. How often did you have to take it easy at you	ır				
work because of problems with your hand(s)/					
wrist(s)?	I	2	3	4	5
4. How often did you accomplish less in your		-		,	
work because of problems with your hand(s)/					
wrist(s)?	ī	2	3	4	5
5 How often did you take longer to do the task	c	5	.,	-	~
in your work because of problems with your	3				
hand(a)/muist(a)?	t	7		4	5
nand(s)/wrist(s)?	1	2	3	4	2

III.	The following	questions	refer to h	now you d	lid in yo	ir <i>norma</i>	l work	(including	both	housework	and	school
	work) during t	the past 4	weeks. (I	Please cir	cle 1 ans	wer for e	ach qu	estion.)				

- IV. The following questions refer to how much *pain* you had in your hand(s)/wrist(s) *during the past week*. (Please circle 1 answer for each question.)
- 1. How often did you have pain in your hand(s)/wrist(s)?
 - 1. Always
 - 2. Often
 - 3. Sometimes
 - 4. Rarely
 - 5. Never

If you answered *never* to *question IV-1* above, please skip the following questions and go to the next page.

2. Please describe the pain you have in your hand(s)/wrist(s).

- 1. Very mild
- 2. Mild
- 3. Moderate
- 4. Severe
- 5. Very severe

		Always	Often	Sometimes	Rarely	Never
3.	How often did the pain in your hand(s)/wrist(s) interfere					
	with your sleep?	I	2	3	4	5
4.	How often did the pain in your hand(s)/wrist(s) interfere					
	with your daily activities (such as eating or bathing)?	1	2	3	4	5
5.	How often did the pain in your hand(s)/wrist(s) make					
	you unhappy?	1	2	3	4	5

586 Chung et al. / Michigan Hand Outcomes Questionnaire

V. A.	The following	questions	refer to	the	appearance	(look)	of	your	right	hand	during	the	past	week.
	(Please circle	1 answer f	for each	ques	stion.)									

		Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
1.	1 was satisfied with the appearance (look) of my <i>right</i> hand.	1	2	3	4	5
2.	The appearance (look) of my <i>right</i> hand sometimes made me					
	uncomfortable in public.	1	2	3	4	5
3.	The appearance (look) of my <i>right</i> hand made me depressed.	I	2	3	4	5
4.	The appearance (look) of my <i>right</i> hand interfered with my					
	normal social activities	Ι	2	3	4	5

B. The following questions refer to the appearance (look) of your *left* hand during the past week. (Please circle 1 answer for each question.)

	Neither					
	Agree	Agree	Agree Nor Disagree	Disagree	Disagree	
1 was satisfied with the appearance (look) of my						
<i>left</i> hand.	1	2	3	4	5	
The appearance (look) of my <i>left</i> hand sometimes						
made me uncomfortable in public.	1	2	3	4	5	
The appearance (look) of my <i>left</i> hand made me						
depressed.	1	2	3	4	5	
The appearance (look) of my <i>left</i> hand interfered						
with my normal social activities	1	2	3	4	5	
	 I was satisfied with the appearance (look) of my <i>left</i> hand. The appearance (look) of my <i>left</i> hand sometimes made me uncomfortable in public. The appearance (look) of my <i>left</i> hand made me depressed. The appearance (look) of my <i>left</i> hand interfered with my normal social activities 	Strongly Agree 1 was satisfied with the appearance (look) of my 1 <i>left</i> hand. 1 The appearance (look) of my <i>left</i> hand sometimes 1 made me uncomfortable in public. 1 The appearance (look) of my <i>left</i> hand made me 1 depressed. 1 The appearance (look) of my <i>left</i> hand interfered 1 with my normal social activities 1	Strongly AgreeAgreeAgreeAgreeAgreeAgreeI was satisfied with the appearance (look) of my left hand.12122The appearance (look) of my left hand sometimes made me uncomfortable in public.12122The appearance (look) of my left hand made me 	Neither Strongly AgreeNeither Agree $Agree$ $Agree$ $Agree$ $Agree$ $Agree$ $Nor Disagree$ I was satisfied with the appearance (look) of my left hand.12123The appearance (look) of my left hand sometimes made me uncomfortable in public.123123The appearance (look) of my left hand made me depressed.123The appearance (look) of my left hand interfered with my normal social activities123	Neither Strongly AgreeNor DisagreeDisagree1was satisfied with the appearance (look) of my left hand.1234123441234412344123441234412344123441234412344123441234412344	

VI. A. The following questions refer to your satisfaction with your *right* hand/wrist during the past week. (Please circle 1 answer for each question.)

		Very Satisfied	Somewhat Satisfied	Neither Satisfied Nor D issatisfied	Somewhat Dissatisfied	Very Dissatisfied
1.	Overall function of your <i>right</i> hand	ł	2	3	4	5
2.	Motion of the fingers in your right					
	hand	1	2	3	4	5
3.	Motion of your <i>right</i> wrist	1	2	3	4	5
4.	Strength of your <i>right</i> hand	1	2	3	4	5
5.	Pain level of your <i>right</i> hand	1	2	3	4	5
6.	Sensation (feeling) of your right					
	hand	1	2	3	4	5

B. The following questions refer to your satisfaction with your *left* hand/wrist during the past week. (Please circle 1 answer for each question.)

		Very Satisfied	Somewhat Satisfied	Neither Satisfied Nor Dissatisfied	Somewhat Dissatisfied	Very Dissatisfied
1.	Overall function of your <i>left</i> hand Motion of the fingers in your <i>left</i>	1	2	3	4	5
2.	hand	1	2	3	4	5
3.	Motion of your <i>left</i> wrist	1	2	3	4	5
4.	Strength of your <i>left</i> hand	1	2	3	4	5
5.	Pain level of your <i>left</i> hand	1	2	3	4	5
6.	Sensation (feeling) of your <i>left</i> hand	1	2	3	4	5

Please provide the following information about yourself. (Please circle 1 answer for each question.)

1. Are you right-handed or left-handed?

- a. Right-handed
- b. Left-handed
- c. Both
- 2. Which hand gives you the most problem?
 - a. Right hand
 - b. Left hand
 - c. Both
- 3. Have you changed your job since you had problem with your hand(s)?
 - a. Yes
 - b. No

Please describe the type of job you did *before* you had problem with your hand(s)._____

Please describe the type of job you are doing now._____

- 4. What is your gender?
 - a. Male
 - b. Female
- 5. What is your ethnic background?
 - a. White
 - b. Black
 - c. Hispanic
 - d. Asian or Pacific Islander
 - e. American Indian or Alaskan Native
 - f. Other (please specify) _
- 6. What is the highest level of education you received?
 - a. Less than high school graduate
 - b. High school graduate
 - c. Some college
 - d. College graduate
 - e. Professional or graduate school
- 7. What is your approximate family income, including wages, disability payment, retirement income, and welfare?
 - a. <\$10,000
 - b. \$10,000-\$19,999
 - c. \$20,000-\$29,999
 - d. \$30,000-\$39,999
 - e. \$40,000-\$49,999
 - f. \$50,000-\$59,999
 - g. \$60,000-\$69,999
 - h. >\$70,000
- 8. Is your injury covered by Workers' Compensation?
 - a. Yes
 - b. No